
Sensitivity Analysis for Atmospheric Infrared Sounder (AIRS) CO_2 Retrieval

Student:

Ilana GAT

Mentors:

Dr. Zhijin LI, JPL

Dr. Paul DIMOTAKIS, Caltech/JPL

SURF: FINAL REPORT

JET PROPULSION LABORATORY

CALIFORNIA INSTITUTE OF TECHNOLOGY

September 27, 2012

Abstract

The Atmospheric Infrared Sounder (AIRS) is a thermal infrared sensor able to retrieve the daily atmospheric state globally for clear as well as partially cloudy field-of-views. The AIRS spectrometer has 2378 channels sensing from 15.4 μm to 3.7 μm , of which a small subset in the 15 μm region has been selected, to date, for CO_2 retrieval. To improve upon the current retrieval method, we extended the retrieval calculations to include a prior estimate component and developed a channel ranking system to optimize the channels and number of channels used. The channel ranking system uses a mathematical formalism to rapidly process and assess the retrieval potential of large numbers of channels. Implementing this system, we identified a larger optimized subset of AIRS channels that can decrease retrieval errors and minimize the overall sensitivity to other iridescent contributors, such as water vapor, ozone, and atmospheric temperature. This methodology selects channels globally by accounting for the latitudinal, longitudinal, and seasonal dependencies of the subset. The new methodology increases accuracy in AIRS CO_2 as well as other retrievals and enables the extension of retrieved CO_2 vertical profiles to altitudes ranging from the lower troposphere to upper stratosphere. The extended retrieval method for CO_2 vertical profile estimation using a maximum-likelihood estimation method. We use model data to demonstrate the beneficial impact of the extended retrieval method using the new channel ranking system on CO_2 retrieval.

Contents

1	Introduction	3
2	Approach	3
2.1	Channel Selection	3
2.2	Maximum Likelihood Estimation for Retrieval	4
2.3	Observing System Simulation Experiments	5
3	Results	6
4	Discussion	7
5	Methods	9
5.1	Previous Methods	9
5.2	Standalone AIRS Radiative-Transfer Algorithm (SARTA)	9
5.3	Channel Selection Method	10
5.4	CO_2 Retrieval Methods	12
5.4.1	Maximum-Likelihood Estimation	12
5.4.2	Minimum Error Variance	27
5.5	MLE vs MEV	28
5.6	Dimension of the Observation	28
5.7	Retrieval of CO_2	30
5.7.1	Retrieved x_{CO_2} Equation	30
5.7.2	Effects of Background and Observational Error on Retrieval Error	31
5.7.3	Background term, B_{CO_2}	31
	Acknowledgments	33
	References	34

1 Introduction

The effects of anthropogenic emissions of various greenhouse gases (GHG), specifically CO_2 , on the atmosphere have yet to be completely determined as the record of the atmosphere's molecular composition is sparse [1]. Developing a comprehensive atmospheric record requires accurate global measurements with high spatial and temporal resolutions which involves using multiple types of sensors such as land, ocean, air, and/or space based sensors. One space based sensor currently in use is the Atmospheric Infrared Sounder—the first orbiting sensor able to gather daily data globally.

In 2002, the Atmospheric Infrared Sounder (AIRS) instrument was launched on NASA's Aqua satellite in a sun-synchronous near-polar orbit [2]. AIRS is the first sensor able to gather daily data that enabled measurements of CO_2 concentration globally (including land, water, and polar regions), extending to coverage under partial cloud cover [3]. Large amounts of data are gathered daily and extending the utilization of this valuable resource is important. At present, only some of the data are used in CO_2 retrieval, bringing up the interesting question: given a large data set, what are the best subsets to analyze in order to optimize accuracy of the results? Data subset selection problems are a topic of much research, as ever more sophisticated sensors are developed with the ability to generate even more data. Currently, there are methods to narrow down all the data gathered from AIRS [3], but ultimately, chosen data sets are resultants of the user's discretion. Data-subset selection has accuracy implications heading sensitivity analysis and uncertainty quantification of the results.

Once data sets are chosen to analyze, a Vanishing Partial Derivatives (VPD) algorithm is utilized for CO_2 retrieval [4]. This method involves many iterations of data perturbations and radiation calculations until the result reaches a minimum, making it a computationally intensive solution. In addition to the computational cost, there is currently no retrieval error quantification for this method, and the retrieved global concentrations are averaged over mid tropospheric layers on a 1° by 1° grid, with stratospheric-layer retrievals in progress.

This project aimed to improve upon the current CO_2 retrieval methodology to obtain higher accuracy in the results as well as retrieval error estimations. We were able to accomplish these goals by developing a new channel selection process and extending the current retrieval algorithm. It is important to note that this new channel selection process was implemented for higher accuracy in the extended retrieval algorithm discussed in this report. We discuss the approach and methodology briefly in Section 2, the findings and results of the project are explained in Section 3, which are discussed then in Section 4. The methodology used is thoroughly explained in Section 5. Throughout the report, we use pressure as a surrogate for altitude, and troposphere to mean altitude ranging from roughly 100 to 850 hPa, and stratosphere representing altitudes ranging from about 1 to 100 hPa.

2 Approach

There are two main contributions from this project: the methodology for channel selection, and the extended retrieval methodology, both of which are required for optimal retrieval results. The selected channels dictate the precision of the data, whereas the retrieval method dictates the accuracy of the overall result.

2.1 Channel Selection

The atmosphere contains particles which scatter and absorb light at various frequencies, and the amount of light transmitted, absorbed, or reflected is a result of the composition of the atmosphere. An observer high in the

atmosphere, or in orbit about the planet, can detect the wavelengths of light, their individual intensities, and use basic spectroscopy to discern what comprises the atmosphere. This is the general theory behind the AIRS instrument.

The AIRS instrument has 2378 infrared channels each sensing the radiance of a particular wavelength. In order to retrieve the vertical profile of an element, the data used must be from channels sensing wavelengths corresponding to that iridescent contributor. To test how sensitive each channel is to various elements, the AIRS team has developed a model called SARTA (Standalone AIRS Radiative-Transfer Algorithm) which is used to calculate the sensitivity matrix for each channel, known as the Jacobian, \mathbf{H}_{CO_2} . The exact methods for these calculations are discussed in more detail in Section 5.3.

$$\mathbf{H}_{CO_2} = \left[\frac{\partial \theta_{\nu_1}(x_T^b, x_q^b, x_{O_3}^b, x_{CO_2}^b)}{\partial x_{CO_2}^t}, \frac{\partial \theta_{\nu_2}(x_T^b, x_q^b, x_{O_3}^b, x_{CO_2}^b)}{\partial x_{CO_2}^t}, \dots, \frac{\partial \theta_{\nu_m}(x_T^b, x_q^b, x_{O_3}^b, x_{CO_2}^b)}{\partial x_{CO_2}^t} \right] \quad (2.1)$$

θ_ν in the above equation represents the radiance calculated for the channel sensing the wavelength ν , and is a function of the background temperature, T , water vapor, q , ozone, O_3 , and carbon dioxide, CO_2 . The superscript of “b” represents the background data, whereas the superscript of “t” represents the “truth.”

The channels were all ranked based on their maximum sensitivity to the element being tested, and the top 80 channels with maximum sensitivities in the mid troposphere and stratosphere were chosen.

2.2 Maximum Likelihood Estimation for Retrieval

There are various optimization techniques that can be used to retrieve the vertical profile of an iridescent contributor, a couple of which are described in Section 5.4. For this application, the maximum likelihood estimation (MLE) is optimal and the reasoning behind this is also discussed in Section 5.4. The MLE method uses Bayesian statistics, and the assumption that the errors all follow a Gaussian distribution, to derive a cost function, $J(x)$, where x is the vertical profile of the variable being retrieved. Equation 2.2 uses the subscript of CO_2 because this project focused on the retrieval of CO_2 , but the subscript could be other elements depending on the retrieval.

$$J(x_{CO_2}) = \frac{1}{2} \underbrace{(x_{CO_2} - x_{CO_2}^b)^T \mathbf{B}_{CO_2}^{-1} (x_{CO_2} - x_{CO_2}^b)}_{\text{Prior estimate component}} + \frac{1}{2} \underbrace{(\theta^{\text{obs}} - \mathbf{H}_{CO_2} x_{CO_2})^T \mathbf{R}^{-1} (\theta^{\text{obs}} - \mathbf{H}_{CO_2} x_{CO_2})}_{\text{Observational component}} \quad (2.2)$$

Equation 2.2 introduces some important variables and notation. The superscript of “b” means “background,” which can also be called the “prior;” it is data gathered from other sources prior to the retrieval process. The matrix \mathbf{B} is the background-error covariance, meaning it is the error in the “prior” data. A superscript of “obs” stands for “observational,” and θ^{obs} is the radiance observed (from the channels used). \mathbf{R} is the radiance observational-error covariance, or the instrument error. The bracket labeled “prior estimate component” is the component of the cost function that relies solely on the prior knowledge and error, or the aleatoric uncertainties. The second part of the cost function relies on the observed, or experimental, radiances, the model of the system—the Jacobian sensitivity matrix, and the instrument error, which gives it the label of the “observational component.” The Jacobian matrix represents our model, and the observational error covariance is a result of calibration, meaning the observational component represents the epistemic uncertainties.

An important property of this cost function is that the mean value, or the expected value, $\langle J(x) \rangle = \frac{n}{2}$, where

n is the number of channels used for the retrieval¹, which means that as long as the number of channels used is held constant, the mean value of the cost function is constant. This concept introduces the balancing act between the prior estimate component and the observational component of the cost function. If there is a larger error in the observation, the observational component would be smaller, and the retrieval would rely more heavily on the background data. If the background data has a high error, then the prior estimate component would be smaller, resulting in the retrieval relying more heavily on the observation. The first term in Equation 2.2 is actually called the “degrees of freedom for signal” and the second term is called the “degrees of freedom for noise” [5]. In short, if there is a small error in the observation/background, the “true” value is known within a small error, or the probability density function, *pdf*, has a narrow width, and the retrieval will lie somewhere between the observation and the background, which is shown in Figure 2.1.

Knowing the impact of the background error realizes the importance of estimating \mathbf{B} correctly². In our results section, we use a constant background error, as well as a smooth background error decreasing quadratically with the logarithmic decrease in pressure (increase in altitude). The motive for choosing two differing background errors is to show the effect of background on the retrieval. The quadratically decreasing error is a fit to what our calculated background error was. These calculations are described in more detail in Section 5.7.3.

Once all the terms (excluding the term we are solving for, x_{CO_2}) are known, we can solve for the retrieved vertical profile of CO₂, x_{CO_2} .

$$x_{CO_2} = x_{CO_2}^b - \Delta x_{CO_2} \quad (2.3)$$

$$\Delta x_{CO_2} = \mathbf{P}_{CO_2} \mathbf{H}_{CO_2}^T \mathbf{R}^{-1} \Delta \theta \quad (2.4)$$

$$\mathbf{P}_{CO_2} = (\mathbf{B}_{CO_2}^{-1} + \mathbf{H}_{CO_2}^T \mathbf{R}^{-1} \mathbf{H}_{CO_2})^{-1} \quad (2.5)$$

The above three equations are derived in Section 5.7.1. \mathbf{P}_{CO_2} in Equation 2.5 is the retrieved error covariance.

2.3 Observing System Simulation Experiments

We used an Observing System Simulation Experiment (OSSE) based approach to test the impact of the new selected channels and the MLE method on the retrieval. Our assumed “true” model data was MOZART (Model for Ozone And Related chemical Tracers) data [6], and the instrument model was represented by the observational error covariance. We added a biased error to the “true” profile which was used as our “prior” profile in

¹The derivation of this is in Section 5.6.

²It is also important to have \mathbf{R} estimated correctly, but this is the error in the instrument—a known and tested quantity.

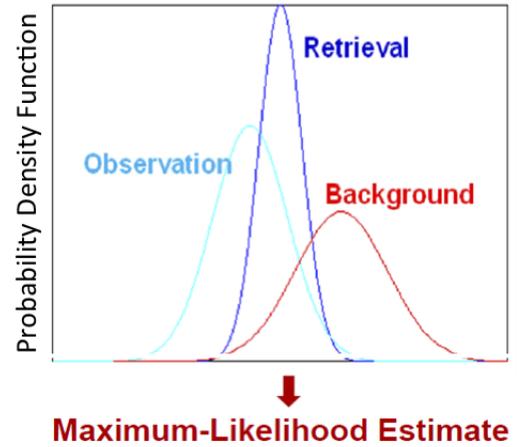


Figure 2.1: Under the assumption that the error distributions are Gaussian, the distribution of where the retrieval will lie, or the probability distribution function, *pdf*, will be in between the *pdf* of the background and the *pdf* of the observation.

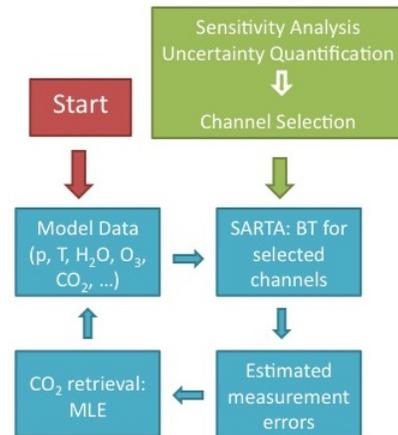


Figure 2.2: OSSE method used for testing methodology. SARTA is the Standalone AIRS Radiative-Transfer Algorithm, and BT stands for brightness temperature.

the retrieval process. The vertical profile was then computed with the MLE method, which was compared to the “true” profile. A flow chart illustrating this method is shown in Figure 2.2.

3 Results

The new channel selection method produces channels that, on average, have higher sensitivity to CO_2 , higher resolution in the sensitivity to CO_2 , and lower sensitivities to H_2O , O_3 , and T_{air} . Figure 3.1 compares the averages of the Jacobian sensitivity matrices for the new selected channels versus the previous selected channels for those with maximum sensitivities in the mid troposphere and stratosphere over a tropical area in the Pacific Ocean. One important finding was that the new list of channels incorporated some of the channels from the previous list, but not all.

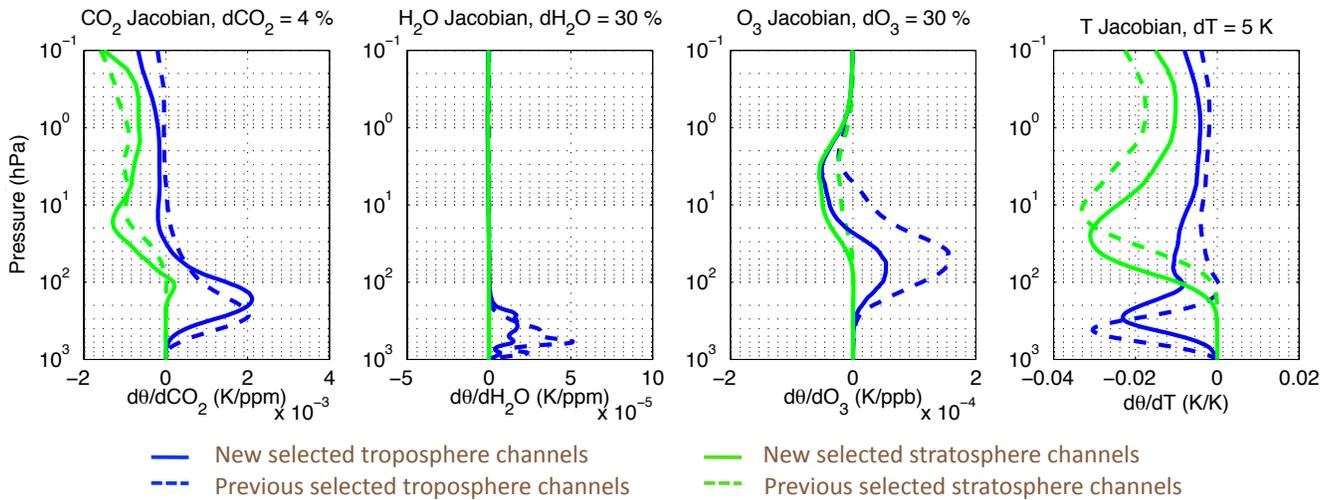
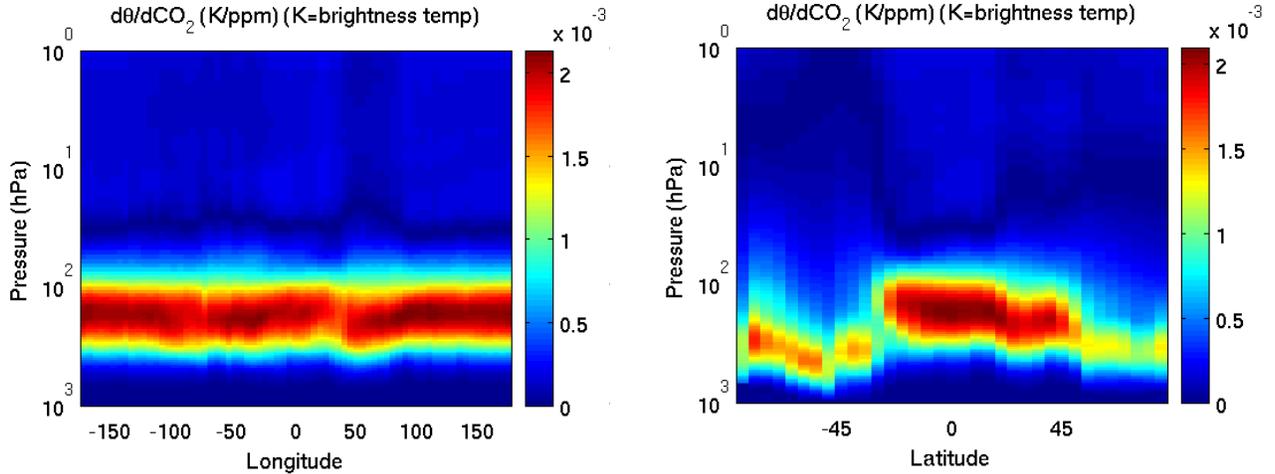


Figure 3.1: These are plots of the average Jacobian sensitivity matrices of selected channels at the equator over the Pacific Ocean. The Jacobians of the new selected troposphere channels, represented by the solid blue line, have a higher maximum sensitivity to CO_2 , as well as lower sensitivity to H_2O , and T relative to the dashed blue lines representing the Jacobians for the previous selected troposphere channels. For O_3 , the Jacobians have lower sensitivity at around 50 hPa, but are more sensitive at around 5 hPa. This is not optimal, but the difference is very small—on the order of 10^{-4} K/ppb, whereas the sensitivity to CO_2 is on the order of 10^{-3} K/ppm, making this slight increase relatively insignificant. As for the green lines, the solid line (representing the new selected stratosphere channels) has a higher maximum with respect to CO_2 than the dashed line (representing the previous stratosphere channels) at around 11 hPa, but dips under the dashed line above that. This means that we will have higher resolution in the stratosphere since there is a defined maximum in the sensitivity versus a relatively constant sensitivity throughout. Again, we minimize the sensitivity to H_2O and T with the new stratosphere channels, and increase the sensitivity slightly to O_3 , but like the troposphere channels, this slight increase is negligible.

We found that the list of optimal channels is dependent on latitude, season, and partially on longitude—the overall list of optimal channels does not change much with longitude, but the ordering of optimal channels does. Also, the shape of the Jacobian sensitivity matrix varies with respect to the same parameters of latitude, longitude, and season. Figure 3.2 depicts the changes in the Jacobian sensitivity matrix over latitudes and longitudes.

Once we were satisfied with the criteria for the channel selection process and the resultant sets of channels,



(a) CO_2 Jacobians near the equator, longitude of -2.5° . (b) CO_2 Jacobians over the Pacific Ocean, latitude of -177.5° .

Figure 3.2: Average of the CO_2 Jacobians for the new selected mid troposphere channels over latitude and longitude. From these figures, it is apparent that the maximum sensitivity has minimal variations over longitude, but large variations, even variations in altitude of maximum sensitivity, over latitude.

we continued on in the OSSE based method to calculate the brightness temperatures for the selected channels, estimate the measurement errors, and retrieve the vertical profiles of CO_2 .

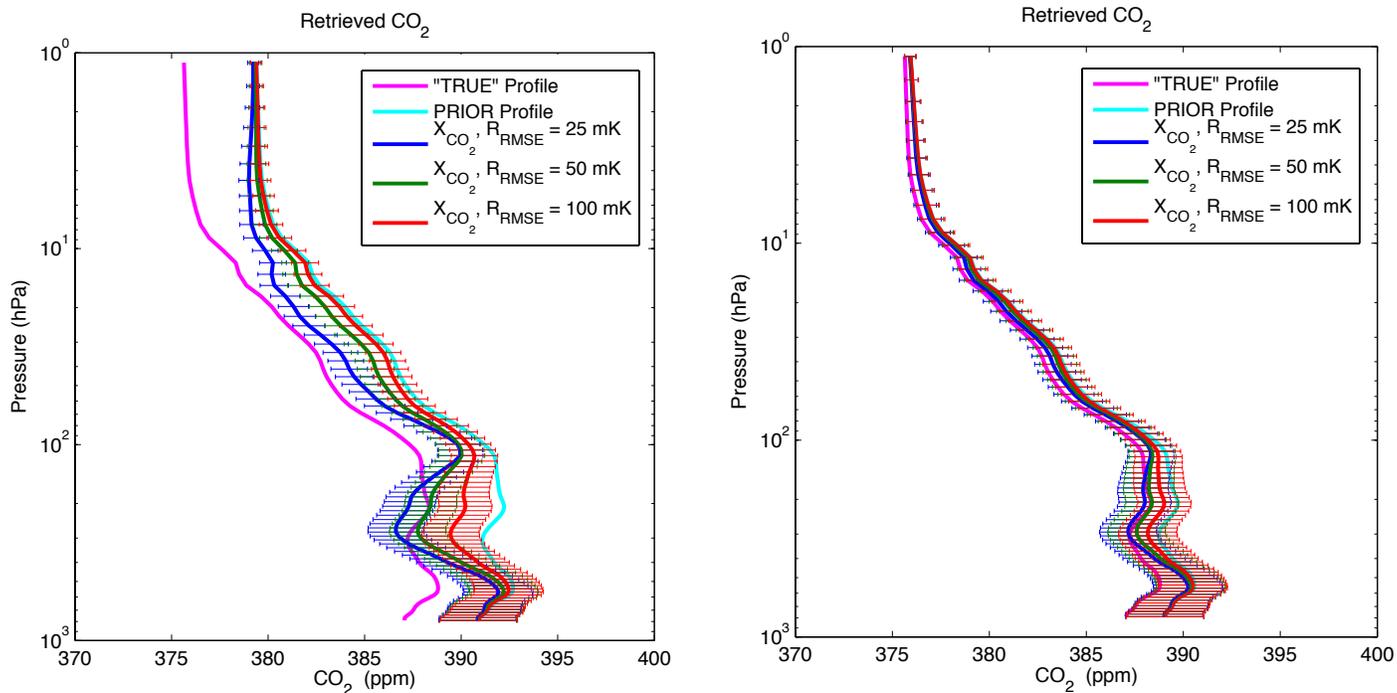
The measurement errors are known within some degree—they are between 25 to 100 mK, and so we calculated the retrieved vertical profile using $R_{RMSE} = 25, 50, 100$ mK, to show the effects of the observational error. The results using the new selected channels are shown in Figure 3.3.

The retrieved CO_2 profiles do deviate from the “prior” profile and come close to the “true” profile, at altitudes at which the Jacobians have high sensitivities. It is important to emphasize that we only used channels with sensitivities in the mid troposphere and stratosphere, which explains altitudes at which the retrieved profile deviates little to no deviation at all from the “prior.” Figure 3.4 shows the retrieved vertical profile using the previous selected channels, which can be compared to Figure 3.3(a) to conclude that using the new selected channels result in a more accurate retrieval.

4 Discussion

By introducing the new channel selection and the MLE methods for retrieval, we are able to obtain higher accuracy in the CO_2 vertical profile as well as a quantitative estimation of retrieval uncertainties. The MLE algorithm also avoids the computationally intensive iterative solution using SARTA that the VPD method requires.

The results of this project show an example of how this new methodology changes the retrievals, and work still needs to be done to implement AIRS data instead of model data. In addition, channels were selected only for the stratosphere and mid troposphere, which will be extended to find channels at more altitudes, extending to the lower troposphere.



(a) Retrieved CO_2 vertical profiles for varying \mathbf{R} matrices using a background error matrix corresponding to 1% error at 2.5° south and 177.5° west.

(b) Retrieved CO_2 vertical profiles for varying \mathbf{R} matrices using a fitted background error matrix at 2.5° south and 177.5° west.

Figure 3.3: Retrieved CO_2 vertical profiles for the two types of backgrounds using the new selected channels.

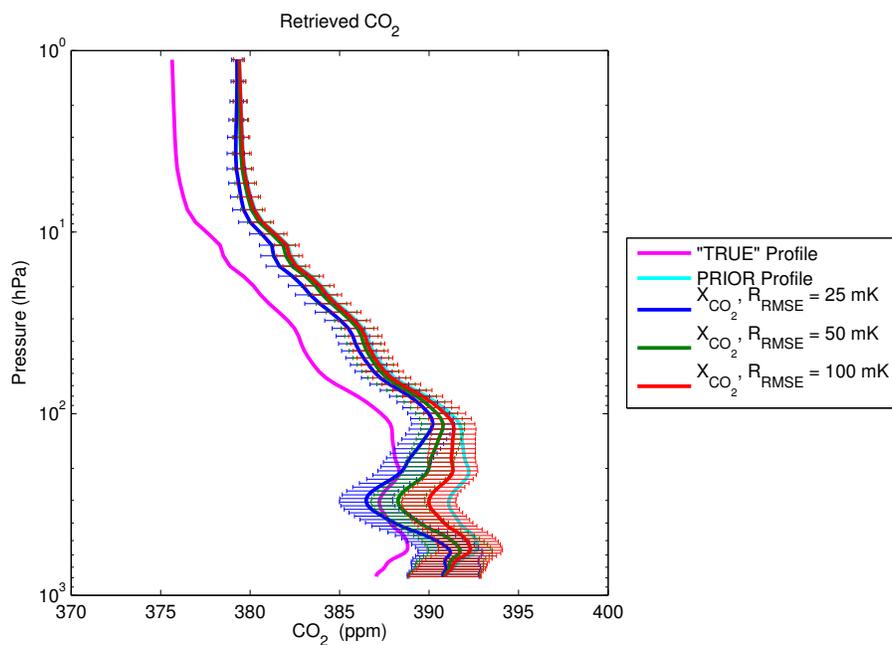


Figure 3.4: Retrievals of CO_2 using the previous selected channels at 2.5° south and 177.5° west. This assumes a constant background error variance of 1%.

5 Methods

Throughout the previous sections, the methodology used was only briefly explained to ensure the mathematical details did not usurp the focus of the report. In this section, we go in to grave details of the derivations of formulas and explanations of methods and algorithms previously mentioned.

5.1 Previous Methods

Channels used in retrieval were previously selected based on a weighting function and a contribution function. The weighting function was a measure of the change in transmittance with respect to the change in pressure, which has a result of a relatively Gaussian shaped curve with a peak at a certain altitude. This weighting function is a function of all iridescent contributors, and consequently, there is no information about the altitude of maximum sensitivity to a particular variable. Each channel has a corresponding weighting function, which gives the altitude at which the channel has maximum sensitivity to all variables. This is how channels were chosen for particular atmospheric levels.

The contribution function is a function of the weighting function as well as the brightness temperature associated with each channel. US Standard Atmosphere was used, with CO_2 as a constant column mixing ratio of 370 ppmv (parts per million volume). The radiances were calculated for channels sensing wavelengths known to interact with CO_2 , and then the CO_2 column mixing ratio was shifted by a constant 15 ppmv to get the change of radiance associated with those channels. This change in radiance was used with the weighting function to obtain the contribution function. Channels with the highest maximum in their contribution function were chosen for CO_2 retrieval at the altitude of maximum contribution. About 60 channel were chosen out of the 2378 for CO_2 retrieval, with roughly 20 channels for the general stratosphere and 14 for the troposphere.

The selected channels were then used in the retrieval process which consisted of a Vanishing Partial Derivatives (VPD) algorithm which was briefly explained in the Introduction section. The VPD methodology uses background data to initiate the iterative steps, but it does not give any weight to the priors in the retrieval process (it consists of Equation 2.2 without the prior estimate component). With this algorithm, the method for channel selection described in this section may be optimal because VPD does not work to calculate the departure of the retrieved CO_2 from the prior CO_2 while all other variables are held constant, as the present version of the MLE method does. Yet the VPD method is extremely computationally expensive, and it does not result in any information about retrieval errors. Also, the retrieved global concentrations were averaged over mid tropospheric layers, with the stratospheric layer retrievals still in progress.

With the current computational means and updated retrieval method, it was important to reselect channels by considering *all* channels and test the addition of the new retrieval method on computational time and estimated retrieval errors.

5.2 Standalone AIRS Radiative-Transfer Algorithm (SARTA)

In order to retrieve the vertical profile of any iridescent contributor using the brightness temperature (or radiance) measured from an instrument, there needs to be a model representing the relation between the radiance observed and the element's concentration. A type of model that gives the relation between an output to various inputs is called a "forward" model, which is what the AIRS team has developed and named "SARTA" (Standalone AIRS

Radiative-Transfer Algorithm).

Given a certain model atmosphere (for this project, we used MOZART as our model), SARTA calculates the radiance that each channel would sense, at the top of the atmosphere (TOA), if this model data were true. Equation 5.1 shows SARTA's calculation in equation form, where $\epsilon_\nu B_\nu(T_s)\tau_\nu(p_s)$ is the surface contribution term, $\frac{d\tau_\nu}{d\ln p}$ is the weighting function, and $B_\nu[T(p)]\left(\frac{d\tau_\nu}{d\ln p}\right)$ is the contribution function. B represents the brightness temperature calculated, T_s is the surface temperature, p is pressure, ν is frequency, and the transmission function is represented by τ_ν .

$$\theta_\nu = \epsilon_\nu B_\nu(T_s)\tau_\nu(p_s) + \int B_\nu[T(p)]\left(\frac{d\tau_\nu}{d\ln p}\right) d\ln p + \rho_\nu H_\nu \tau_\nu(p_s) + (1 - \epsilon_\nu)R_\nu \tau_\nu(p_s) \quad (5.1)$$

To obtain a model of each channel's reaction to a specific variable, that specific variable is perturbed while holding all others constant, which results in a new atmospheric model that is then passed through SARTA to calculate the new observed radiance of each channel. Calculating the difference between the new observed radiance and the old observed radiance for each channel results in knowledge about how sensitive each channel is to that specific variable. To improve our knowledge even more, instead of perturbing the entire profile of the specific variable, every layer of the profile is held constant while only one layer is perturbed. SARTA calculates the new brightness temperature for that specific perturbation which results in information about how sensitive each channel is to that specific variable at a specific altitude. Looping over all layers, we now have a model of the reaction of each channel to that specific variable at each atmospheric layer. This result is called the sensitivity matrix, or the Jacobian, which was mentioned in Section 2.1, and displayed in Equation 2.1. Equation 2.1 shows the radiance, θ_ν , is calculated for each channel's frequency, ν , and is a function of the elements in the atmosphere such as temperature, T , water vapor, Q , ozone, O_3 , and carbon dioxide, CO_2 .

The AIRS team has developed a program that computes these Jacobians, which is called the "profile alpha tool." This exact algorithm is schematically shown in Figure 5.1 which is a variation of the chart shown in the AIRS Science Team Meeting presentation [7].

5.3 Channel Selection Method

Once we have run SARTA to calculate the Jacobians for all AIRS channels, the channels are ranked in order of maximum sensitivity to the element being retrieved. We developed the following criteria for channels to be considered:

1. The maximum percent change in radiance per percent change in water vapor must be less than 10% of the maximum percent change in radiance per percent change in carbon dioxide.
2. The maximum percent change in radiance per percent change in ozone must be less than 10% of the maximum percent change in radiance per percent change in carbon dioxide.

No exact restriction was placed on sensitivity to temperature, although, channels that satisfied the above criteria seemed to have lower sensitivity to temperature as well. Also, there were two rankings—one for channels with maximum sensitivity in the stratosphere and another for channels with maximum sensitivity in the mid troposphere.

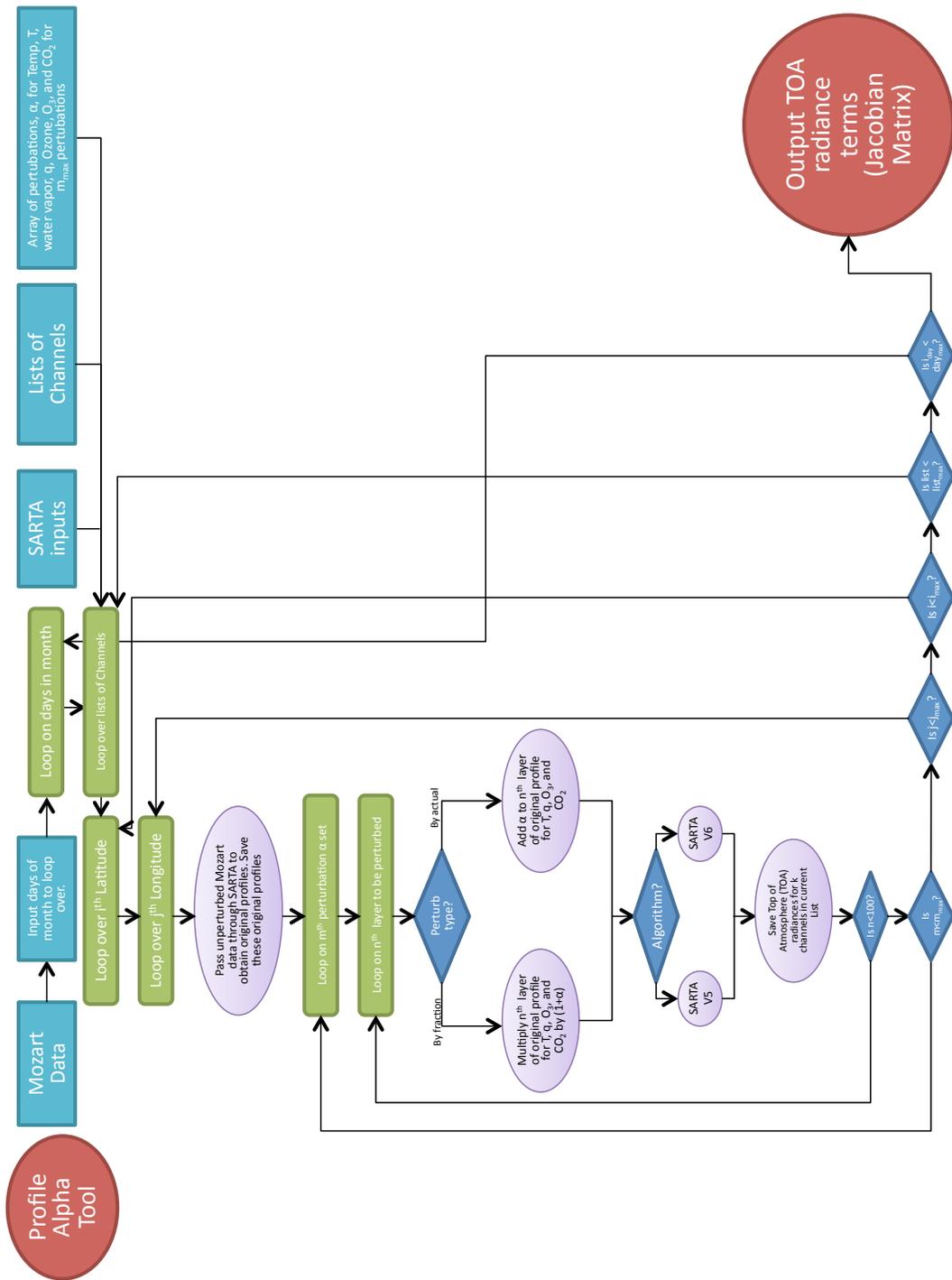


Figure 5.1: AIRS profile alpha tool algorithm. It calculates the Jacobians, Equation 2.1, for each channel, element, day, latitude, and longitude inputted to the program. The element can be perturbed by an actual amount (having correct units) or by a fractional amount. Also, there are two versions of SARTA, and the one used is an input parameter.

Once all 2378 channels were ranked, the top 40 channels for each ranking were chosen, and the averages of their Jacobians is what is plotted in Figures 3.2 and 3.1. To fully understand the shape of each channel's Jacobian, and how that corresponds to the actual concentration of CO₂, H₂O, O₃, and T, Figure 5.2 has these elements plotted together.

Earlier in this report, it was mentioned that the optimal set of channels, as well as the altitude of maximum sensitivity, was dependent on latitude and partially on longitude. This is best understood by comparing Figure 5.2 with Figures 5.3 and 5.4 which show the changes in the Jacobians and optimal channels over latitude, and comparing Figure 5.2 with Figure 5.5 to understand the longitudinal variations.

The variations of maximum sensitivity and altitude of maximum sensitivity over latitude and longitude required further investigations than the one dimensional figures. Figures 5.6 through 5.9 illustrate the latitudinal and longitudinal variations of maximum sensitivity for the mid troposphere and stratosphere. Figures 5.10 and 5.11 demonstrate the latitudinal and longitudinal effects on the altitude of maximum sensitivity.

After studying the various dependencies on the maximum sensitivity and optimal sets of channels, we developed a sort of database to use for retrieval which has the top 80 channels (40 for the stratosphere and 40 for the mid troposphere) for each latitude and longitude. The next step is to develop a larger database that takes in to account variations over seasons as well.

5.4 CO₂ Retrieval Methods

There are a few main methods in atmospheric sounding used for retrieval: the VPD method previously described, a maximum likelihood estimation method, and a minimum variance solution. All three methods are optimization methods, and the optimal method to use is dependent on the application. The latter two methods follow a Bayesian approach—approaches that identify a class of possible states that are consistent with the available information using a measurement, information about that measurement's error, a forward model describing the relation between the measurement and the unknown state, as well as any other previous information that may be provided [5]. One of the states is then identified as the actual solution with some error. The maximum likelihood solution assumes a probability density function (*pdf*), and then seeks a mean state averaged over the *pdf*. The minimum variance solution finds the state that minimizes the total error in the retrieval [5]. The following subsections describe these two methods and their derivations.

5.4.1 Maximum-Likelihood Estimation

The maximum likelihood estimation (MLE) is based on Bayes' theorem, which is written as:

$$P(\mathbf{x}|\mathbf{y}) = \frac{P(\mathbf{y}|\mathbf{x})P(\mathbf{x})}{P(\mathbf{y})} \quad (5.2)$$

where $P(\mathbf{x})$ is the prior *pdf* of state \mathbf{x} , $P(\mathbf{y})$ is the prior *pdf* of the measurement, $P(\mathbf{y}|\mathbf{x})$ is the conditional *pdf* of \mathbf{y} given \mathbf{x} , and $P(\mathbf{x}|\mathbf{y})$ is the conditional *pdf* of \mathbf{x} given \mathbf{y} [5]. We assume that each *pdf* follows a Gaussian distribution:

$$P(x) \propto e^{-\frac{(x-x_0)^2}{2\sigma^2}} \quad (5.3)$$

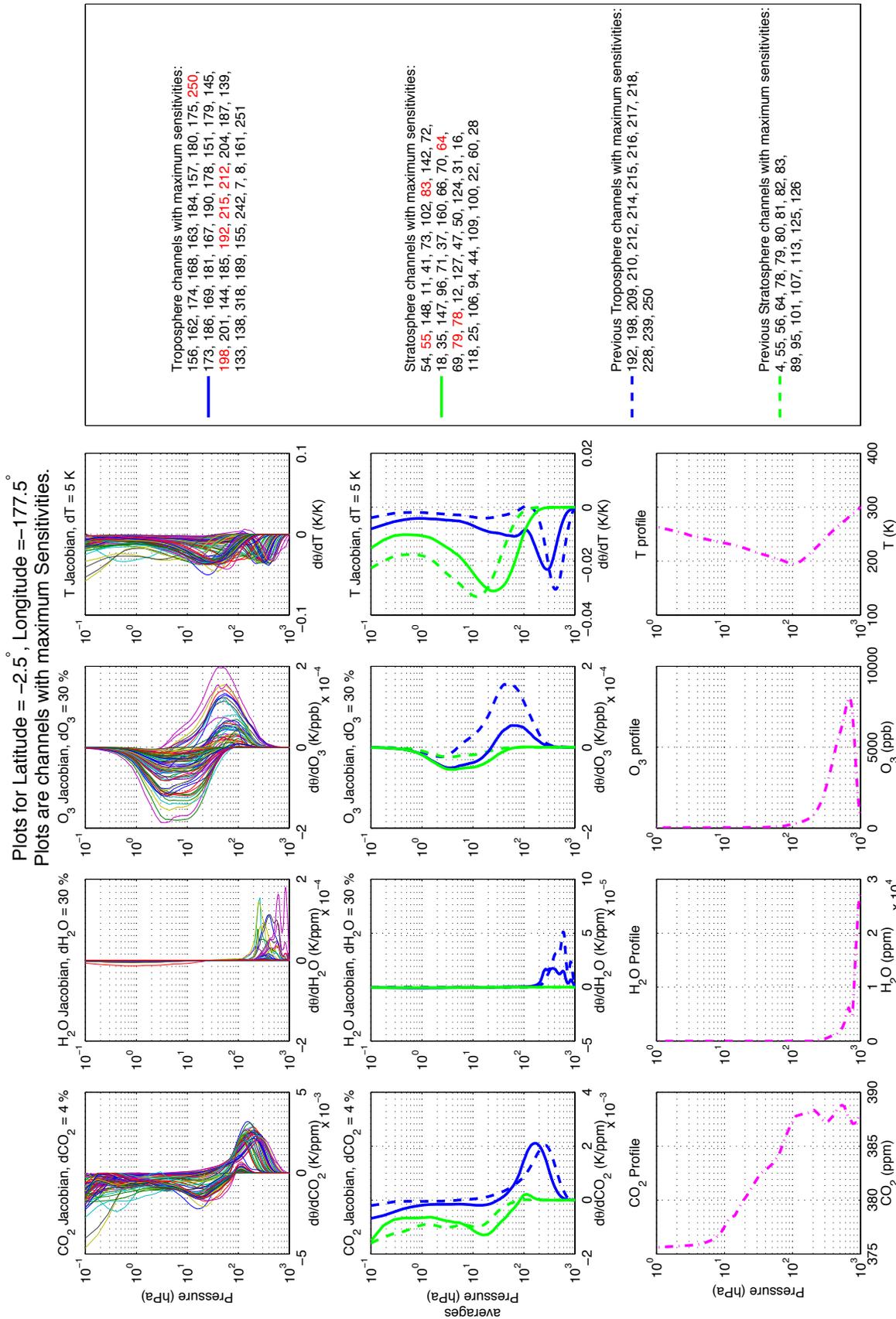


Figure 5.2: Jacobians of all new chosen channels are plotted in the top row, the averages of those channels versus the averages of the previous selected channels are plotted in the middle row. The bottom row plots are the background vertical profiles used in the calculations. The legend shows which channels' Jacobians were averaged to form the line from the middle row of plots. Dashed lines represent previous selected channels, whereas solid lines represent new selected channels. Green lines represent channels with maximum sensitivity in the stratosphere as opposed to blue lines showing channels with maximum sensitivity in the mid troposphere. Channel numbers listed in red are ones that are in the previous selected list as well. This is at a point in the Pacific Ocean over the equator.

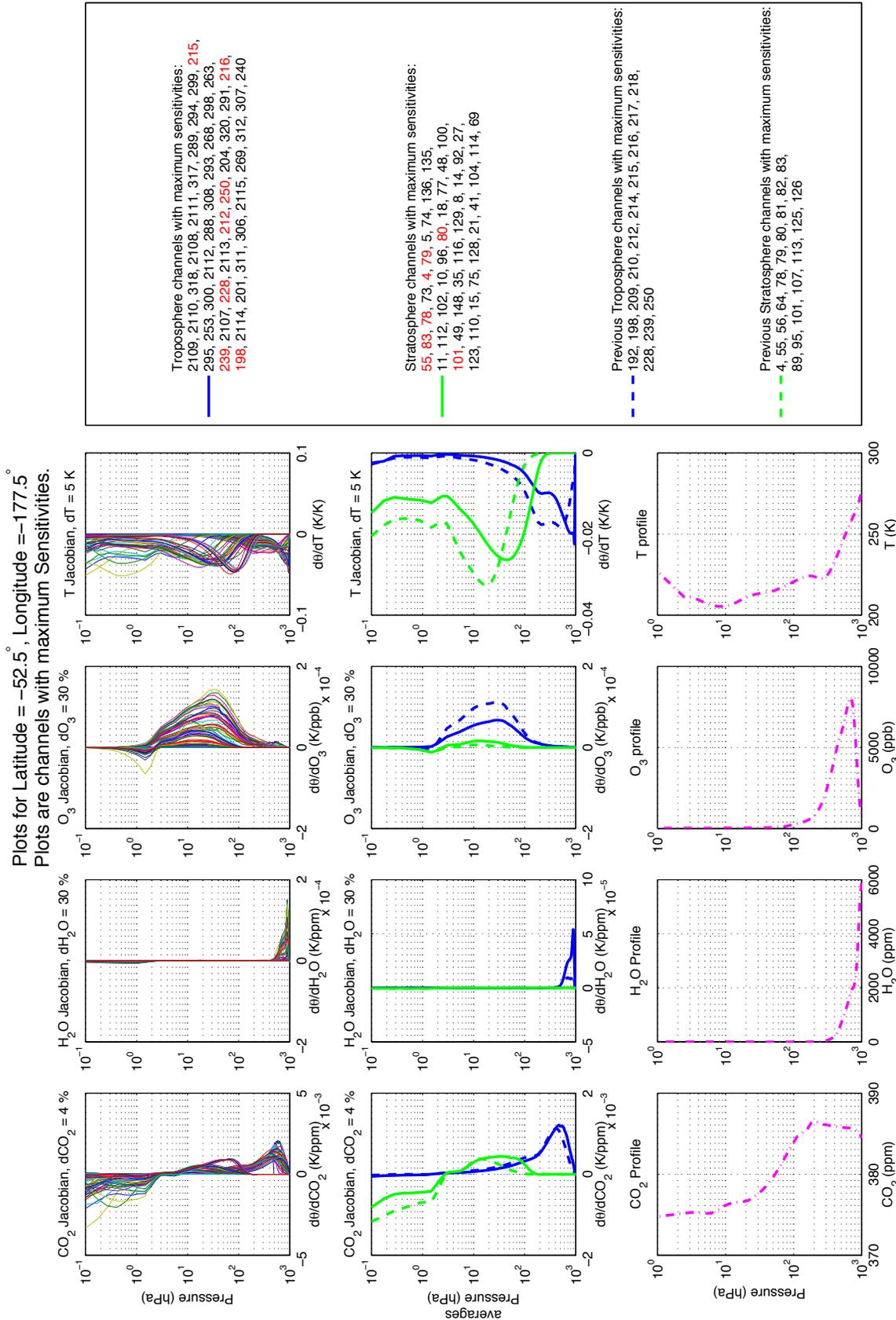


Figure 5.3: Jacobians of all new chosen channels are plotted in the top row, the averages of those channels versus the averages of the previous selected channels are plotted in the middle row. The bottom row plots are the background vertical profiles used in the calculations. The legend shows which channels' Jacobians were averaged to form the line from the middle row of plots. Dashed lines represent previous selected channels, whereas solid lines represent new selected channels. Green lines represent channels with maximum sensitivity in the stratosphere as opposed to blue lines showing channels with maximum sensitivity in the mid troposphere. Channel numbers listed in red are ones that are in the previous selected list as well. This is at a point in the Pacific Ocean in the lower hemisphere.

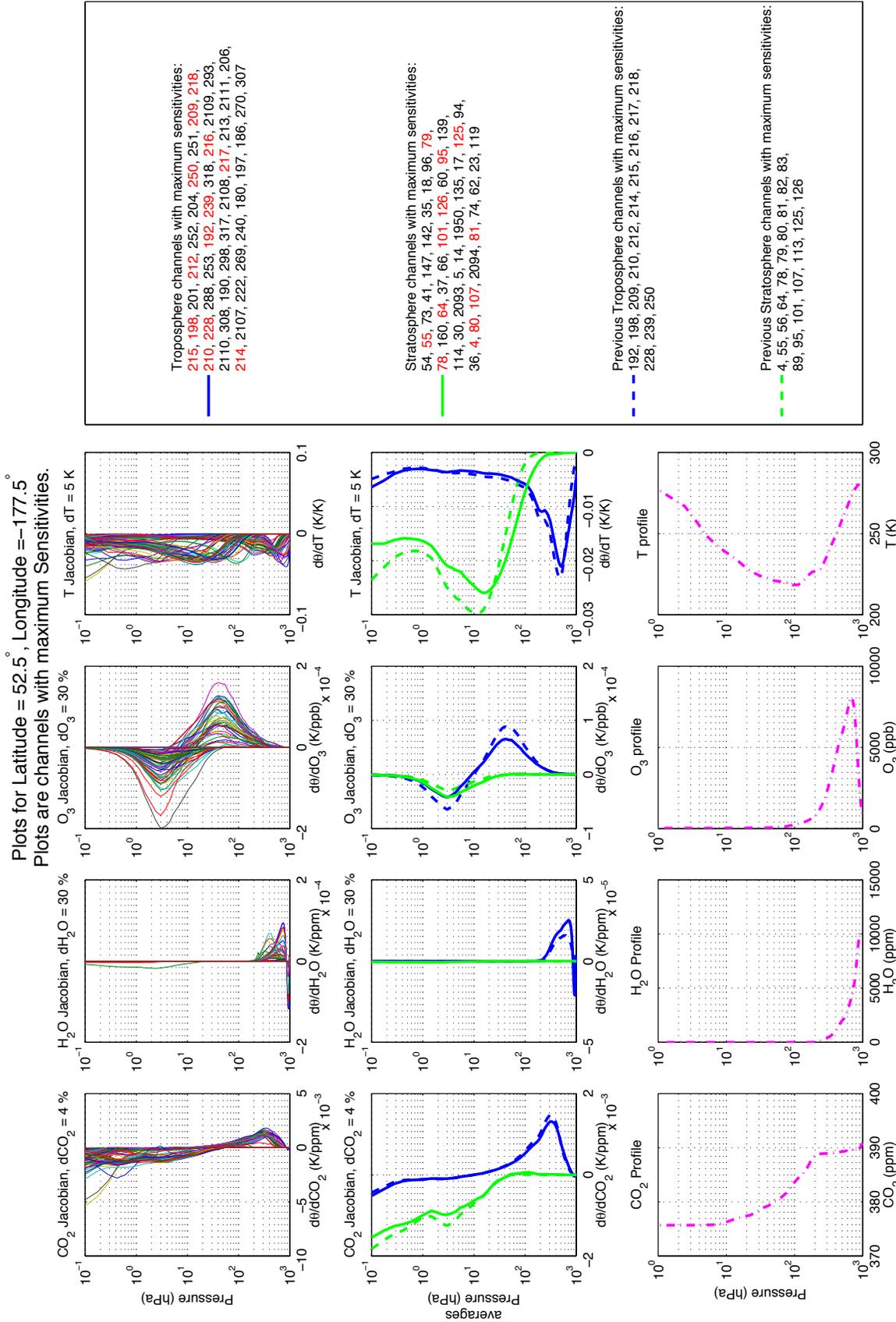


Figure 5.4: Jacobians of all new chosen channels are plotted in the top row, the averages of those channels versus the averages of the previous selected channels are plotted in the middle row. The bottom row plots are the background vertical profiles used in the calculations. The legend shows which channels' Jacobians were averaged to form the line from the middle row of plots. Dashed lines represent previous selected channels, whereas solid lines represent new selected channels. Green lines represent channels with maximum sensitivity in the stratosphere as opposed to blue lines showing channels with maximum sensitivity in the mid troposphere. Channel numbers listed in red are ones that are in the previous selected list as well. This is at a point in the Pacific Ocean in the upper hemisphere.

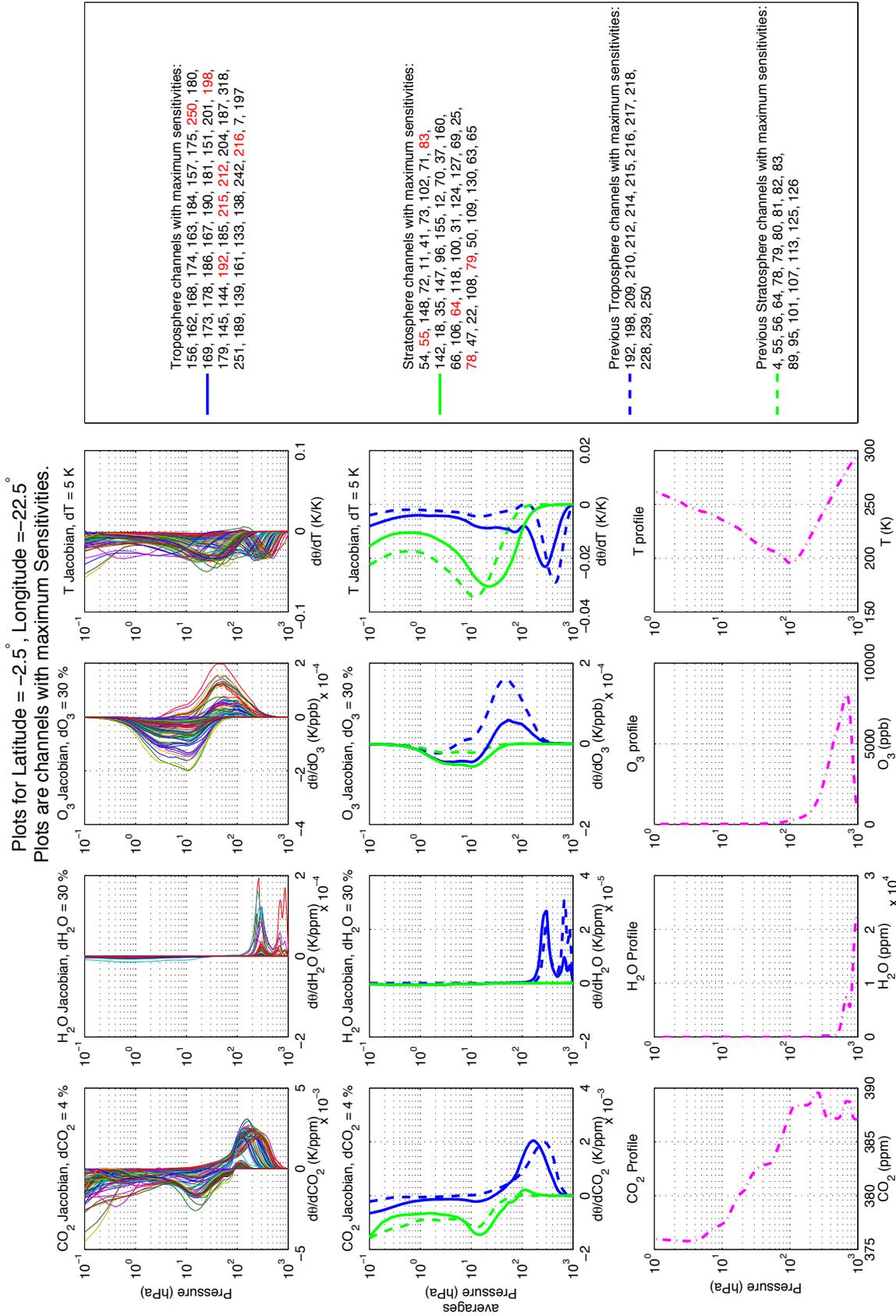


Figure 5.5: Jacobians of all new chosen channels are plotted in the top row, the averages of those channels versus the averages of the previous selected channels are plotted in the middle row. The bottom row plots are the background vertical profiles used in the calculations. The legend shows which channels' Jacobians were averaged to form the line from the middle row of plots. Dashed lines represent previous selected channels, whereas solid lines represent new selected channels. Green lines represent channels with maximum sensitivity in the stratosphere as opposed to blue lines showing channels with maximum sensitivity in the mid troposphere. Channel numbers listed in red are ones that are in the previous selected list as well. This is at a point in the Atlantic Ocean over the equator.

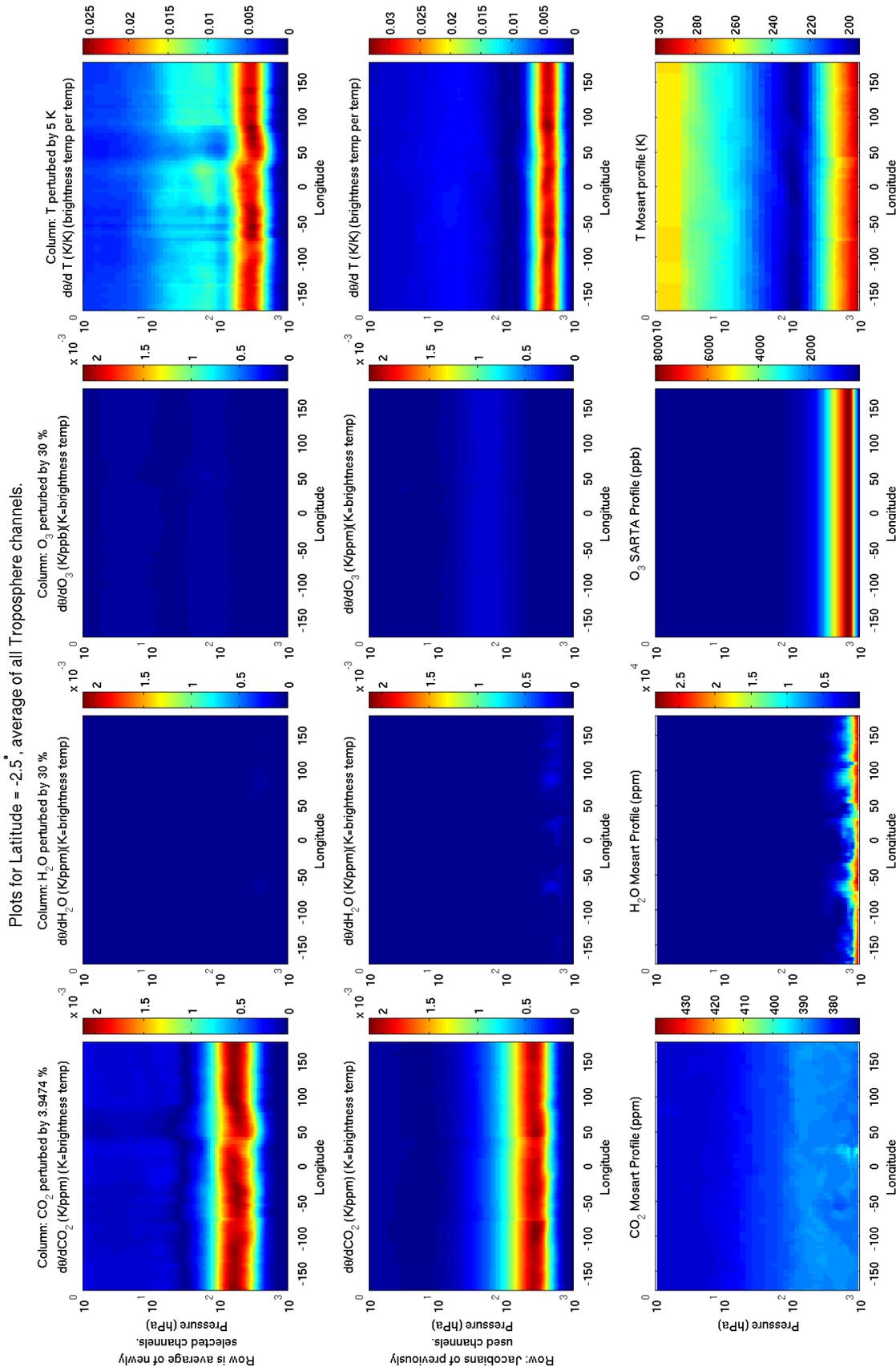


Figure 5.6: This figure shows the averages of the new selected mid troposphere channels (top row), the averages of the previous selected mid troposphere channels (middle row), and the profiles used to calculate these Jacobians (bottom row) over the equator. As is apparent by the figure, the new channel selection results in higher sensitivities to CO₂, and lower sensitivities to H₂O, O₃, and T. From this figure, it is also apparent that there are sensitivity variations over longitude, but they are minimal.

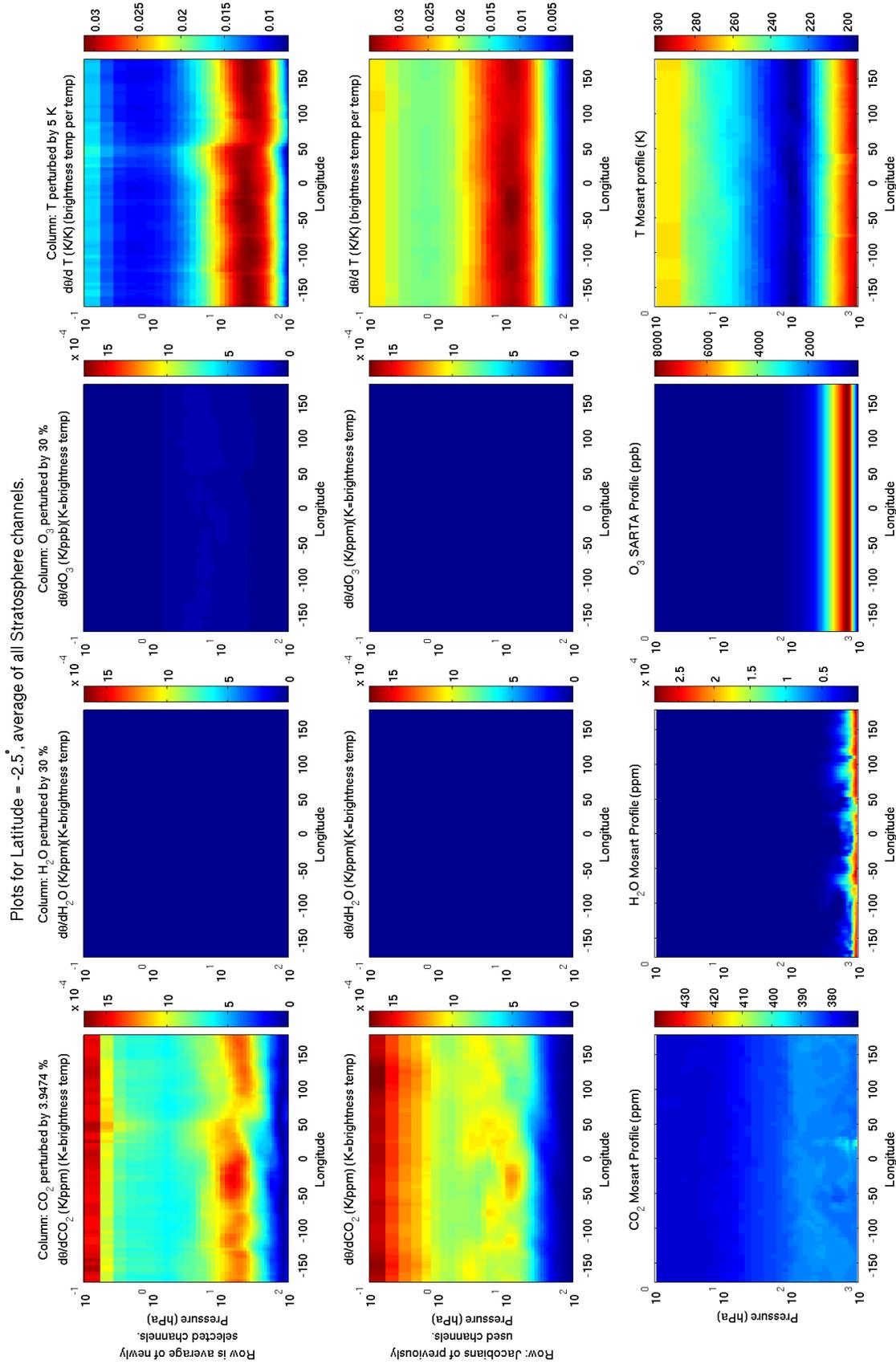


Figure 5.7: This figure shows the averages of the new selected stratosphere channels (top row), the averages of the previous selected stratosphere channels (middle row), and the profiles used to calculate these Jacobians (bottom row) over the equator. As is apparent by the figure, the new channel selection results in higher sensitivities to CO₂, and lower sensitivities to H₂O, O₃, and T. From this figure, it is also apparent that there are sensitivity variations over longitude, and these variations are larger than those in the troposphere, but they still are relatively small.

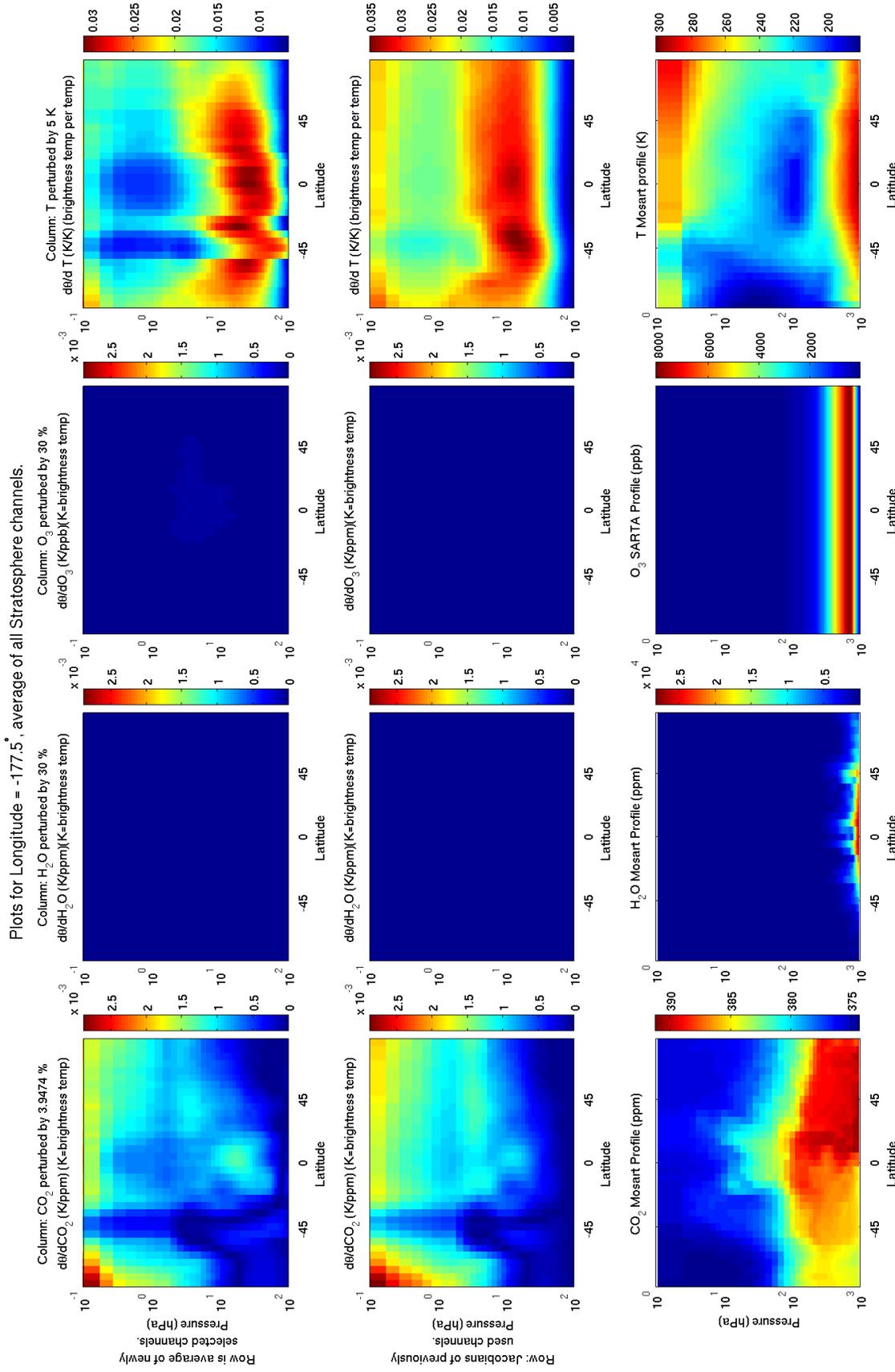


Figure 5.9: This figure shows the averages of the new selected stratosphere channels (top row), the averages of the previous selected stratosphere channels (middle row), and the profiles used to calculate these Jacobians (bottom row) over the Atlantic Ocean. As is apparent by the figure, the new channel selection results in higher sensitivities to CO₂, and lower sensitivities to H₂O, O₃, and T. This figure illustrates the substantial latitudinal variations in the Jacobians.

Plots of Average of Troposphere Channels
 For each latitude and longitude, the pressure level plotted is where the newly selected channels' Jacobians are maximum.

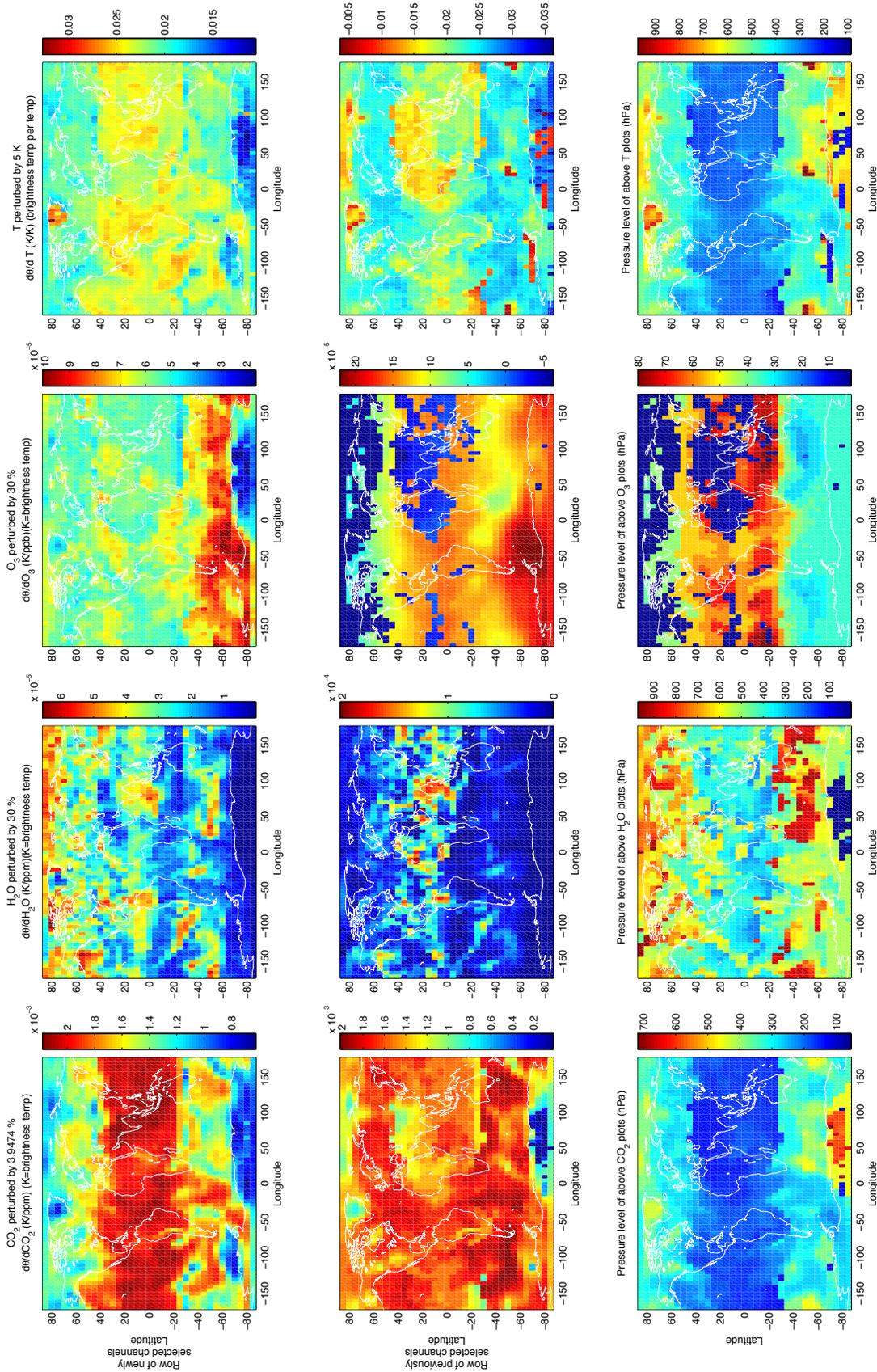


Figure 5.10: The global plots in this figure show maximum sensitivity of the new selected mid troposphere channels (top row) and the previous selected mid troposphere channels (middle row) plotted at the altitude at which the average of the new selected mid troposphere channels' Jacobians have a maximum. The bottom row plots the pressure level (or altitude) at which the two graphs above it are plotted. This is just another method of showing that the new selected channels have higher sensitivities to CO₂, and lower sensitivities to H₂O, O₃, and T, and it is interesting to see the altitude variations of these maxima.

Plots of Average of Stratosphere Channels
 For each latitude and longitude, the pressure level plotted is where the newly selected channels' Jacobians are maximum.

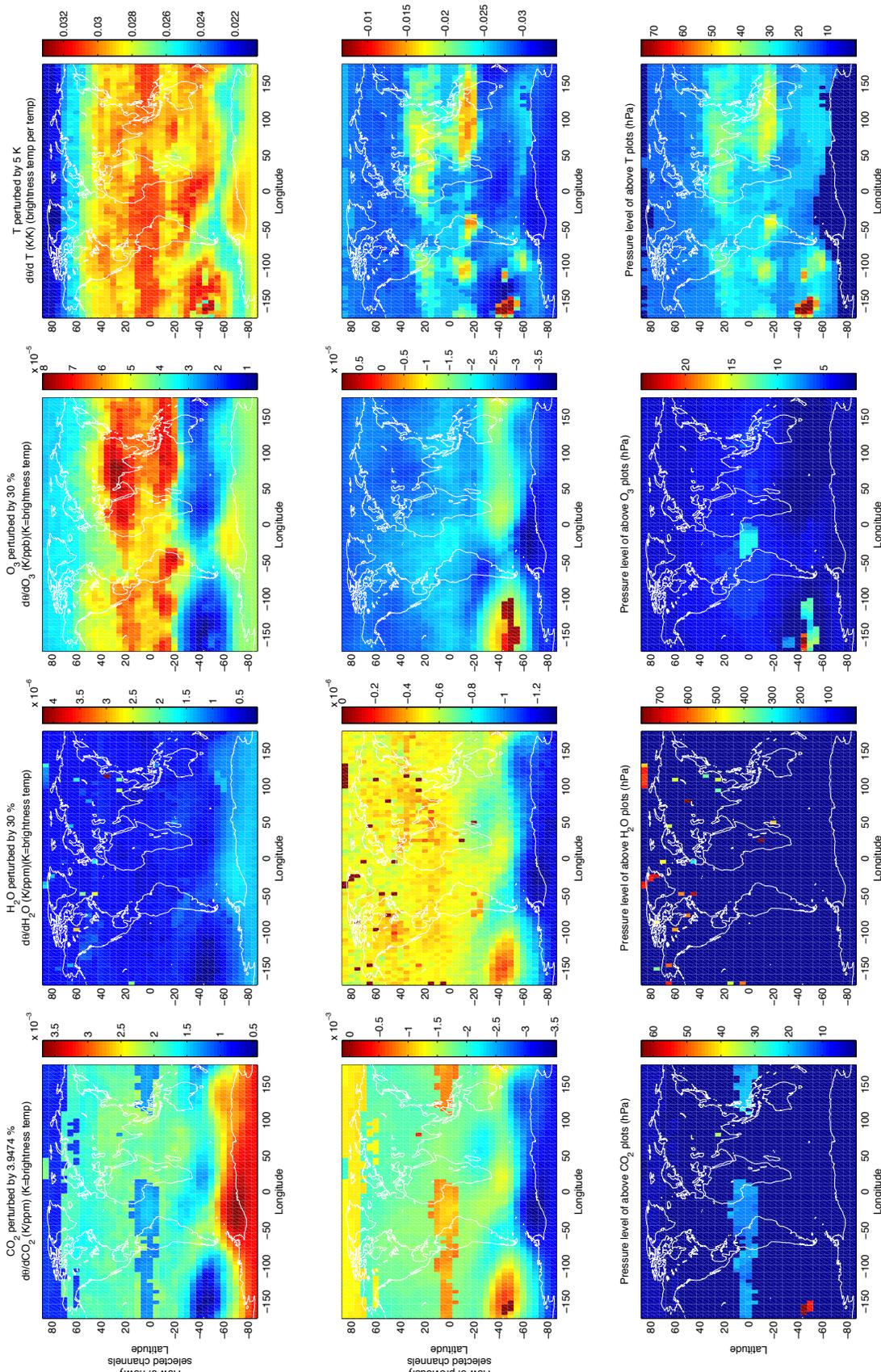


Figure 5.11: The global plots in this figure show maximum sensitivity of the new selected stratosphere channels (top row) and the previous selected stratosphere channels (middle row) plotted at the altitude at which the average of the new selected stratosphere channels' Jacobians have a maximum. The bottom row plots the pressure level (or altitude) at which the two graphs above it are plotted. This is just another method of showing that the new selected channels have higher sensitivities to CO_2 , and lower sensitivities to H_2O , O_3 , and T , and it is interesting to see the altitude variations of these maxima.

where σ is the variance and x is a vector (they are no longer in bold since we will be representing matrices with bold notation for the rest of this derivation). When matrices are involved, this becomes a bit more complicated. We no longer have just σ^2 , but we have a covariance matrix that represents σ^2 .

In an idealized situation, each data point is independent of every other data point resulting in a diagonal covariance matrix, \mathbf{B} , with components $\{\sigma_0^2, \sigma_1^2, \dots, \sigma_k^2\}$. Our atmosphere is much more complicated than that. Each data point is dependent on measurements and other information below and above that point, resulting in an added correlation term in the covariance matrix. The covariance matrix can then be written as, $\mathbf{B} = \mathbf{\Sigma}^T \mathbf{C} \mathbf{\Sigma}$, where $\mathbf{\Sigma}$ is a diagonal matrix with elements $\{\sigma_0, \sigma_1, \dots, \sigma_k\}$ and \mathbf{C} represents the correlation between data points. More about the meaning of this term is explained in Section 5.7.3.

x in Equation 5.3 is a column vector with 100 entries representing the concentration of molecules in the atmosphere, where each of the 100 entries is the concentration at a certain altitude. The number of 100 was chosen from precedent. Through the rest of this derivation, x will be written as x_{var} where “var” represents temperature, T , water vapor, q , ozone, O_3 , carbon dioxide, CO_2 , etc.

Transforming Equation 5.3 into matrix form for a CO_2 calculation, we obtain Equation 5.4 where \mathbf{B}_{CO_2} is the background-error covariance matrix, or the aleatoric uncertainty.

$$P(x_T^b, x_q^b, x_{O_3}^b, \dots, x_{CO_2}) \propto e^{-\frac{1}{2}(x_{CO_2} - x_{CO_2}^b)^T \mathbf{B}_{CO_2}^{-1} (x_{CO_2} - x_{CO_2}^b)} \quad (5.4)$$

Equation 5.4 uses the superscript b instead of the subscript 0 used in Equation 5.3 to represent that it is the background term. Background data are prior data gathered before any model is implemented or data are analyzed. The background data are data obtained from weather forecasts, other experiments, observations, etc. They are data gathered from some source other than AIRS that we can use as a reference when analyzing the AIRS data.

Equation 5.4 still is not the entire picture. It is only the $P(\mathbf{x})$ term in Equation 5.2. To get a better estimate of the error in the model, $P(\mathbf{y}|\mathbf{x})$ from Equation 5.2 must be found. In this application, it is a comparison between the radiance calculated from the model and the observed/measured radiance (which will be represented with the superscript “obs”). If the model and the assumed values of the various quantities are perfect, the radiance observed is equal to the radiance calculated from the model. In mathematical terms, if $h(x)$ is the model, then $y = h(x)$. Using a Taylor expansion, $y = h(x) + h'(x)(x - x_b)$, or $\Delta y = h'(x)\Delta x$. Relating this back to the current application, observed radiance can be represented as $\Delta\theta^{\text{obs}}$, and correspond to Δy . The derivative of the model, $h'(x)$, corresponds to the derivative of our SARTA model, \mathbf{H} , the Jacobian matrix. Letting $\Delta\theta^{\text{obs}} = \theta^{\text{obs}} - \theta(x_T, x_q, x_{O_3}, \dots, x_{CO_2})$ and $\Delta x_{CO_2} = x_{CO_2} - x_{CO_2}^b$ Equation 5.5 is obtained.

$$\Delta\theta^{\text{obs}} = \mathbf{H}_{CO_2} \Delta x_{CO_2} \quad (5.5)$$

Equation 5.5 is a linear approximation, and to simplify the derivations presented later, the Δ can be dropped off both sides of the equation to yield Equation 5.6.

$$\theta^{\text{obs}} = \theta(x_T, x_q, x_{O_3}, \dots, x_{CO_2}) = \mathbf{H}_{CO_2} x_{CO_2} \quad (5.6)$$

Under the assumption that the measurement errors are Gaussian, $P(\mathbf{y}|\mathbf{x})$ representing the actual *pdf* of the model

results in Equation 5.4.

$$P(x_T^b, x_q^b, x_{O_3}^b, \dots, x_{CO_2}) \propto e^{-\frac{1}{2}(\theta^{\text{obs}} - \mathbf{H}_{CO_2} x_{CO_2})^T \mathbf{R}^{-1} (\theta^{\text{obs}} - \mathbf{H}_{CO_2} x_{CO_2})} \quad (5.7)$$

\mathbf{R} in the above equation is the observational-error covariance matrix, which is a measured quantity gathered from calibrating the instrument—it is the error associated with each channel. θ^{obs} is a vector representing observed radiances, which are calculated for each AIRS channel used. In equation form, $\theta^{\text{obs}} = \{\theta_{\nu_1}^{\text{obs}}, \theta_{\nu_2}^{\text{obs}}, \dots, \theta_{\nu_m}^{\text{obs}}\}$ where ν_m represents the wavelength detected by channel m .

Since Equation 5.4 is the $P(\mathbf{x})$ term from Equation 5.2, and Equation 5.7 is the $P(\mathbf{y}|\mathbf{x})$ term from Equation 5.2, they can be multiplied together to obtain an equation for the *pdf* of the retrieval.

$$P(x_T^b, x_q^b, x_{O_3}^b, \dots, x_{CO_2}) \propto e^{-\frac{1}{2}[(\theta^{\text{obs}} - \mathbf{H}_{CO_2} x_{CO_2})^T \mathbf{R}^{-1} (\theta^{\text{obs}} - \mathbf{H}_{CO_2} x_{CO_2}) + (x_{CO_2} - x_{CO_2}^b)^T \mathbf{B}_{CO_2}^{-1} (x_{CO_2} - x_{CO_2}^b)]} \quad (5.8)$$

Equation 5.8 is a Gaussian curve ranging over all possible states. The most probable state is the mean value of x , which is also the x value that minimizes the absolute value of the term in the exponential. The term in the exponential is called the “cost function,” and is represented by $J(x_{CO_2})$. Equation 5.9 describes the cost function, which was described briefly earlier in this report (Equation 2.2).

$$J(x_{CO_2}) = \frac{1}{2}(x_{CO_2} - x_{CO_2}^b)^T \mathbf{B}_{CO_2}^{-1} (x_{CO_2} - x_{CO_2}^b) + \frac{1}{2}(\theta^{\text{obs}} - \mathbf{H}_{CO_2} x_{CO_2})^T \mathbf{R}^{-1} (\theta^{\text{obs}} - \mathbf{H}_{CO_2} x_{CO_2}) \quad (5.9)$$

To minimize the cost function (Equation 5.9), its derivative is taken with respect to x_{CO_2} and is set to zero.

$$\frac{\partial J}{\partial x_{CO_2}} = \mathbf{B}_{CO_2}^{-1} (x_{CO_2} - x_{CO_2}^b) + \mathbf{H}_{CO_2}^T \mathbf{R}^{-1} (\mathbf{H}_{CO_2} x_{CO_2} - \theta^{\text{obs}}) = 0 \quad (5.10)$$

The above equation can then be solved for x_{CO_2} .

$$x_{CO_2} (\mathbf{B}_{CO_2}^{-1} + \mathbf{H}_{CO_2}^T \mathbf{R}^{-1} \mathbf{H}_{CO_2}) - \mathbf{B}_{CO_2}^{-1} x_{CO_2}^b - \mathbf{H}_{CO_2}^T \mathbf{R}^{-1} \theta^{\text{obs}} = 0 \quad (5.11)$$

$$x_{CO_2} = (\mathbf{B}_{CO_2}^{-1} + \mathbf{H}_{CO_2}^T \mathbf{R}^{-1} \mathbf{H}_{CO_2})^{-1} (\mathbf{B}_{CO_2}^{-1} x_{CO_2}^b + \mathbf{H}_{CO_2}^T \mathbf{R}^{-1} \theta^{\text{obs}}) \quad (5.12)$$

The Sherman-Morrison-Woodbury formula can be introduced to simplify Equation 5.12.

$$(\mathbf{A} + \mathbf{X} \mathbf{B} \mathbf{X}^T)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{X} (\mathbf{B}^{-1} + \mathbf{X}^T \mathbf{A}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{A}^{-1} \quad (5.13)$$

Using the Sherman-Morrison-Woodbury formula (Equation 5.13) to simplify the term $(\mathbf{B}_{CO_2}^{-1} + \mathbf{H}_{CO_2}^T \mathbf{R}^{-1} \mathbf{H}_{CO_2})^{-1}$, the following equation is obtained:

$$(\mathbf{B}_{CO_2}^{-1} + \mathbf{H}_{CO_2}^T \mathbf{R}^{-1} \mathbf{H}_{CO_2})^{-1} = \mathbf{B}_{CO_2} - \mathbf{B}_{CO_2} \mathbf{H}_{CO_2}^T (\mathbf{R} + \mathbf{H}_{CO_2} \mathbf{B}_{CO_2} \mathbf{H}_{CO_2}^T)^{-1} \mathbf{H}_{CO_2} \mathbf{B}_{CO_2}. \quad (5.14)$$

Multiplying the above result by the rest of Equation 5.12, $(\mathbf{B}_{CO_2}^{-1}x_{CO_2}^b + \mathbf{H}_{CO_2}^T\mathbf{R}^{-1}\theta^{obs})$ yields Equation 5.15.

$$\begin{aligned} x_{CO_2} &= \mathbf{B}_{CO_2}\mathbf{B}_{CO_2}^{-1}x_{CO_2}^b - \mathbf{B}_{CO_2}\mathbf{H}_{CO_2}^T(\mathbf{R} + \mathbf{H}_{CO_2}\mathbf{B}_{CO_2}\mathbf{H}_{CO_2}^T)^{-1}\mathbf{H}_{CO_2}\mathbf{B}_{CO_2}\mathbf{B}_{CO_2}^{-1}x_{CO_2}^b \\ &\quad + \mathbf{B}_{CO_2}\mathbf{H}_{CO_2}^T\mathbf{R}^{-1}\theta^{obs} - \mathbf{B}_{CO_2}\mathbf{H}_{CO_2}^T(\mathbf{R} + \mathbf{H}_{CO_2}\mathbf{B}_{CO_2}\mathbf{H}_{CO_2}^T)^{-1}\mathbf{H}_{CO_2}\mathbf{B}_{CO_2}\mathbf{H}_{CO_2}^T\mathbf{R}^{-1}\theta^{obs} \end{aligned} \quad (5.15)$$

The first term in Equation 5.15 can be simplified by implementing $\mathbf{B}_{CO_2}\mathbf{B}_{CO_2}^{-1} = \mathbf{I}$ where \mathbf{I} is the identity matrix. The second half of Equation 5.15 needs linear algebra for simplification.

$$\begin{aligned} &\mathbf{B}_{CO_2}\mathbf{H}_{CO_2}^T(\mathbf{R}^{-1}\theta^{obs} - (\mathbf{R} + \mathbf{H}_{CO_2}\mathbf{B}_{CO_2}\mathbf{H}_{CO_2}^T)^{-1}\mathbf{H}_{CO_2}\mathbf{B}_{CO_2}\mathbf{H}_{CO_2}^T\mathbf{R}^{-1}\theta^{obs}) \\ &= \mathbf{B}_{CO_2}\mathbf{H}_{CO_2}^T(\mathbf{R} + \mathbf{H}_{CO_2}\mathbf{B}_{CO_2}\mathbf{H}_{CO_2}^T)^{-1}(\mathbf{R}\mathbf{R}^{-1}\theta^{obs} + \mathbf{H}_{CO_2}\mathbf{B}_{CO_2}\mathbf{H}_{CO_2}^T\mathbf{R}^{-1}\theta^{obs} - \mathbf{H}_{CO_2}\mathbf{B}_{CO_2}\mathbf{H}_{CO_2}^T\mathbf{R}^{-1}\theta^{obs}) \\ &= \mathbf{B}_{CO_2}\mathbf{H}_{CO_2}^T(\mathbf{R} + \mathbf{H}_{CO_2}\mathbf{B}_{CO_2}\mathbf{H}_{CO_2}^T)^{-1}\theta^{obs} \end{aligned} \quad (5.16)$$

Using the result from Equation 5.16 with Equation 5.15 and combining like terms, a more simplified equation for x_{CO_2} is obtained.

$$x_{CO_2} = x_{CO_2}^b + \mathbf{B}_{CO_2}\mathbf{H}_{CO_2}^T(\mathbf{R} + \mathbf{H}_{CO_2}\mathbf{B}_{CO_2}\mathbf{H}_{CO_2}^T)^{-1}(\theta^{obs} - \mathbf{H}_{CO_2}x_{CO_2}^b) \quad (5.17)$$

In order to prevent confusion throughout the rest of the derivation, a couple labeling schemes will be used. x_{CO_2} in the above derivation is the retrieved value, which will be distinguished by a superscript “re.” Also, any value considered to be the “true” value will be denoted by a superscript “t.” The error in retrieval is then represented as the difference between the values used or derived in/from the model and the true values.

$$x_{CO_2}^{re} - x_{CO_2}^t = \epsilon^{re}, \quad x_{CO_2}^b - x_{CO_2}^t = \epsilon^b, \quad \theta^{obs} - \theta^t = \epsilon^{obs} \quad (5.18)$$

The “true” values are then subtracted from both sides of Equation 5.17.

$$\begin{aligned} x_{CO_2} - x_{CO_2}^t &= x_{CO_2}^b - x_{CO_2}^t + \mathbf{B}_{CO_2}\mathbf{H}_{CO_2}^T(\mathbf{R} + \mathbf{H}_{CO_2}\mathbf{B}_{CO_2}\mathbf{H}_{CO_2}^T)^{-1}(\theta^{obs} - \theta^t + \theta^t - \mathbf{H}_{CO_2}x_{CO_2}^b) \\ \epsilon^{re} &= \epsilon^b + \mathbf{B}_{CO_2}\mathbf{H}_{CO_2}^T(\mathbf{R} + \mathbf{H}_{CO_2}\mathbf{B}_{CO_2}\mathbf{H}_{CO_2}^T)^{-1}(\epsilon^{obs} + \theta^t - \mathbf{H}_{CO_2}x_{CO_2}^b) \end{aligned} \quad (5.19)$$

To simplify the above equation further, the following assumption is made: if the model is perfect, from Equation 5.6 it follows that $\theta^t = \mathbf{H}_{CO_2}x_{CO_2}^t$. Implementing this assumption into the above equation, an equation for the retrieval error is derived.

$$\epsilon^{re} = \epsilon^b + \mathbf{B}_{CO_2}\mathbf{H}_{CO_2}^T(\mathbf{R} + \mathbf{H}_{CO_2}\mathbf{B}_{CO_2}\mathbf{H}_{CO_2}^T)^{-1}(\epsilon^{obs} - \mathbf{H}_{CO_2}\epsilon^b) \quad (5.20)$$

Equation 5.20 gives the retrieval root-mean-square error. The last piece of this error derivation is to find the retrieved CO₂ error covariance matrix, \mathbf{P}_{CO_2} . To find this matrix, the inner product of the retrieval error and the transpose of the retrieval error is taken, which is true for any of the error covariance matrices.

$$\mathbf{B}_{CO_2} = \langle \epsilon^b(\epsilon^b)^T \rangle, \quad \mathbf{R} = \langle \epsilon^{obs}(\epsilon^{obs})^T \rangle, \quad \mathbf{P}_{CO_2} = \langle \epsilon^{re}(\epsilon^{re})^T \rangle \quad (5.21)$$

To calculate \mathbf{P}_{CO_2} , the transpose of Equation 5.20 is needed.

$$(\epsilon^{re})^T = (\epsilon^b)^T + ((\epsilon^{obs})^T - (\epsilon^b)^T \mathbf{H}_{CO_2}^T) [(\mathbf{R}^T + \mathbf{H}_{CO_2} \mathbf{B}_{CO_2}^T \mathbf{H}_{CO_2}^T)^{-1} \mathbf{H}_{CO_2} \mathbf{B}_{CO_2}^T] \quad (5.22)$$

The inner product of Equations 5.20 and 5.22, using the formulas from Equation 5.21, yields an equation for \mathbf{P}_{CO_2} .

$$\begin{aligned} \mathbf{P}_{CO_2} &= \langle \epsilon^{re} (\epsilon^{re})^T \rangle = \epsilon^b \epsilon^{bT} + \epsilon^b ((\epsilon^{obs})^T - \epsilon^{bT} \mathbf{H}_{CO_2}^T) (\mathbf{R}^T + \mathbf{H}_{CO_2} \mathbf{B}_{CO_2}^T \mathbf{H}_{CO_2}^T)^{-1} \mathbf{H}_{CO_2} \mathbf{B}_{CO_2}^T \\ &\quad + \mathbf{B}_{CO_2} \mathbf{H}_{CO_2}^T (\mathbf{R} + \mathbf{H}_{CO_2} \mathbf{B}_{CO_2} \mathbf{H}_{CO_2}^T)^{-1} (\epsilon^{obs} - \mathbf{H}_{CO_2} \epsilon^b) \epsilon^{bT} \\ &\quad + \mathbf{B}_{CO_2} \mathbf{H}_{CO_2}^T (\mathbf{R} + \mathbf{H}_{CO_2} \mathbf{B}_{CO_2} \mathbf{H}_{CO_2}^T)^{-1} (\epsilon^{obs} - \mathbf{H}_{CO_2} \epsilon^b) (\epsilon^{bT} - \epsilon^{bT} \mathbf{H}_{CO_2}^T) (\mathbf{R}^T + \mathbf{H}_{CO_2} \mathbf{B}_{CO_2}^T \mathbf{H}_{CO_2}^T)^{-1} \mathbf{H}_{CO_2} \mathbf{B}_{CO_2}^T \\ &= \mathbf{B}_{CO_2} - \mathbf{B}_{CO_2} \mathbf{H}_{CO_2}^T (\mathbf{R}^T + \mathbf{H}_{CO_2} \mathbf{B}_{CO_2}^T \mathbf{H}_{CO_2}^T)^{-1} \mathbf{H}_{CO_2} \mathbf{B}_{CO_2}^T \\ &\quad - \mathbf{B}_{CO_2} \mathbf{H}_{CO_2}^T (\mathbf{R} + \mathbf{H}_{CO_2} \mathbf{B}_{CO_2} \mathbf{H}_{CO_2}^T)^{-1} \mathbf{H}_{CO_2} \mathbf{B}_{CO_2} \\ &\quad + \mathbf{B}_{CO_2} \mathbf{H}_{CO_2}^T (\mathbf{R} + \mathbf{H}_{CO_2} \mathbf{B}_{CO_2} \mathbf{H}_{CO_2}^T)^{-1} (\mathbf{R} + \mathbf{H}_{CO_2} \mathbf{B}_{CO_2} \mathbf{H}_{CO_2}^T) (\mathbf{R}^T + \mathbf{H}_{CO_2} \mathbf{B}_{CO_2}^T \mathbf{H}_{CO_2}^T)^{-1} \mathbf{H}_{CO_2} \mathbf{B}_{CO_2}^T \\ &= \mathbf{B}_{CO_2} - \mathbf{B}_{CO_2} \mathbf{H}_{CO_2}^T [(\mathbf{R}^T + \mathbf{H}_{CO_2} \mathbf{B}_{CO_2}^T \mathbf{H}_{CO_2}^T)^{-1} + (\mathbf{R} + \mathbf{H}_{CO_2} \mathbf{B}_{CO_2} \mathbf{H}_{CO_2}^T)^{-1} \\ &\quad - (\mathbf{R}^T + \mathbf{H}_{CO_2} \mathbf{B}_{CO_2}^T \mathbf{H}_{CO_2}^T)^{-1}] \mathbf{H}_{CO_2} \mathbf{B}_{CO_2}^T \end{aligned} \quad (5.23)$$

The terms corresponding to $\langle \epsilon^b (\epsilon^{obs})^T \rangle$ and $\langle \epsilon^{obs} (\epsilon^b)^T \rangle$ are set to zero in the above because $\langle J(x) \rangle = \frac{n}{2}$, which was explained briefly in Section 2.2 and will be described in more detail in Section 5.6. For the expectation value of the cost function to be independent of changing \mathbf{B}_{CO_2} and \mathbf{R} , it must be true that if one error is larger, the other error is smaller. Under this assumption, if ϵ^b is large, ϵ^{obs} must be small, which results in $\langle \epsilon^b (\epsilon^{obs})^T \rangle = 0$ and $\langle \epsilon^{obs} (\epsilon^b)^T \rangle = 0$.

Continuing on with the derivation of \mathbf{P}_{CO_2} , canceling out like terms and then using the Sherman-Morrison-Woodbury formula (Equation 5.13) with the result of Equation 5.23, results in the final answer that represents the retrieved CO₂ error covariance—Equation 5.24.

$$\begin{aligned} \mathbf{P}_{CO_2} &= \mathbf{B}_{CO_2} - \mathbf{B}_{CO_2} \mathbf{H}_{CO_2}^T (\mathbf{R} + \mathbf{H}_{CO_2} \mathbf{B}_{CO_2} \mathbf{H}_{CO_2}^T)^{-1} \mathbf{H}_{CO_2} \mathbf{B}_{CO_2} \\ &= (\mathbf{B}_{CO_2}^{-1} + \mathbf{H}_{CO_2}^T \mathbf{R}^{-1} \mathbf{H}_{CO_2})^{-1} \end{aligned} \quad (5.24)$$

Overall, the goal is to minimize \mathbf{P}_{CO_2} , or maximize $(\mathbf{B}_{CO_2}^{-1} + \mathbf{H}_{CO_2}^T \mathbf{R}^{-1} \mathbf{H}_{CO_2})$, which is done by minimizing \mathbf{B}_{CO_2} and/or maximizing $\mathbf{H}_{CO_2}^T \mathbf{R}^{-1} \mathbf{H}_{CO_2}$. Minimizing the background error covariance term means using more accurate data as the prior, which is not a result of the model or the instrument. What *is* a result of the model and instrument is the second term $\mathbf{H}_{CO_2}^T \mathbf{R}^{-1} \mathbf{H}_{CO_2}$. To maximize that term, the channels used need to have large sensitivities (maximize the Jacobians in the \mathbf{H}_{CO_2} term) and/or a large number of channels should be used. A large number of channels increases that term because of the following— $(\mathbf{H}_{CO_2}^T \mathbf{R}^{-1} \mathbf{H}_{CO_2})_{j,j} = \mathbf{H}_{j,i} \mathbf{R}_{i,i}^{-1} \mathbf{H}_{i,j}$, where $j = 1, 2, \dots, \text{nlevs}$, and $i = 1, 2, \dots, \text{nchans}$, “nlevs” is the number of atmospheric levels used in the calculation, and “nchans” is the number of channels used. The matrix multiplication shows that the term sums over i , meaning a larger “nchans” results in a larger sum, yielding a smaller \mathbf{P}_{CO_2} quantity. To summarize—choosing a larger number of channels can minimize the retrieval error, but those channels must have relatively large sensitivities for $\mathbf{H}_{CO_2}^T \mathbf{R}^{-1} \mathbf{H}_{CO_2}$ to be large. This also introduces more noise since each channel has a particular error associated with it, which brings the conclusion that the optimal solution is a balance between using more channels and adding

noise versus using fewer channels and adding more uncertainty from lack of data. This conclusion is the foundation for the emphasis on sensitivity analysis and uncertainty quantification of channels with the new channel selection process.

5.4.2 Minimum Error Variance

In the previous subsection, we assessed errors using a maximum-likelihood estimation, but there is another method to estimating errors which does not rely on a Gaussian *pdf*—the minimum error variance (MEV) solution. This method uses the background assumption, the observed radiance, and the model to find the minimum retrieval error variance. It begins with an assumed equation for the retrieval of x^{re} being a function of the background, and the observation multiplied by some value \mathbf{K} .

$$x^{\text{re}} = x^b + \mathbf{K}(\theta^{\text{obs}} - \mathbf{H}x^b) \quad (5.25)$$

To find the retrieval error covariance associated with this method, relatively similar steps are taken as the previous section.

$$x^{\text{re}} - x^t = x^b - x^t + \mathbf{K}(\theta^{\text{obs}} - \theta^t + \theta^t - \mathbf{H}x^b) \quad (5.26)$$

Using the relations in Equation 5.18 as well as the assumption mentioned that if the model is perfect, $\theta^t = \mathbf{H}_{CO_2}x^t_{CO_2}$, an equation for the retrieval error is obtained.

$$\epsilon^{\text{re}} = \epsilon^b + \mathbf{K}(\epsilon^{\text{obs}} - \mathbf{H}\epsilon^b) \quad (5.27)$$

$$(\epsilon^{\text{re}})^T = \epsilon^{bT} + ((\epsilon^{\text{obs}})^T - \mathbf{H}^T\epsilon^{bT})\mathbf{K}^T \quad (5.28)$$

$$\begin{aligned} \mathbf{P} = \langle \epsilon^{\text{re}}(\epsilon^{\text{re}})^T \rangle &= \epsilon^b\epsilon^{bT} + \epsilon^b((\epsilon^{\text{obs}})^T - \epsilon^{bT}\mathbf{H}^T)\mathbf{K}^T + \mathbf{K}(\epsilon^{\text{obs}} - \mathbf{H}\epsilon^b)\epsilon^{bT} + \mathbf{K}(\epsilon^{\text{obs}} - \mathbf{H}\epsilon^b)((\epsilon^{\text{obs}})^T - \epsilon^{bT}\mathbf{H}^T)\mathbf{K}^T \\ &= \mathbf{B} - \mathbf{B}\mathbf{H}^T\mathbf{K}^T - \mathbf{K}\mathbf{H}\mathbf{B} + \mathbf{K}\mathbf{R}\mathbf{K}^T + \mathbf{K}\mathbf{H}\mathbf{B}\mathbf{H}^T\mathbf{K}^T \\ &= \mathbf{B} - \mathbf{B}\mathbf{H}^T\mathbf{K}^T - \mathbf{K}\mathbf{H}\mathbf{B} + \mathbf{K}(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)\mathbf{K}^T \end{aligned} \quad (5.29)$$

The final term in Equation 5.29 is not the full solution because the meaning of \mathbf{K} has yet to be described. To find \mathbf{K} , the trace of \mathbf{P} is taken and minimized with respect to \mathbf{K} . The trace is taken because this solution looks for the error variance—which are the diagonal terms of \mathbf{P} .

$$\begin{aligned} \frac{\partial \text{Tr}(\mathbf{P})}{\partial \mathbf{K}} &= \text{Tr}(\partial_{\mathbf{K}}(\mathbf{B})) - \text{Tr}(\partial_{\mathbf{K}}(\mathbf{B}\mathbf{H}^T\mathbf{K}^T)) - \text{Tr}(\partial_{\mathbf{K}}(\mathbf{K}\mathbf{H}\mathbf{B})) + \text{Tr}(\partial_{\mathbf{K}}(\mathbf{K}\mathbf{R}\mathbf{K}^T)) + \text{Tr}(\partial_{\mathbf{K}}(\mathbf{K}\mathbf{H}\mathbf{B}\mathbf{H}^T\mathbf{K}^T)) \\ &= 0 \end{aligned} \quad (5.30)$$

The derivatives are put inside the trace because of the identity $\partial_x \text{Tr}(\mathbf{A}(x)) = \text{Tr}(\partial_x \mathbf{A}(x))$. The first term in Equation 5.30 is zero because $\mathbf{B} \neq \mathbf{B}(\mathbf{K})$. The second and third terms are simple derivatives.

$$\text{Tr}(\partial_{\mathbf{K}}(\mathbf{B}\mathbf{H}^T\mathbf{K}^T)) - \text{Tr}(\partial_{\mathbf{K}}(\mathbf{K}\mathbf{H}\mathbf{B})) = -\mathbf{H}(\mathbf{B}^T + \mathbf{B}) \quad (5.31)$$

The fourth term in Equation 5.30 can be solved using the chain rule.

$$\text{Tr}(\partial_{\mathbf{K}}(\mathbf{K}\mathbf{R}\mathbf{K}^T)) = \text{Tr}(\partial_{\mathbf{K}}(\mathbf{K})\mathbf{R}\mathbf{K}^T + \mathbf{K}\mathbf{R}\partial_{\mathbf{K}}(\mathbf{K}^T)) = \mathbf{R}\mathbf{K}^T + \mathbf{R}^T\mathbf{K}^T = (\mathbf{R} + \mathbf{R}^T)\mathbf{K}^T \quad (5.32)$$

The last term is more tricky, but again the chain rule is used.

$$\begin{aligned} \text{Tr}(\partial_{\mathbf{K}}(\mathbf{K}\mathbf{H}\mathbf{B}\mathbf{H}^T\mathbf{K}^T)) &= \text{Tr}((\partial_{\mathbf{K}}\mathbf{K})\mathbf{H}\mathbf{B}\mathbf{H}^T\mathbf{K}^T + \mathbf{K}\mathbf{H}\mathbf{B}\mathbf{H}^T(\partial_{\mathbf{K}}\mathbf{K}^T)) \\ &= \mathbf{H}\mathbf{B}\mathbf{H}^T\mathbf{K}^T + (\mathbf{K}\mathbf{H}\mathbf{B}\mathbf{H}^T)^T \\ &= \mathbf{H}(\mathbf{B} + \mathbf{B}^T)\mathbf{H}^T\mathbf{K}^T \end{aligned} \quad (5.33)$$

\mathbf{R} and \mathbf{B} are symmetric matrices, which means $\mathbf{R} = \mathbf{R}^T$ and $\mathbf{B} = \mathbf{B}^T$. Using these relations, Equation 5.30 becomes:

$$\frac{\partial \text{Tr}(\mathbf{P})}{\partial \mathbf{K}} = -2\mathbf{H}\mathbf{B}^T + 2\mathbf{R}^T\mathbf{K}^T + 2\mathbf{H}\mathbf{B}^T\mathbf{H}^T\mathbf{K}^T = 0. \quad (5.34)$$

Solving for \mathbf{K} then yields:

$$\begin{aligned} [(\mathbf{R}^T + \mathbf{H}\mathbf{B}^T\mathbf{H}^T)\mathbf{K}^T]^T &= (\mathbf{H}\mathbf{B}^T)^T \\ \mathbf{K}(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T) &= \mathbf{B}\mathbf{H}^T \\ \mathbf{K} &= \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}. \end{aligned} \quad (5.35)$$

Plugging \mathbf{K} back into Equation 5.25, an equation that minimizes the retrieval error is realized.

$$x^{\text{re}} = x^b + \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}(\theta^{\text{obs}} - \mathbf{H}x^b) \quad (5.36)$$

It is important to note that this is the same equation found in the previous section for x^{re} (Equation 5.17), but Equation 5.36 was derived with no assumptions about the shape of the *pdf*.

5.5 MLE vs MEV

Comparing the two methods described in the previous two subsections, it is not quite clear which one is “better” to use. The maximum likelihood estimation only works if there is some *pdf* assumed, but there really is no known *pdf* of errors in this application—a Gaussian is just assumed. If the *pdf* is some other distribution, the derivation and resulting error for the MLE is incorrect. The minimum error variance makes no assumptions about the probability distribution, which essentially makes it the better optimization technique, but the MEV solution still relies on some specified distribution to find the covariance matrices (\mathbf{B} , \mathbf{R} , and \mathbf{H}) needed to derive the retrieval error covariance matrix. Because of the lack of knowledge in the distribution, we end up assuming a Gaussian distribution. This assumption turns the MEV solution into the same solution as with the MLE.

5.6 Dimension of the Observation

There is one more important aspect needing explaining: $\langle J(x) \rangle = \frac{n}{2}$. To understand what that means, it first needs to be derived. Under the assumption that the *pdf* is a Gaussian, the average of the cost function from Section 5.4.1 is taken, using Equation 5.17 (or Equation 5.36). First an equation for $J(x^{\text{re}})$ is needed which is

obtained by plugging Equation 5.36 into Equation 5.9.

$$\begin{aligned}
J(x^{\text{re}}) &= \frac{1}{2} \left[(\mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}(\theta^{\text{obs}} - \mathbf{H}x^b))^T \mathbf{B}^{-1} \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}(\theta^{\text{obs}} - \mathbf{H}x^b) \right] \\
&\quad + \frac{1}{2} \left[(\theta^{\text{obs}} - \mathbf{H}(x^b + \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}(\theta^{\text{obs}} - \mathbf{H}x^b)))^T \mathbf{R}^{-1} \right. \\
&\quad \left. (\theta^{\text{obs}} - \mathbf{H}(x^b + \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}(\theta^{\text{obs}} - \mathbf{H}x^b))) \right]
\end{aligned} \tag{5.37}$$

The first term in Equation 5.37 can be simplified using some matrix manipulations and algebra.

$$\begin{aligned}
&(\mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}(\theta^{\text{obs}} - \mathbf{H}x^b))^T \mathbf{B}^{-1} \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}(\theta^{\text{obs}} - \mathbf{H}x^b) \\
&= (\theta^{\text{obs}} - \mathbf{H}x^b)^T (\mathbf{R}^T + \mathbf{H}\mathbf{B}^T \mathbf{H}^T)^{-1} \mathbf{H}\mathbf{B}^T \mathbf{B}^{-1} \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}(\theta^{\text{obs}} - \mathbf{H}x^b) \\
&= (\theta^{\text{obs}} - \mathbf{H}x^b)^T (\mathbf{R}^T + \mathbf{H}\mathbf{B}^T \mathbf{H}^T)^{-1} \mathbf{H}\mathbf{B}^T \mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}(\theta^{\text{obs}} - \mathbf{H}x^b)
\end{aligned} \tag{5.38}$$

The second term in Equation 5.37 can also be simplified with some algebra.

$$\begin{aligned}
&(\theta^{\text{obs}} - \mathbf{H}(x^b + \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}(\theta^{\text{obs}} - \mathbf{H}x^b)))^T \mathbf{R}^{-1} (\theta^{\text{obs}} - \mathbf{H}(x^b + \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}(\theta^{\text{obs}} - \mathbf{H}x^b))) \\
&= ((\theta^{\text{obs}})^T + (\mathbf{H}x^b)^T - (\theta^{\text{obs}} - \mathbf{H}x^b)^T (\mathbf{R}^T + \mathbf{H}\mathbf{B}^T \mathbf{H}^T)^{-1} \mathbf{H}\mathbf{B}^T \mathbf{H}^T) \mathbf{R}^{-1} \\
&\quad (\theta^{\text{obs}} - \mathbf{H}x^b + \mathbf{H}\mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}(\theta^{\text{obs}} - \mathbf{H}x^b)) \\
&= (\theta^{\text{obs}} - \mathbf{H}x^b)^T (\mathbf{I} - (\mathbf{R}^T + \mathbf{H}\mathbf{B}^T \mathbf{H}^T)^{-1} \mathbf{H}\mathbf{B}^T \mathbf{H}^T) \mathbf{R}^{-1} (\theta^{\text{obs}} - \mathbf{H}x^b + \mathbf{H}\mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}(\theta^{\text{obs}} - \mathbf{H}x^b))
\end{aligned} \tag{5.39}$$

Plugging the simplified first and second parts (Equation 5.38 and 5.39) back into Equation 5.37 yields Equation 5.40.

$$\begin{aligned}
J(x^{\text{re}}) &= \frac{1}{2} (\theta^{\text{obs}} - \mathbf{H}x^b)^T \left[(\mathbf{R}^T + \mathbf{H}\mathbf{B}^T \mathbf{H}^T)^{-1} \mathbf{H}\mathbf{B}^T \mathbf{H}^T (\mathbf{I} + \mathbf{R}^{-1} \mathbf{H}\mathbf{B}\mathbf{H}^T) (\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1} + \mathbf{R}^{-1} \right. \\
&\quad \left. - \mathbf{R}^{-1} \mathbf{H}\mathbf{B}\mathbf{H}^T (\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1} - (\mathbf{R}^T + \mathbf{H}\mathbf{B}^T \mathbf{H}^T)^{-1} \mathbf{H}\mathbf{B}^T \mathbf{H}^T \mathbf{R}^{-1} \right] (\theta^{\text{obs}} - \mathbf{H}x^b)
\end{aligned} \tag{5.40}$$

The terms inside the brackets in the above equation can be simplified even further. Manipulating the equation by multiplying it by $\mathbf{R}\mathbf{R}^{-1}$ does not change the actual equation and helps simplify the terms.

$$\begin{aligned}
&(\mathbf{R}^T + \mathbf{H}\mathbf{B}^T \mathbf{H}^T)^{-1} \mathbf{H}\mathbf{B}^T \mathbf{H}^T (\mathbf{I} + \mathbf{R}^{-1} \mathbf{H}\mathbf{B}\mathbf{H}^T) (\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1} \\
&= (\mathbf{R}^T + \mathbf{H}\mathbf{B}^T \mathbf{H}^T)^{-1} \mathbf{H}\mathbf{B}^T \mathbf{H}^T \mathbf{R} (\mathbf{I} + \mathbf{R}^{-1} \mathbf{H}\mathbf{B}\mathbf{H}^T) (\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1} \mathbf{R}^{-1} \\
&= (\mathbf{R}^T + \mathbf{H}\mathbf{B}^T \mathbf{H}^T)^{-1} \mathbf{H}\mathbf{B}^T \mathbf{H}^T (\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T) (\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1} \mathbf{R}^{-1} \\
&= (\mathbf{R}^T + \mathbf{H}\mathbf{B}^T \mathbf{H}^T)^{-1} \mathbf{H}\mathbf{B}^T \mathbf{H}^T \mathbf{R}^{-1}
\end{aligned} \tag{5.41}$$

Plugging the above result back in to Equation 5.40, the result is able to be simplified even further.

$$\begin{aligned}
J(x^{\text{re}}) &= \frac{1}{2}(\theta^{\text{obs}} - \mathbf{H}x^b)^T \left[(\mathbf{R}^T + \mathbf{H}\mathbf{B}^T\mathbf{H}^T)^{-1}\mathbf{H}\mathbf{B}^T\mathbf{H}^T\mathbf{R}^{-1} + \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{H}\mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1} \right. \\
&\quad \left. - (\mathbf{R}^T + \mathbf{H}\mathbf{B}^T\mathbf{H}^T)^{-1}\mathbf{H}\mathbf{B}^T\mathbf{H}^T\mathbf{R}^{-1} \right] (\theta^{\text{obs}} - \mathbf{H}x^b) \\
&= \frac{1}{2}(\theta^{\text{obs}} - \mathbf{H}x^b)^T \left[\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{H}\mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1} \right] (\theta^{\text{obs}} - \mathbf{H}x^b) \\
&= \frac{1}{2}(\theta^{\text{obs}} - \mathbf{H}x^b)^T \left[\mathbf{R}^{-1}(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T) - \mathbf{R}^{-1}\mathbf{H}\mathbf{B}\mathbf{H}^T \right] (\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1} (\theta^{\text{obs}} - \mathbf{H}x^b) \\
&= \frac{1}{2}(\theta^{\text{obs}} - \mathbf{H}x^b)^T (\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1} (\theta^{\text{obs}} - \mathbf{H}x^b)
\end{aligned} \tag{5.42}$$

Next, the expectation value is taken, which is the trace of $J(x^{\text{re}})$.

$$\begin{aligned}
\langle J(x^{\text{re}}) \rangle &= \text{Tr} \left[\frac{1}{2}(\theta^{\text{obs}} - \mathbf{H}x^b)^T (\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1} (\theta^{\text{obs}} - \mathbf{H}x^b) \right] \\
&= \frac{1}{2} \text{Tr} \left[(\theta^{\text{obs}} - \mathbf{H}x^b)(\theta^{\text{obs}} - \mathbf{H}x^b)^T (\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1} \right]
\end{aligned} \tag{5.43}$$

The second step in the above equation is allowed because the trace function allows cyclic permutations. From here, the true values for the radiance and concentration are added and subtracted, and the equation is simplified using Equations 5.18 and 5.21.

$$\begin{aligned}
\langle J(x^{\text{re}}) \rangle &= \frac{1}{2} \text{Tr} \left[(\theta^{\text{obs}} - \theta^t + \theta^t - \mathbf{H}x^b)(\theta^{\text{obs}} - \theta^t + \theta^t - \mathbf{H}x^b)^T (\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1} \right] \\
&= \frac{1}{2} \text{Tr} \left[(\epsilon^{\text{obs}} - \mathbf{H}\epsilon^b)(\epsilon^{\text{obs}} - \mathbf{H}\epsilon^b)^T (\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1} \right] \\
&= \frac{1}{2} \text{Tr} \left[(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1} \right] \\
&= \frac{1}{2} \text{Tr}(\mathbf{I})
\end{aligned} \tag{5.44}$$

In the end, $\langle J(x^{\text{re}}) \rangle = \frac{1}{2} \text{Tr}(\mathbf{I}) = \frac{n}{2}$ where n is just a number and represents the dimension of the observation—it is the number of channels used.

5.7 Retrieval of CO₂

For this project, we used the MLE retrieval method. This method has been described in much length in previous sections, but the only missing components are the derivations of Equations 2.3 and 2.4, the effects of various \mathbf{B}_{CO_2} and \mathbf{R} matrices, and an explanation of the background term calculation.

5.7.1 Retrieved x_{CO_2} Equation

Equations 2.3 and 2.4 have a relatively straight forward derivation. Starting from the cost function, Equation 5.9, substitution of $\Delta x_{\text{CO}_2} = x_{\text{CO}_2}^b - x_{\text{CO}_2}$, and a minimization with respect to Δx_{CO_2} , yields an equation for Δx_{CO_2} .

$$\begin{aligned}
J(x_{\text{CO}_2}) &= \frac{1}{2} \Delta x_{\text{CO}_2}^T \mathbf{B}_{\text{CO}_2}^{-1} \Delta x_{\text{CO}_2} + \frac{1}{2} (\theta^{\text{obs}} + \theta^b - \theta^b - \mathbf{H}_{\text{CO}_2} x_{\text{CO}_2})^T \mathbf{R}^{-1} (\theta^{\text{obs}} + \theta^b - \theta^b - \mathbf{H}_{\text{CO}_2} x_{\text{CO}_2}) \\
&= \frac{1}{2} \Delta x_{\text{CO}_2}^T \mathbf{B}_{\text{CO}_2}^{-1} \Delta x_{\text{CO}_2} + \frac{1}{2} (\mathbf{H}_{\text{CO}_2} \Delta x_{\text{CO}_2} - \Delta \theta)^T \mathbf{R}^{-1} (\mathbf{H}_{\text{CO}_2} \Delta x_{\text{CO}_2} - \Delta \theta)
\end{aligned} \tag{5.45}$$

$$\begin{aligned}
\frac{\partial J}{\partial \Delta x_{CO_2}} &= \Delta x_{CO_2} \mathbf{B}_{CO_2}^{-1} + (\mathbf{H}_{CO_2} \Delta x_{CO_2} - \Delta \theta) \mathbf{R}^{-1} \mathbf{H}_{CO_2} = 0 \\
\Delta x_{CO_2} (\mathbf{B}_{CO_2}^{-1} + \mathbf{H}_{CO_2}^T \mathbf{R}^{-1} \mathbf{H}_{CO_2}) &= \mathbf{H}_{CO_2}^T \mathbf{R}^{-1} \Delta \theta \\
\Delta x_{CO_2} &= (\mathbf{B}_{CO_2}^{-1} + \mathbf{H}_{CO_2}^T \mathbf{R}^{-1} \mathbf{H}_{CO_2})^{-1} \mathbf{H}_{CO_2}^T \mathbf{R}^{-1} \Delta \theta \\
\Delta x_{CO_2} &= \mathbf{P}_{CO_2} \mathbf{H}_{CO_2}^T \mathbf{R}^{-1} \Delta \theta
\end{aligned} \tag{5.46}$$

Lastly, Equation 2.3 is just a manipulation of $\Delta x_{CO_2} = x_{CO_2}^b - x_{CO_2}$.

$$x_{CO_2} = x_{CO_2}^b - \Delta x_{CO_2}$$

5.7.2 Effects of Background and Observational Error on Retrieval Error

The effects of the background error covariance, \mathbf{B}_{CO_2} , and observational error covariance, \mathbf{R} , on the retrieval error covariance, \mathbf{P}_{CO_2} was discussed in Section 5.4.1. To illustrate the effects, Figure 5.12 shows a figure of $\mathbf{B}_{RMSE} - \mathbf{P}_{RMSE}$ for two different \mathbf{B}_{CO_2} matrices and three different \mathbf{R} matrices. The background error covariances chosen for this explanation had constant variances, corresponding to roughly 1% and 2% error. The three values chosen for the observational error covariance were 25 mK, 50 mK, and 100 mK. The actual value of \mathbf{R} varies with channel, but each channel's error is within or even less than this range. In an effort to not underestimate the error, this range was considered optimal. The graph in Figure 5.12 shows the minimization of error, meaning larger value is optimal. But, it is important to mention that even if the error minimization is larger with a higher background error, the overall error is not definitively minimal. Looking at the far right green curve of Figure 5.12, even with the largest minimization, the error is still roughly 7 ppm, versus the far left blue curve, which at most has minimum error of roughly 3.6 ppm. The blue curve still overall has lower error. In short, if the background error is large, a smaller observational error results in the retrieval error being minimized, but the minimization of error does not overcome the large error with which the retrieval began.

5.7.3 Background term, \mathbf{B}_{CO_2}

The background error covariance matrix is a function of the variance, as well as the correlation, terms that were briefly mentioned in Section 5.4.1.

$$\mathbf{B}_{CO_2} = \mathbf{\Sigma}^T \mathbf{C} \mathbf{\Sigma} \tag{5.47}$$

$$\mathbf{\Sigma}_{i,i} = \left[\frac{1}{M} \sum_{m=1}^M (x_i^m - \bar{x}_i)^2 \right]^{1/2} \tag{5.48}$$

$$\mathbf{C}_{i,j} = \frac{\frac{1}{M} \sum_{m=1}^M (x_i^m - \bar{x}_i)(x_j^m - \bar{x}_j)}{\mathbf{\Sigma}_{i,i} \mathbf{\Sigma}_{j,j}} \tag{5.49}$$

$$\bar{x}_k = \frac{1}{M} \sum_{i=1}^M x_k \tag{5.50}$$

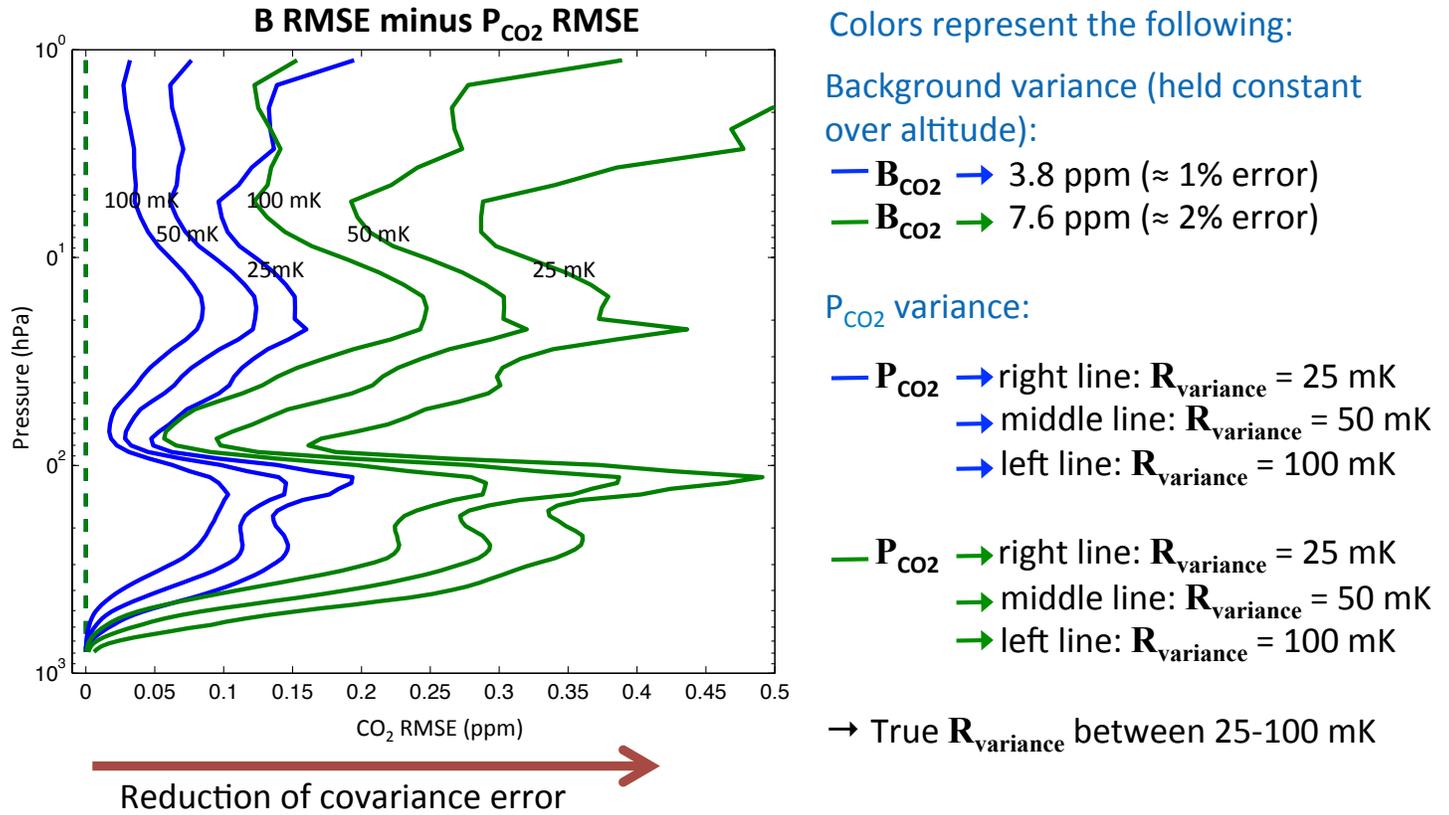


Figure 5.12: Chart showing the effects of B_{CO_2} and R on the minimization of P_{CO_2} .

M in the above equations is the number of different background data used. To obtain accurate statistics, the background error should be calculated using data from multiple sources, multiple years, etc. For this project, we decided to use data within a certain radius to gather ample statistics—our background data consisted of data 5° North/South and 30° East/West from the point of retrieval. We focused on retrieving a profile over the Pacific Ocean, for these reasons specifically—the vertical profiles do not change much within this degree range over the ocean. Also, to ensure we were not underestimating the variance, we doubled our best estimate. The correlation matrix (at 177.5° West and 2.5° South), calculated using this method, is shown in Figure 5.13.

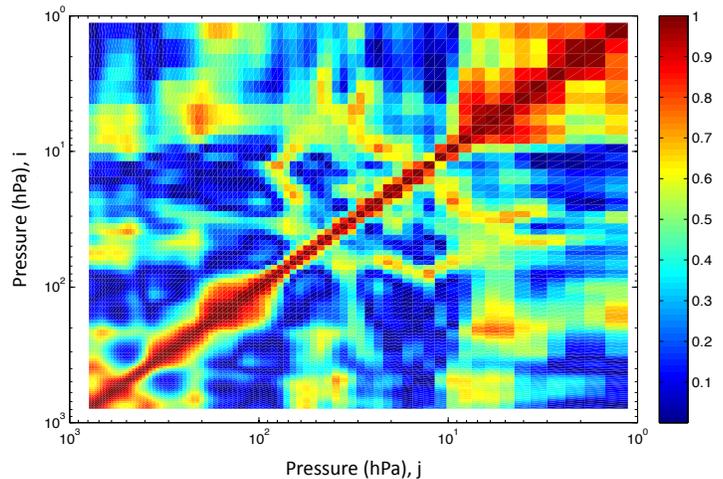


Figure 5.13: Correlation matrix calculated from Equation 5.49. Axes show i^{th} and j^{th} components.

The correlation matrix shown in Figure 5.13 is very noisy, and is not exactly accurate. There should be more correlation through the stratosphere, and there should not be any correlation between concentrations high in the

stratosphere and low in the troposphere, and Figure 5.13 shows that there are (those are the yellow to red spots at the off diagonal corners). This is a result of our choice of background data, and to compensate for this error, we fit a smooth curve to the error variance, which is shown in Figure 5.14.

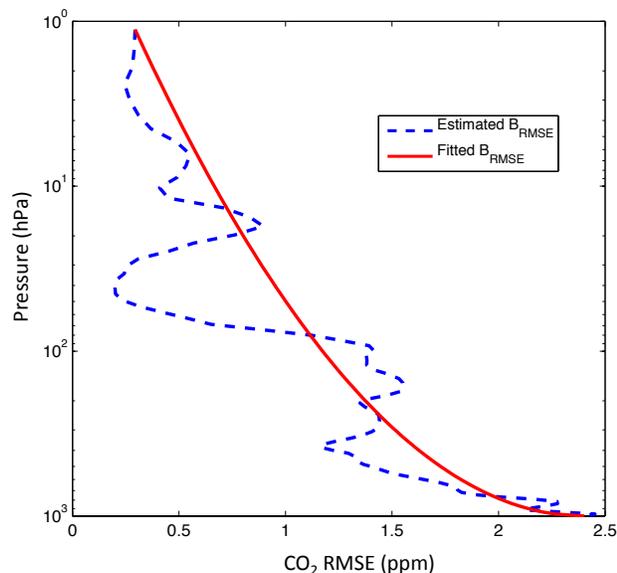


Figure 5.14: Calculated and fitted B_{RMSE} .

The fitted curve shown in Figure 5.14 is what was used to produce the retrieval results specified in Section 3.

Acknowledgments

The author gratefully acknowledges shared financial support from the California Institute of Technology Student Faculty Programs (SFP) Summer Undergraduate Research Fellowship (SURF) as well as the Jet Propulsion Laboratory (JPL) Research and Technology Development Fund (R&TD). I thank Professor Paul Dimotakis (mentor), Zhijin Li (mentor), and Edward Olsen for their tremendous help, advice, and guidance, as well as Luke Chen for developing the software to

extract a global selection of vertical profiles from ECMWF data and the associated profile of CO_2 from the MOZART model. I also thank Stephen Licata for developing the software that processed the vertical profiles through the radiative transfer forward model to generate the model's Jacobian functions. Their software development tasks were carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

References

- [1] P. Dimotakis, B. Walker, K. Jonietz, and D. Rotman, "A greenhouse-gas information system: Monitoring and validating emissions reporting and mitigation," tech. rep., Jet Propulsion Laboratory, Sandia National Laboratories, Los Alamos National Laboratory, and Lawrence Livermore National Laboratory, 2011.
- [2] M. Chahine, L. Chen, P. Dimotakis, X. Jiang, Q. Li, E. Olsen, T. Pagano, J. Randerson, and Y. Yung, "Satellite remote sounding of mid-tropospheric CO_2 ," *Geophysical Research Letters*, vol. 35, 2008.
- [3] M. T. Chahine, T. S. Pagano, H. H. Aumann, R. Atlas, C. Barnet, J. Blaisdell, L. Chen, M. Divakarla, E. J. Fetzer, M. Goldberg, C. Gautier, S. Granger, S. Hannon, F. W. Irion, R. Kakar, E. Kalnay, B. H. Lambrigtsen, S. Y. Lee, J. Le Marshall, W. W. McMillan, L. McMillin, E. T. Olsen, H. Revercomb, P. Rosenkranz, W. L. Smith, D. Staelin, L. L. Strow, D. Susskind, J. and Tobin, W. Wolf, and L. Zhou, "The atmospheric infrared sounder (airs): improving weather forecasting and providing new insights into climate," *Bulletin of the American Meteorological Society*, vol. 87, pp. 911–926, 2006.

- [4] M. Chahine, C. Barnet, E. Olsen, L. Chen, and E. Maddy, "On the determination of atmospheric minor gases by the method of vanishing partial derivatives with application to CO_2 ," *Geophysical Research Letters*, vol. 32, 2005.
- [5] C. D. Rodgers, *Inverse Methods for Atmospheric Sounding: Theory and Practice*. World Scientific, 2000.
- [6] "Near earth system laboratory, atmospheric chemistry division."
- [7] P. Dimotakis, L. Chen, Z. Li, S. Licata, E. Olsen, and T. Pagano, "Retrieval of vertical t , H_2O , O_3 , and CO_2 profiles, with sensitivity analysis (sa) and uncertainty quantification (uq)," in *AIRS Science Team Meeting*, (Caltech, Pasadena, CA), April 2012.
- [8] X. Jiang, Q. Li, M. C. Liang, R. L. Shia, M. Chahine, E. Olsen, L. Chen, and Y. Yung, "Simulation of upper tropospheric CO_2 from chemistry and transport models," *Global Biogeochemical Cycles*, vol. 22, 2008.
- [9] X. Jiang, M. Chahine, E. Olsen, L. Chen, and Y. Yung, "Interannual variability of mid-tropospheric CO_2 from atmospheric infrared sounder," *Geophysical Research Letters*, vol. 37, 2010.
- [10] T. Pagano, M. Chahine, and E. Olsen, "Seven years of observations of mid-tropospheric CO_2 from the atmospheric infrared sounder," in *61st International Astronautical Congress*, International Astronautical Federation, 2010.
- [11] K. Li, B. Tian, D. Waliser, and Y. Yung, "Tropical mid-tropospheric CO_2 variability driven by the Madden-Julian oscillation," *PNAS*, 2010.
- [12] J. Wang, X. Jiang, M. Chahine, M. C. Liang, E. Olsen, L. Chen, S. Licata, T. Pagano, and Y. Yung, "The influence of tropospheric biennial oscillation on mid-tropospheric CO_2 ," *Geophysical Research Letters*, vol. 38, 2011.
- [13] J. Liu, I. Fung, E. Kalnay, J. S. Kang, E. Olsen, and L. Chen, "Simultaneous assimilation of AIRS CO_2 and meteorological observations in a carbon climate model with an ensemble Kalman filter," tech. rep., Jet Propulsion Laboratory, California Institute of Technology; University of California, Berkeley; University of Maryland, College Park, 2011.
- [14] T. Pagano, E. Olsen, M. Chahine, A. Ruzmaikin, H. Nguyen, and X. Jiang, "Monthly representations of mid-tropospheric carbon dioxide from the atmospheric infrared sounder," *SPIE*, vol. 8158, 2011.
- [15] S. C. Wofsy and the HIPPO Science Team, "HiAPER pole-to-pole observations (HIPPO): fine-grained, global-scale measurements of climatically important atmospheric gases and aerosols," *Philosophical Transactions of the Royal Society*, vol. 369, pp. 2073–2086, 2011.
- [16] X. Jiang, J. Wang, E. Olsen, M. Liang, T. Pagano, L. Chen, S. Licata, and Y. Yung, "Influence of El Niño on mid-tropospheric CO_2 from atmospheric infrared sounder and model." Submitted to *JAS*, 2012.