Atmospheric
Measurement
Techniques
Discussions
Open Access
EGU

An SNR-Optimized Scanning Strategy for Geostationary Carbon Cycle Observatory (GeoCarb) Instrument

Jeffrey Nivitanont[1] and Sean Crowell[1]

[1]GeoCarb, University of Oklahoma, 301 David L. Boren Blvd., 4 Partners Place, Suite 1120, Norman, OK 73072

Correspondence: Jeffrey Nivitanont (jeffniv@ou.edu)

Abstract. The Geostationary Carbon Observatory (GeoCarb) will make measurements of greenhouse gases over the land mass in the western hemisphere. The extreme flexibility of observing from geostationary orbit induces an optimization problem, as operators must choose what to observe and when. We express this problem in terms of an optimal subcovering problem, and use an Incremental Optimization (IO) algorithm to create a scanning strategy that minimizes expected error as a function of

5   the signal-to-noise ratio (SNR), and show that this method outperforms the "human selected" strategy in terms of global error distributions.

1   Introduction

Understanding the effects of anthropogenic carbon dioxide (CO_2) on the carbon cycle requires us to understand the spatial distribution of atmospheric CO_2 concentrations to identify natural and anthropogenic sources and sinks. In addition to a sparse

10   in situ sampling network, ground-based remote sensing measurements are currently obtained from the Total Column Carbon Observing Network (TCCON) and space-based measurements from the Orbiting Carbon Observatory (OCO-2) (Eldering et al. (2017a), Eldering et al. (2017b), Crisp et al. (2017), Crisp et al. (2008), Crisp et al. (2004)) and Greenhouse Gases Observing Satellite (GOSAT) (Kuze et al. (2009), Yokota et al. (2009), Hammerling et al. (2012)). These instruments have provided a wealth of data for understanding the global carbon cycle in the recent years. However, these instruments have spatial and

15   temporal limitations. The repeat cycles of the space-based instruments force the spatial and temporal interpolation of the atmospheric CO_2 concentrations within their respective cycles, 3 days for GOSAT (Kuze et al. (2009)) and 16 days for OCO-2 (Miller et al. (2007)). The sparsity of the TCCON measurement sites restricts the latitudinal range of observations. The new Geostationary Carbon Observatory (GeoCarb) (Polonsky et al. (2014)) will allow us to augment the current remote sensors on the ground and in space.

20   Recently selected as the NASA's Earth Venture Mission-2 (EVM-2), GeoCarb is set to launch into geostationary orbit in 2022 to be positioned at $85°W$ with the mission of improving the understanding of the carbon cycle. Building on the work of OCO-2, GeoCarb will observe reflected sunlight daily over the Americas, and retrieve the column average dry air mole fraction of carbon dioxide (XCO$_2$), carbon monoxide (XCO), methane (XCH$_4$), and solar-induced fluorescence (SIF). GeoCarb views reflected sunlight from Earth through a narrow slit that projects on the Earth's surface to an area measuring

25   about 1,740 miles (2,800 kilometers) from north to south and about 3.7 miles (6 kilometers) from east to west. The instrument

makes measurements along the slit with a 4.5 second integration. Instrument pointing is accomplished by way of two scanning mirrors that shift the field of view (FoV) north-south and east-west. The pointing system is extremely flexible, and observations can be made at any location at any time with sufficient solar illumination. This flexibility induces an optimization problem: where should the instrument take measurements at a given time throughout the day?

5   Determining when and where to make daily scans with GeoCarb's observing capabilities is mathematically similar to a CO_2 observation network optimization problem for establishing new observation sites. Selecting the optimal location of new observing stations has been shown to be feasible by utilizing various optimization algorithms. There have been previous studies performed on the problem of optimizing CO_2 observation networks utilizing computationally expensive evolutionary algorithms [i.e. Simulated Annealing (Rayner et al. (1996); Gloor et al. (2000)) and Genetic Algorithm (Nickless et al. (2018))]

10   and one utilizing a deterministic, incremental algorithm (Patra and Maksyutov (2002)). All of the previous studies utilized their optimization routines to minimize CO_2 measurement uncertainty as a function of signal-to-noise ratio (SNR).

In this paper, a deterministic, incremental optimization routine (IO) is utilized to find a scanning strategy for GeoCarb that minimizes expected measurement uncertainty as a function of SNR. Section 2 gives background information on the GeoCarb mission and the objectives for this paper. Section 3 explains the process used to create the SNR-optimized algorithm and how

15   the expected error is calculated from the simulated retrievals. In Section 4, the distribution of global error of the algorithm-selected strategy is compared to a baseline strategy to determine the performance of the algorithm and the sensitivity of algorithm to inputs is investigated, with results discussed in Sect. 5. We offer concluding statements in Sect. 6 and future research goals.

2 Background

20   GeoCarb will be hosted on a SES Government Solutions (http://www.ses-gs.com) communications satellite in geostationary orbit at 85°W. It will measure reflected near-infrared sunlight in the O_2 band at 0.76μm to measure total column O_2, weak and strong CO_2 bands at 1.61μm and 2.06μm to measure XCO_2, and the CH_4 and CO bands 2.32μm for measuring XCH_4 and XCO. The O_2 spectral band is identical to that of the OCO-2 mission and allows for determination of mixing ratios and the measurement of SIF, as well as additional information on aerosol and cloud contamination of retrievals. The baseline mission

25   for GeoCarb aims to produce column-averaged mixing ratios of CO2, CH4 and CO with accuracy per sample of 0.7% (≈ 2.7 ppm), 1% (≈ 18 ppb) and 10% (≈ 10 ppb), respectively. Geostationary orbit offers two main advantages over low Earth orbit (LEO). First, the signal-to-noise (SNR) is proportional to the square root of the dwell time for detectors limited by photon shot noise. Geostationary orbits enables longer observation times, thereby increasing SNR. Second, areas with high and uncertain anthropogenic emissions of CO_2, CH_4 and CO may be targeted with contiguous sampling, relatively small spatial footprints

30   and fine temporal resolution allowing for several observations per day on continental scales are possible.

We are presented the problem of finding an optimized scanning strategy for the GeoCarb satellite instrument. The underlying abstract mathematical problem related to optimizing the scanning pattern is the Geometric Set Cover problem (Hetland (2014)). Given a finite set of points in space and a set of subsets, the objective is to find a minimal set of scan blocks whose union covers

Atmospheric
Measurement
Techniques
Discussions

all points in the space. This idea is identical to a network optimization problem comparing the coverage area of a potential

network observation site to a geometric subset in the space to be covered. The task of determining the locations of new

observation sites so that the total number of required sites to cover an area is minimal is solved similarly. Therefore, we looked

at methods used for optimizing network observation sites for our application. However, our motivation extends beyond just

5 finding a minimal covering set. Our goal is to find a minimal covering set that is operationally efficient and minimizes global

measurement error for the GeoCarb instrument.

Specific to the instrument's application, the geometric subsets are 5-minute East-to-West scan blocks, shown in Fig. 1, and

the points in space that we are trying to cover is the North American and South American land masses between $50°N$ and

$50°S$. Because measurement errors are influenced by parameters that vary in space and time, the solutions are in the form of

10 ordered sets where the scan blocks are ordered by time of execution. With the simplifying assumptions of making our problem

computationally tractable and minimizing scan coverage over the ocean, we propose a candidate set of 135 scan blocks (Fig.

1). This is a much larger candidate set than those of the network optimization studies that utilized evolutionary algorithms

(Rayner et al. (1996); Gloor et al. (2000)). Therefore, the computationally efficient Incremental Optimization (IO) procedure

was implemented to select scan blocks that minimize our objective function at each increment in time.

15 ## 3   Methods

3.1   Scan Blocks

We assume that GeoCarb will process commands in terms of 5-minute scan blocks, during which the instrument steps the slit

from east to west. The set of considered scan blocks, shown in Fig. 1, purposely excludes potential scans that are primarily

over the ocean, as GeoCarb will not be able to make retrievals over water surfaces due to lack of signal in the CO_2 spectral

20 bands. The potential scans are also largely restricted to land between $50°N$ and below $50°S$ due to larger solar zenith angles

at the higher latitudes. Each slit observation (i.e. 1016 individual soundings) is assumed to take 5 seconds, after which the slit

moves to the west by half a slit width.

3.2   Science Operations Timeline

A goal of this study is to create a scanning strategy that views all land masses of interest within the time window of usable

25 daylight. To determine what time of day to begin the scanning process, Macapá, Brazil and Mexico City, Mexico were chosen

as geographic reference points (Fig. 1) to determine the beginning and ending time, respectively, of the usable daylight time

frame. Macapá is located at $(0°, 50°W)$ at the mouth of the Amazon river and being on the equator gives us a consistent

starting time relative to airmass factor (AF), a function of solar zenith angle (SZA) and the sensor zenith angle (ZA), where

$AF = \frac{1}{\cos(SZA)} + \frac{1}{\cos(ZA)}$. Located at $(19.5°N, 99.25°W)$, Mexico City, Mexico is an ideal reference point to determine when

30 the window of usable daylight ended because it is longitudinally centered in the North American land mass while being close

enough to the equator for the calculated airmass factors to remain consistent through the winter months. The scanning strategy

calculates the starting time when Macapá first exceeds a starting threshold for AF and the ending time when Mexico City drops below an ending threshold for AF to determine when the usable daylight time window is over. As a result of parameter exploration experiments described in section 3.7, the suggested starting threshold is AF $= 2.6$ for the Summer Solstice and AF $= 2.7$ for the Autumn Equinox for minimum variance in predicted errors.

3.3 Uncertainty in Retrieved Gas Concentrations

GeoCarb retrieves gas concentrations using reflected sunlight. The radiance, I, observed by GeoCarb is an aggregate of insolation and atmospheric and land surface processes that absorb, reflect, and scatter photons. The impact of these processes is parameterized using a simple model from Polonsky et al. (2014) that incorporates the effects of surface albedo and attenuation by aerosols over the sun-Earth-satellite path described by the solar zenith angle (SZA) and the sensor zenith angle (ZA):

$$I = F_{sun}\alpha\cos(SZA)e^{-AF\tau} \tag{1}$$

where F_{sun} is the band-specific solar irradiance, α is the band-specific surface albedo, and τ is the optical depth (OD) of atmospheric scatterers (e.g. aerosols). For our simple model, we assumed a cloud-free atmosphere, whereas in the operational environment, clouds play a major role in retrieval quality due to poorly understood 3-D scattering effects. As can be readily verified, larger zenith angles lead to reduced signal for constant scatterer OD, as does smaller surface albedo. Note that τ is a quantity with significant spatial and temporal variability, as aerosol concentrations are modified by atmospheric dynamics, emissions, and chemistry. Typical values of τ in successful retrievals for OCO-2 are less than 0.6 for nadir soundings near the equator and decrease as AF increases. Similarly, surface albedo varies with land cover type on small spatial scales, and throughout the year with vegetation density. The OD term was set to $\tau = 0.3$ as it was previously found to be a reasonable estimate for a "clear" sky retrieval (Crisp et al. (2004), O'Dell et al. (2012)).

An important indicator of observation quality is the signal-to-noise ratio (SNR). In the case of GeoCarb, the signal is referred to as I, and the instrument noise equivalent spectral radiance, N, as

$$N(I) = \sqrt{N_0{}^2 + N_1 I} \tag{2}$$

where N_0 and N_1 are parameters that empirically capture the effects of the instrument design (e.g. telescope length, detector noise) on overall instrument noise (O'Dell et al. (2012)). The O_2 A-band (0.763μm) specific constants, $N_0 = 0.1819$ and $N_1 = 0.003295$ are used in the noise model, which are figures derived from airborne trials with the Tropospheric Infrared Mapping Spectrometers (TIMS) by Lockheed Martin (Kumer et al. (2013)), that were later revised in Polonsky et al. (2014). The SNR is then defined as $\frac{I}{N}$.

In O'Brien et al. (2016), the authors fitted an empirical model to predict the observational uncertainty as a function of the measurement SNR. They found that the function

$$\sigma = \frac{1}{\frac{1}{14} + (0.0039)SNR}. \tag{3}$$

Atmospheric
Measurement
Techniques
Discussions

best captures the observation error distribution as SNR increases. The same model is used to connect SNR and uncertainty for later scanning strategy distributions. The distribution of σ is treated as the metric against which a particular scanning sequence is evaluated.

3.4 Objective Function

5 Examining the definition of SNR, it is easy to see that

$$SNR = \frac{I}{N} \approx k\sqrt{I}, \text{ where } k \text{ is a constant.} \tag{4}$$

Since the goal is to ultimately maximize I, we define an objective function that minimizes its multiplicative inverse. In addition to minimizing SNR, two constraints were included in the objective function to prevent erratic behavior in the scanning strategy. An overlap term was introduced to minimize repeated coverage of regions. A distance term, δ, was also included, which is the

10 squared linear distance from the boundary of the last selected scan block to a candidate scan block. The objective function, c, to be minimized is given by

$$c(s,t) = \text{median}(e^{AF(s,t,x,y)}\alpha(x,y)^{-1})(1 + \frac{(s \cap \mathbb{I}) + \delta^2}{s \cap \mathbb{E}}), \tag{5}$$

where

$\quad s = $ A candidate scan block.

15 $\quad t = $ Time.

$\quad \mathbb{E} = $ Uncovered land mass

$\quad \mathbb{I} = $ The set of selected scan blocks

$\quad \alpha(x,y) = $ Surface albedo of a point in scan block.

$\quad \delta = $ Distance from last selected scan block.

20 $\quad AF(s,t,x,y) = $ Airmass factor of a point in a scan block with respect to time.

The median of $e^{AF(s,t,x,y)}\alpha(x,y)^{-1}$ is used because we assume that the distributions of airmass factor and surface albedo are non-Gaussian within the scan blocks. The high variability of both parameters are described in Section 3.4.2

3.4.1 Surface Albedo

The MCD43C3 Version 6 White Sky Albedo MODIS band 6 data set (Schaaf and Wang (2015)) was utilized for obtaining

25 surface albedo, α. The MODIS BRDF/Albedo product combines multiband, atmospherically corrected surface reflectance data from the MODIS and MISR instruments to fit a Bidirectional Reflectance Distribution Function (BRDF) in seven spectral bands at a 1 km spatial resolution on a 16-day cycle (Lucht et al. (2000)). The White Sky Albedo measure is a bihemispherical reflectance obtained by integrating the BRDF over all viewing and irradiance directions. These albedo measures are purely properties of the surface, therefore they are compatible with any atmospheric specification to provide true surface albedo as an

30 input to regional and global climate models. The native data was aggregated to the $0.5°$ spatial resolution, and interpolated in time to daily resolution.

Atmospheric
Measurement
Techniques
Discussions
Open Access

3.4.2 Seasonal Variation of Parameters

Since AF is affected by the sun's position and albedo is affected by the density of vegetation, there are large seasonal variations in both of these variables, shown in Fig. 2 and 3. However, there is little to no variation between day-to-day comparisons of these variables. It suffices then, and gives an added advantage of being computationally efficient, to calculate separate scanning
5   strategies for each month rather than day.

3.5 Optimization Algorithms

The time-dependency of the scanning strategy requires the solutions to be represented as ordered scan blocks of the candidate set. Therefore, the sum of permutations $\sum_{k=1}^{135} \frac{135!}{(135-k)!}$ gives approximately 7×10^{230} possible solutions. Since it is computationally intractable to evaluate all possible solutions, a Greedy heuristic algorithm was employed to find a minimal covering
10   set as a lower-bound estimate for set cardinality, and then it was modified to an Incremental Optimization (IO) algorithm to find a scanning strategy optimizing for SNR.

3.5.1 Greedy Algorithm

Viewing the North American and South American land masses as a uniform space to be covered without considering any additional constraints, the problem is a Geometric Set Cover problem where the goal is to find a minimal cardinality covering
15   set that we will call optimal. It is well-known that there are no known analytical solutions to the Set Cover problem, as it is one of Karp's 21 NP-Complete problems, and the optimization version is NP-Hard (Karp (1972)). However, there exists a heuristic method for finding a solution called the Greedy algorithm that selects the cover with the largest intersection with the uncovered space recursively until the space is covered (Hetland (2014)). The pseudo-code is shown in Algorithm 1. The Greedy algorithm is computationally efficient, though it is difficult to verify that the solution it finds is the *optimal* solution. The Greedy algorithm
20   is suitable for our application because it reduces the set of candidate blocks at each iteration by removing the selected scan blocks and this ensures that there are no repeated scan blocks in a scanning strategy. Running the Greedy heuristic with no objective function shows that the area of interest can be covered using 83 scan blocks. Therefore, we took this as the lower bound of covering set size.

Algorithm 1 Greedy Algorithm

    $\mathbf{E} \leftarrow$ Space to be covered
    $\mathbf{S} \leftarrow$ Set of scan blocks where $E \subseteq \bigcup s_i \in \mathbf{S}$
    $\mathbf{I} \leftarrow \emptyset$
    while $\mathbf{E} \subsetneq \mathbf{I}$ **do**
        Find $s^* \in \mathbf{S}$ such that $s^* \cap \mathbf{E}$ is maximal for all $s_i \in \mathbf{S}$
        Append s^* to $\mathbf{I}$
        Remove s^* from $\mathbf{S}$
    end while

Atmospheric
Measurement
Techniques
Discussions

Open Access

3.5.2 Incremental Optimization

The Greedy algorithm was modified to select the scan block that minimizes the objective function at each iteration to satisfy operational constraints. Presented in Patra and Maksyutov (2002), this modification to the Greedy algorithm makes it an Incremental Optimization (IO) algorithm because its goal is to minimize the objective function at each increment of time to

5 find the global optimum. Like the Greedy algorithm, IO has the advantage of being computationally inexpensive. However, it may find local optima only and produce sub-optimal solutions depending on the nature of the problem. Usually to avoid this issue, it is common to introduce small perturbations at each increment, such as in evolutionary algorithms (i.e. simulated annealing and genetic algorithm). It has been shown that IO yields results that are nearly as good as evolutionary algorithms while using a fraction of the computational power (Nickless et al. (2018)). For GeoCarb's application, we were looking at the

10 global distribution of errors, σ, and therefore were not concerned about local optima. An additional constraint was added that requires the algorithm to cover South America before switching to North America to further prevent erratic scanning behavior. The pseudo-code procedure of the algorithm is shown in Algorithm 2.

Algorithm 2 Incremental Optimization Algorithm

    $\mathbf{E} \leftarrow$ North American and South American land masses between $50°N$ and $50°S$

    $\mathbf{S} \leftarrow$ Set of 5-minute east-to-west scan blocks where $E \subseteq \bigcup s_i \in \mathbf{S}$

    $\mathbf{I} \leftarrow \emptyset$

    $\mathbf{C} \leftarrow$ Objective Function

    while $\mathbf{E} \subsetneq \mathbf{I}$ **do**

        Find $s^* \in \mathbf{S}$ such that $\mathbf{C}(s^*)$ is the minimum of the set $\{\mathbf{C}(s_i) : s_i \in \mathbf{S}\}$

        Append s^* to $\mathbf{I}$

        Remove s^* from $\mathbf{S}$

    end while

3.6 Evaluating the Optimized Scanning Strategy

To determine a "best" algorithm-selected scanning strategy, the global distribution of error, σ (Eq. (3)), of the algorithm-selected

15 strategy is compared to a baseline scanning strategy that we considered the "obvious" choice if chosen by a human, shown in Fig. 4. We say the baseline strategy is the "obvious" choice because it tracks the sun's path and covers the entire area of interest in five coherent regions in the order of Tropical South America East, Tropical South America West, Temperate South America, Tropical North America, and Temperate North America. For scanning start and stop times of the baseline strategy, the same times used by the IO algorithm are used for the baseline strategy (1230 UTC for the Autumn Equinox and 1315 UTC for the

20 Summer Solstice).

Atmospheric
Measurement
Techniques
Discussions
Open Access
EGU

3.7 Parameter Exploration

In Equation (5), the overlap and distance terms have equal weighting in the objective function. The effects of weighting these terms on the global distribution of errors were investigated by adding (w_o, w_d) constant weight terms to Eq. (5) as new input parameters resulting in Eq. (6). Utilizing Equation 6, the algorithm now has three inputs, w_o, w_d, and the starting airmass factor

5   threshold. A Monte Carlo experiment was performed to explore the parameter space of weights and determine the distribution of sample error statistics across the range of possible starting thresholds. For both w_o and w_d, 1000 weights each were randomly sampled from a uniform distribution between 0 and 10. This process was repeated for starting AF thresholds starting from 2.5 increasing by 0.1 to 3.5 and found that the minimum variance of predicted errors occurs when the starting AF threshold is 2.6 for the Summer Solstice and 2.7 for the Autumn Equinox, shown in Fig. 5. Figure 6 shows that the weight values that

10   correspond to minimum error distribution medians and variance are when w_o, w_d are both equal to one.

$$\mathrm{c}(s,t,w_o,w_d) = \mathrm{median}(e^{AF(s,t,x,y)}\alpha(x,y)^{-1})(1 + \frac{w_o(s \cap \mathbb{I}) + w_d\delta^2}{s \cap \mathbb{E}}). \qquad (6)$$

4   Results

4.1   Global error

Based on the parameter exploration results, the global distributions of error were investigated for simulations during the Sum-

15   mer Solstice with a starting AF threshold of 2.6 and simulations during the Autumn Equinox with a starting AF threshold of 2.7. The algorithm-selected scanning strategies consistently matched or exceeded the overall performance of the baseline scanning pattern, shown in Fig. 7 and 8. The region where the most significant improvement is seen is in the Amazon during the Autumn Equinox, refer to Fig. 9. Additionally, the potential of both scanning strategies to yield observations where the SNR was greater than 100 (yielding a predicted error of $\sigma = 2.17$) was analyzed for both strategies. A comparison of the

20   distributions of medians of usable soundings obtained from the parameter exploration experiment is shown in Fig. 10. During the Summer Solstice with a starting AF threshold of 2.6, the algorithm-selected strategy yielded approximately 3.79 million usable soundings versus the baseline strategy, which yielded approximately 3.02 million usable soundings. Similarly during the Autumn Equinox with a starting AF threshold of 2.7, the algorithm-selected strategy yielded approximately 4.31 million usable soundings versus 3.04 million usable soundings from the baseline. It is important to note that these figures are within

25   the cloud-free environment of our model and we expect the yield of usable soundings to be significantly less during operations, but that will likely affect both the baseline and optimal strategies similarly.

4.2   Sensitivity Analysis

To quantify the algorithm's sensitivity to input parameters, the method of standardized regression coefficients (SRC) was utilized (Helton et al. (2006)). SRCs are the regression coefficients of a linear model fitted to the standardized dependent variable,

$Y_Z = \frac{Y - \bar{Y}}{\sigma_Y}$, using standardized independent variables, $X_Z = \frac{X - \bar{X}}{\sigma_X}$. The dependent variable in this case is the predicted error and the independent variables are w_o, w_d, and the starting AF threshold. The standardization of variables allows for measuring the effect of the input parameters without their dependency on units (i.e. km^2, ppm). The coefficient of determination, R^2, of the SRC model tells us how much of the variability in the sample statistics is explained by the model. R^2 is defined as the

5   Modeled Sum of Squares (MSS) divided by the Total Sum of Squares (TSS), where

$$\text{MSS} = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})$$

$$\text{TSS} = \sum_{i=1}^{n}(Y_i - \bar{Y})$$

$$R^2 = \frac{\text{MSS}}{\text{TSS}}$$

and $\hat{Y}$ = model predicted values, $\bar{Y}$ = mean Error, Y = observed values, n = number of observations. The method of SRC was

10   chosen for the sensitivity analysis by convenience of readily available simulation data from the parameter exploration experiment.

In the results, both the median and variance of the global error are found to be sensitive to starting AF thresholds as seen in Fig. 5 and 10, and Tables 1 and 2. The starting AF thresholds affect the scanning strategy as a whole by shifting the scanning time frame. This sensitivity was expected considering that airmass factors rely on time and play a large role in the calculation

15   of radiance (Eq. (1)). Because SRCs determine the effect of the input parameters in the presence of others, the SRCs fitted to a linear model of predicted error with respect to w_o and w_d were also analyzed within the Monte Carlo samples of starting AF threshold equal to 2.7 for the Autumn Equinox and starting AF threshold equal to 2.6 for the Summer Solstice.

Within the specified starting AF threshold of 2.7 for the Autumn Equinox, moderate effects of the weights were found on the sample global error distribution. The values in Table 3 show that the SRC model explains approximately half of the variability

20   in median of global error distributions, $R^2 = 0.552$, and the parameter with the largest effect on the variance is the distance, δ. With respect to variance of global error distributions, the SRC model explains less than half of the variability with $R^2 = 0.384$. Again, the parameter with the largest effect is the distance term.

Within the specified starting AF threshold of 2.6 for the Summer Solstice, the effects of the weights on the sample global error distribution are small. The SRC model explains approximately a quarter of the variability in median of global error

25   distributions, $R^2 = 0.242$, and approximately 15% of the variability with $R^2 = 0.148$, shown in Table 4. The parameter with the largest effect is the overlap term for both variance and median of error distributions.

Small R^2 values signify little sensitivity to the independent variables or a nonlinear relationship between the independent and dependent variables. Visual analysis of the scatter plots of the distributions of sample statistics versus weights (Fig. 6) does not imply a nonlinear relationship between the weights and sample statistics.

Atmospheric
Measurement
Techniques
Discussions

5 Conclusions

We illustrate an efficient technique that selects a covering set that also minimizes global measurement error. The Incremental Optimization routine gives us an optimized scanning strategy that performs better than the baseline scanning strategy relative to the global distribution of error and number of usable soundings. We also found that by optimizing for the global distribution

5 of error, we obtained an improvement in regional errors as well, seen in Fig. 8. There may be better objective functions to optimize, and the structure of the algorithm does allow human intervention beyond the scope of this work. For example, in the event of natural phenomena such as wildfires, droughts, and volcanic eruptions, the algorithm can be modified to capture prioritized scanning regions during the minimum predicted error time for those regions.

At the moment, our model does not take into account the effect of clouds on retrieval quality. It is known that clouds play

10 a significant role in scattering effects and influences the calculation of radiance (Eq. (1)). In a case study by Polonsky et al. (2014) that included clouds and aerosols in the atmosphere, they found that the number of usable soundings passing their post-processing filter (PPF) of AOD < 0.1 was between 8.1% to 20% of total simulated soundings. We believe that an AOD threshold of 0.1 is too conservative for a clear sky AOD threshold, therefore it was relaxed to 0.3 as previously done by Rayner et al. (2014) and O'Dell et al. (2012). In O'Dell et al. (2012), they found that 22% of scenes were classified correctly as

15 "clear" when they used an AOD threshold of 0.3. Because $\tau = 0.3$ in our calculation of radiance (Eq. (1)), we estimate that the true number of usable soundings will be around 20% of our original estimates of daily usable soundings in Sect. 4.1. Going forward, cloud products from CALIPSO will be incorporated to better simulate operational conditions. This will yield more robust scanning strategies and estimates of usable soundings.

The SNR-Optimized scanning strategy outperforms the human-selected strategy first proposed for the GeoCarb scientific

20 observation plan. The algorithm selects strategies that consistently match or exceed the performance of the baseline scanning pattern with respect to global error. The algorithm-selected strategies yield an 18% increase of soundings with a SNR>100 during the Summer and a 41% increase during the Autumn over the baseline strategy.

Author contributions. SC conceptualized the goals of this project. JN and SC developed the methodology together. JN developed the algorithm code, model code, and performed the simulations.

References

Crisp, D., Atlas, R., Breon, F.-M., Brown, L., Burrows, J., Ciais, P., Connor, B., Doney, S., Fung, I., Jacob, D., Miller, C., O'Brien, D., Pawson, S., Randerson, J., Rayner, P., Salawitch, R., Sander, S., Sen, B., Stephens, G., Tans, P., Toon, G., Wennberg, P., Wofsy, S., Yung, Y., Kuang, Z., Chudasama, B., Sprague, G., Weiss, B., Pollock, R., Kenyon, D., and Schroll, S.: The Orbiting Carbon Observatory (OCO) mission, Advances in Space Research, 34, 700–709, https://doi.org/10.1016/J.ASR.2003.08.062, 2004.

Crisp, D., Miller, C. E., and DeCola, P. L.: NASA Orbiting Carbon Observatory: measuring the column averaged carbon dioxide mole fraction from space, Journal of Applied Remote Sensing, 2, 023 508, https://doi.org/10.1117/1.2898457, 2008.

Crisp, D., Pollock, H. R., Rosenberg, R., Chapsky, L., Lee, R. A. M., Oyafuso, F. A., Frankenberg, C., O'Dell, C. W., Bruegge, C. J., Doran, G. B., Eldering, A., Fisher, B. M., Fu, D., Gunson, M. R., Mandrake, L., Osterman, G. B., Schwandner, F. M., Sun, K., Taylor, T. E., Wennberg, P. O., and Wunch, D.: The on-orbit performance of the Orbiting Carbon Observatory-2 (OCO-2) instrument and its radiometrically calibrated products, Atmospheric Measurement Techniques, 10, 59–81, https://doi.org/10.5194/amt-10-59-2017, 2017.

Eldering, A., O'Dell, C. W., Wennberg, P. O., Crisp, D., Gunson, M. R., Viatte, C., Avis, C., Braverman, A., Castano, R., Chang, A., Chapsky, L., Cheng, C., Connor, B., Dang, L., Doran, G., Fisher, B., Frankenberg, C., Fu, D., Granat, R., Hobbs, J., Lee, R. A. M., Mandrake, L., McDuffie, J., Miller, C. E., Myers, V., Natraj, V., O'Brien, D., Osterman, G. B., Oyafuso, F., Payne, V. H., Pollock, H. R., Polonsky, I., Roehl, C. M., Rosenberg, R., Schwandner, F., Smyth, M., Tang, V., Taylor, T. E., To, C., Wunch, D., and Yoshimizu, J.: The Orbiting Carbon Observatory-2: first 18 months of science data products, Atmospheric Measurement Techniques, 10, 549–563, https://doi.org/10.5194/amt-10-549-2017, 2017a.

Eldering, A., Wennberg, P. O., Crisp, D., Schimel, D. S., Gunson, M. R., Chatterjee, A., Liu, J., Schwandner, F. M., Sun, Y., O'Dell, C. W., Frankenberg, C., Taylor, T., Fisher, B., Osterman, G. B., Wunch, D., Hakkarainen, J., Tamminen, J., and Weir, B.: The Orbiting Carbon Observatory-2 early science investigations of regional carbon dioxide fluxes, Science, 358, eaam5745, https://doi.org/10.1126/science.aam5745, 2017b.

Gloor, M., Fan, S.-M., Pacala, S., and Sarmiento, J.: Optimal sampling of the atmosphere for purpose of inverse modeling: A model study, Global Biogeochemical Cycles, 14, 407–428, https://doi.org/10.1029/1999GB900052, 2000.

Hammerling, D. M., Michalak, A. M., O'Dell, C., and Kawa, S. R.: Global CO $_2$ distributions over land from the Greenhouse Gases Observing Satellite (GOSAT), Geophysical Research Letters, 39, https://doi.org/10.1029/2012GL051203, 2012.

Helton, J. C., Johnson, J. D., Sallaberry, C. J., and Storlie, C. B.: Survey of sampling-based methods for uncertainty and sensitivity analysis, Reliability Engineering and System Safety, 91, 1175–1209, https://doi.org/10.1016/j.ress.2005.11.017, 2006.

Hetland, M. L.: Greed Is Good? Prove It!, pp. 139–161, Apress, Berkeley, CA, https://doi.org/10.1007/978-1-4842-0055-1_7, 2014.

Karp, R. M.: Reducibility among Combinatorial Problems, pp. 85–103, Springer US, Boston, MA, https://doi.org/10.1007/978-1-4684-2001-2_9, 1972.

Kumer, J. J. B., Rairden, R. L., Roche, A. E., Chevallier, F., Rayner, P. J., and Moore, B.: Progress in development of Tropospheric Infrared Mapping Spectrometers (TIMS): GeoCARB Greenhouse Gas (GHG) application, Proc.SPIE, 8867, 8867 – 8867 – 16, https://doi.org/10.1117/12.2022668, 2013.

Kuze, A., Suto, H., Nakajima, M., and Hamazaki, T.: Thermal and near infrared sensor for carbon observation Fourier-transform spectrometer on the Greenhouse Gases Observing Satellite for greenhouse gases monitoring, Appl. Opt., 48, 6716–6733, https://doi.org/10.1364/AO.48.006716, 2009.

Atmospheric
Measurement
Techniques
Discussions
Open Access

Lucht, W., Schaaf, C. B., and Strahler, A. H.: An algorithm for the retrieval of albedo from space using semiempirical BRDF models, IEEE Transactions on Geoscience and Remote Sensing, 38, 977–998, https://doi.org/10.1109/36.841980, 2000.

Miller, C. E., Crisp, D., DeCola, P. L., Olsen, S. C., Randerson, J. T., Michalak, A. M., Alkhaled, A., Rayner, P., Jacob, D. J., Suntharalingam, P., Jones, D. B. A., Denning, A. S., Nicholls, M. E., Doney, S. C., Pawson, S., Boesch, H., Connor, B. J., Fung, I. Y., O'Brien, D., Salawitch,

5   R. J., Sander, S. P., Sen, B., Tans, P., Toon, G. C., Wennberg, P. O., Wofsy, S. C., Yung, Y. L., and Law, R. M.: Precision requirements for space-based data, Journal of Geophysical Research: Atmospheres, 112, https://doi.org/10.1029/2006JD007659, 2007.

Nickless, A., Rayner, P. J., Erni, B., and Scholes, R. J.: Comparison of the genetic algorithm and incremental optimisation routines for a Bayesian inverse modelling based network design, Inverse Problems, 34, 055 006, 2018.

O'Brien, D. M., Polonsky, I. N., Utembe, S. R., and Rayner, P. J.: Potential of a geostationary geoCARB mission to estimate surface

10   emissions of CO_2, CH_4 and CO in a polluted urban environment: case study Shanghai, Atmospheric Measurement Techniques, 9, 4633–4654, https://doi.org/10.5194/amt-9-4633-2016, 2016.

O'Dell, C. W., Connor, B., Bösch, H., O'Brien, D., Frankenberg, C., Castano, R., Christi, M., Eldering, D., Fisher, B., Gunson, M., McDuffie, J., Miller, C. E., Natraj, V., Oyafuso, F., Polonsky, I., Smyth, M., Taylor, T., Toon, G. C., Wennberg, P. O., and Wunch, D.: The ACOS CO2retrieval algorithm-Part 1: Description and validation against synthetic observations, Atmospheric Measurement Techniques, 5, 99–

15   121, https://doi.org/10.5194/amt-5-99-2012, https://www.atmos-meas-tech.net/5/99/2012/, 2012.

Patra, P. K. and Maksyutov, S.: Incremental approach to the optimal network design for CO2 surface source inversion, Geophysical Research Letters, 29, 971–974, https://doi.org/10.1029/2001GL013943, 2002.

Polonsky, I. N., O'Brien, D. M., Kumer, J. B., O'Dell, C. W., and the geoCARB Team: Performance of a geostationary mission, geoCARB, to measure CO_2, CH_4 and CO column-averaged concentrations, Atmospheric Measurement Techniques, 7, 959–981,

20   https://doi.org/10.5194/amt-7-959-2014, 2014.

Rayner, P. J., Enting, I. G., and Trudinger, C. M.: Optimizing the CO2 observing network for constraining sources and sinks, Tellus B, 48, 433–444, https://doi.org/10.1034/j.1600-0889.1996.t01-3-00003.x, 1996.

Rayner, P. J., Utembe, S. R., and Crowell, S.: Constraining regional greenhouse gas emissions using geostationary concentration measurements: a theoretical study, Atmospheric Measurement Techniques, 7, 3285–3293, https://doi.org/10.5194/amt-7-3285-2014, 2014.

25 Schaaf, C. and Wang, Z.: MCD43C3 MODIS/Terra+Aqua BRDF/Albedo Albedo Daily L3 Global 0.05Deg CMG V006 [Data set], https://doi.org/10.5067/MODIS/MCD43C3.006, 2015.

Yokota, T., Yoshida, Y., Eguchi, N., Ota, Y., Tanaka, T., Watanabe, H., and Maksyutov, S.: Global Concentrations of CO2 and CH4 Retrieved from GOSAT: First Preliminary Results, SOLA, 5, 160–163, https://doi.org/10.2151/sola.2009-041, 2009.

Atmospheric
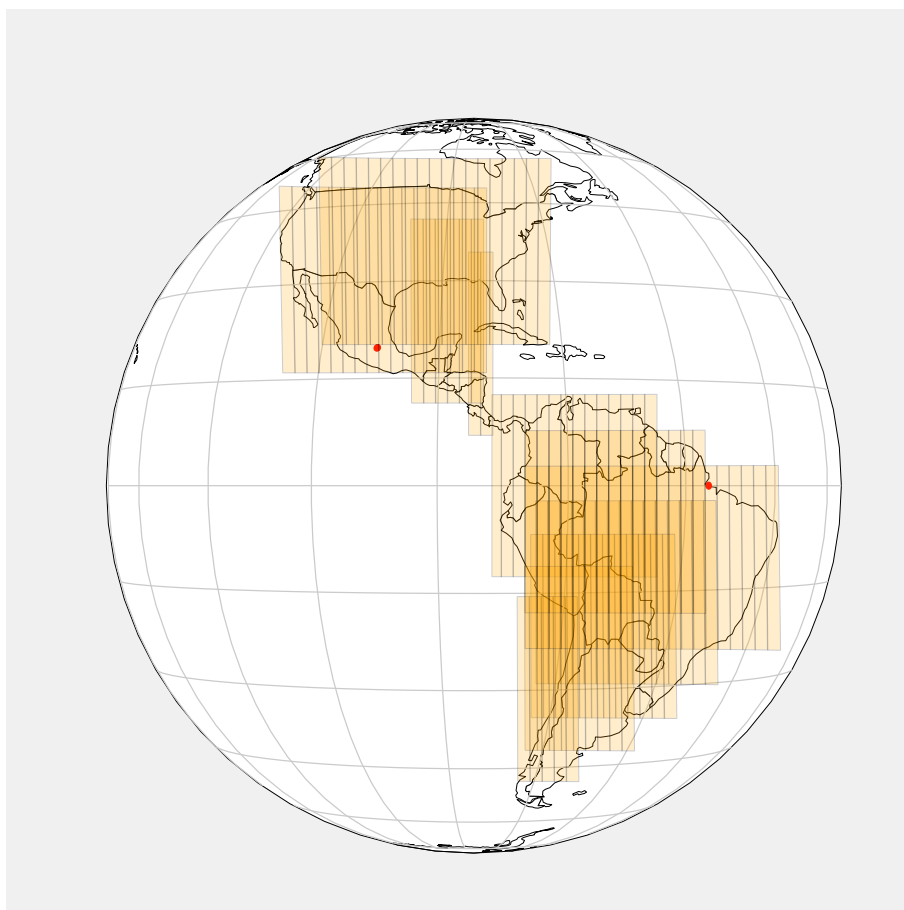Measurement
Techniques

Discussions

Figure 1. Candidate scan blocks are shown here in yellow. In red are the geographic reference points, Mexico City, Mexico and Macapá, Brazil, for starting and stopping airmass thresholds.

Standardized Regression Coefficients, Summer Solstice			
Input Parameters	Variance	Median	Expected Usable Observations
R^2	0.939	0.935	0.311
Starting Threshold	0.9679	0.9618	-0.2769
w_o	0.0615	-0.0621	-0.0634
w_d	-0.0040	-0.0716	0.4839

Table 1. The SRCs show that the starting AF threshold has major effect on variance and median of global error distributions.

Atmospheric
Measurement
Techniques
Discussions
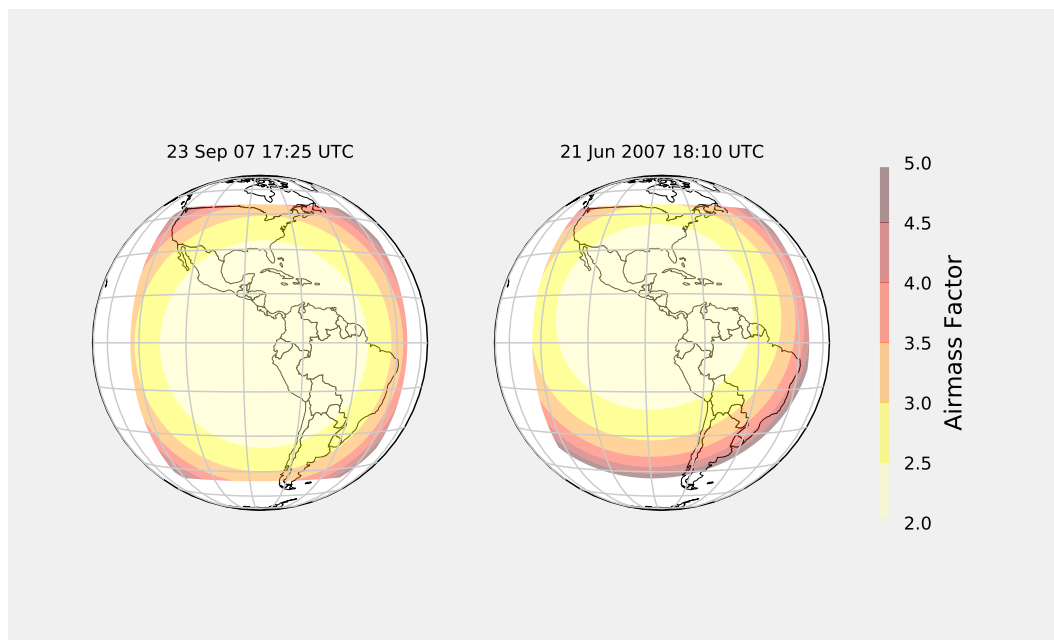
Figure 2. Comparing airmass factors in June (left) and December (right).
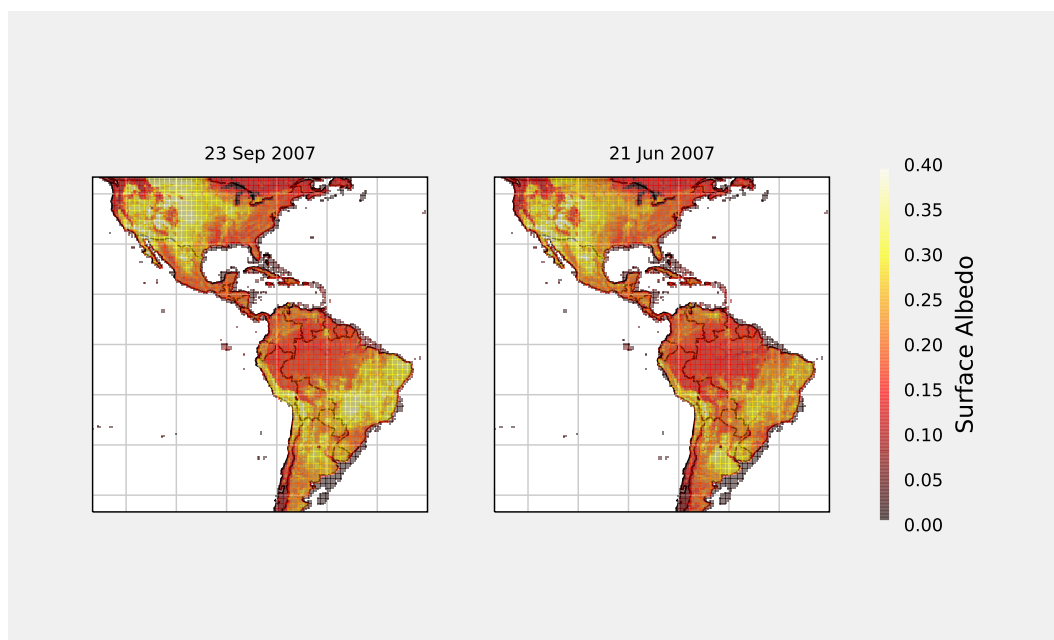
Figure 3. Comparing Surface Albedo during the Autumn Equinox (left) and the Summer Solstice (right).
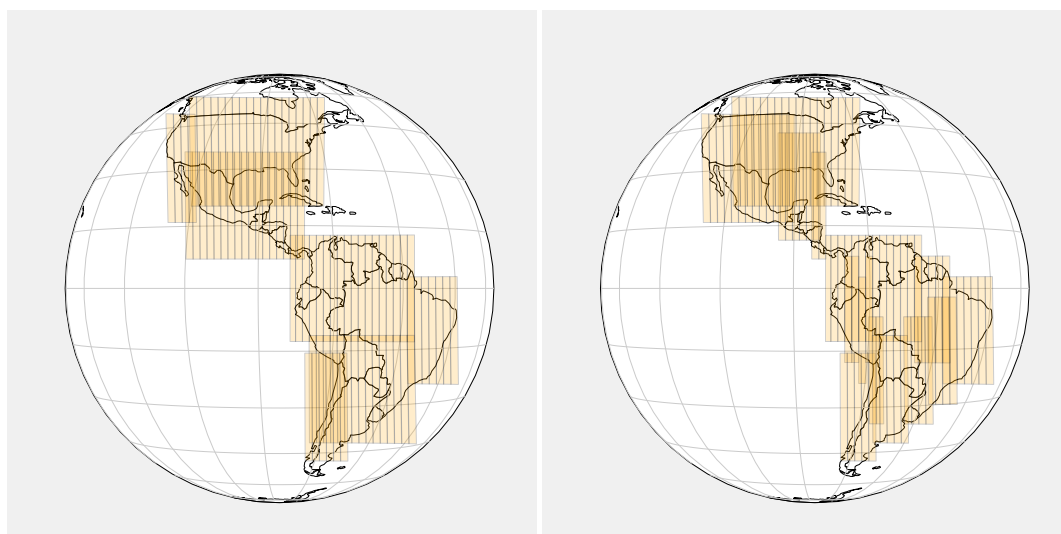
Figure 4. The "obvious" choice (left) compared to an algorithm-selected strategy (right).

Standardized Regression Coefficients, Autumn Equinox			
Input Parameters	Variance	Median	Expected Usable Observations
R^2	0.646	0.977	0.208
Starting Threshold	0.7997	0.9770	-0.4473
w_o	-0.0163	0.0056	-0.0579
w_d	0.0757	0.1455	0.0772

Table 2. The SRCs show that the starting AF threshold has major effect on median and variance of global error distributions

SRC for Starting Threshold = 2.7, Autumn Equinox			
Input Parameters	Variance	Median	Expected Usable Observations
R^2	0.384	0.552	0.497
w_o	-0.3051	0.2502	-0.2310
w_d	0.5450	0.6950	0.6702

Table 3. The SRCs show that the distance term has a moderate effect on the median of global error distributions and some effect on variance of global error distributions during the Autumn Equinox.
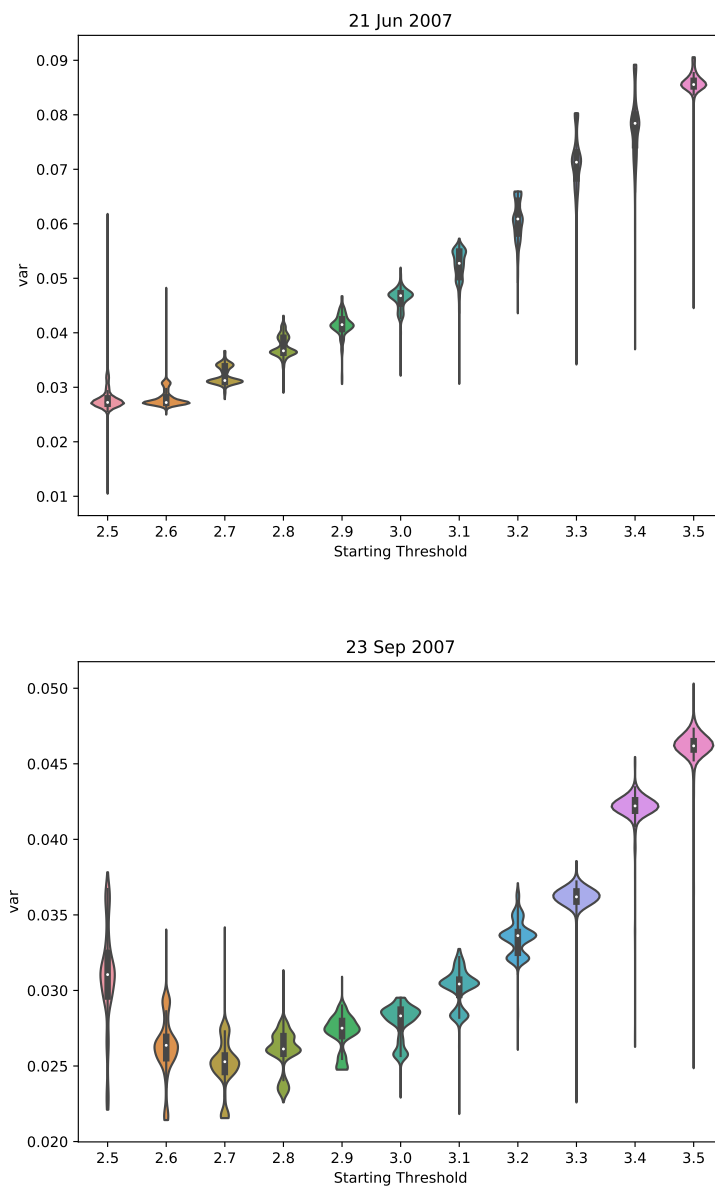
Figure 5. Violin plots show the effect of starting threshold on variance of errors: Summer Solstice (top) and Autumn Equinox (bottom).
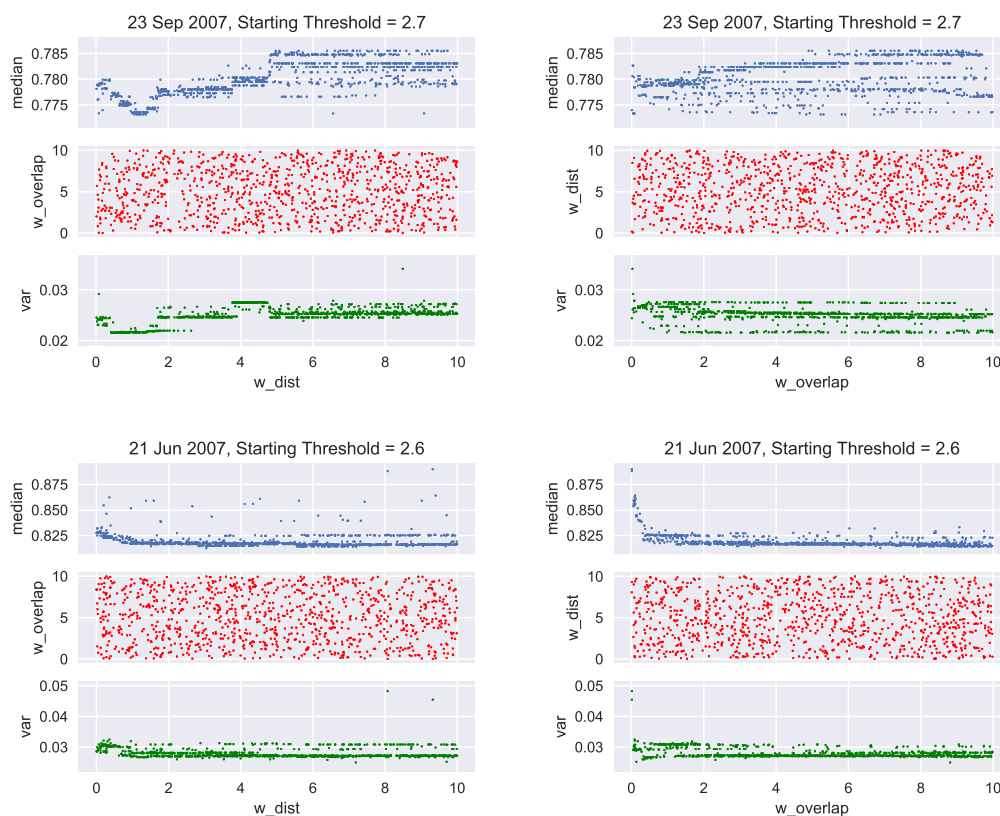
Figure 6. Scatter plots of the parameter exploration results show that equal weighting for the distance and overlap terms in the objective function results in optimal error distributions.

SRC for Starting Threshold = 2.6, Summer Solstice			
Input Parameters	Variance	Median	Expected Usable Observations
R^2	0.148	0.242	0.284
w_o	-0.3481	-0.4833	-0.3911
w_d	-0.1717	-0.1064	0.3519

Table 4. The coefficients of determination tell us that the weighting has little effect on global error distribution during the Summer Solstice.

17
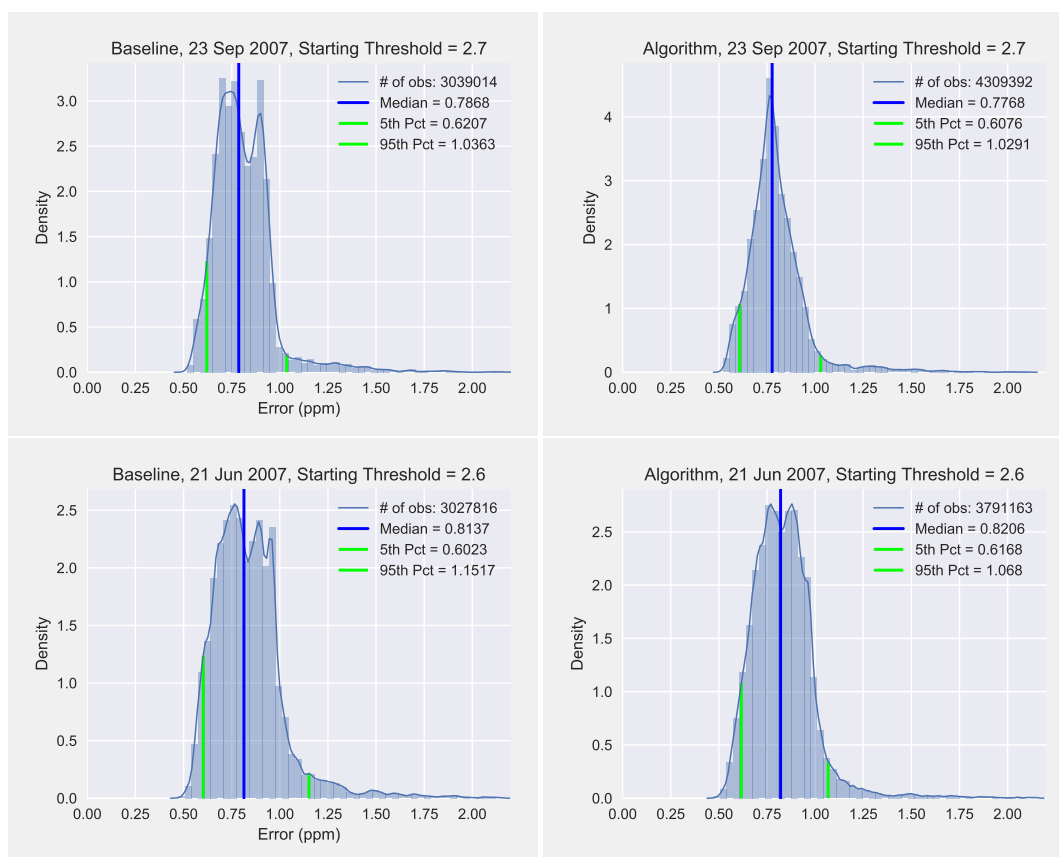
Figure 7. Global error distributions of observations with a predicted SNR > 100 of the baseline strategy (left) and algorithm-selected strategy (right) for the minimum variance starting thresholds. The number of observations is significantly greater in the algorithm-selected strategies.
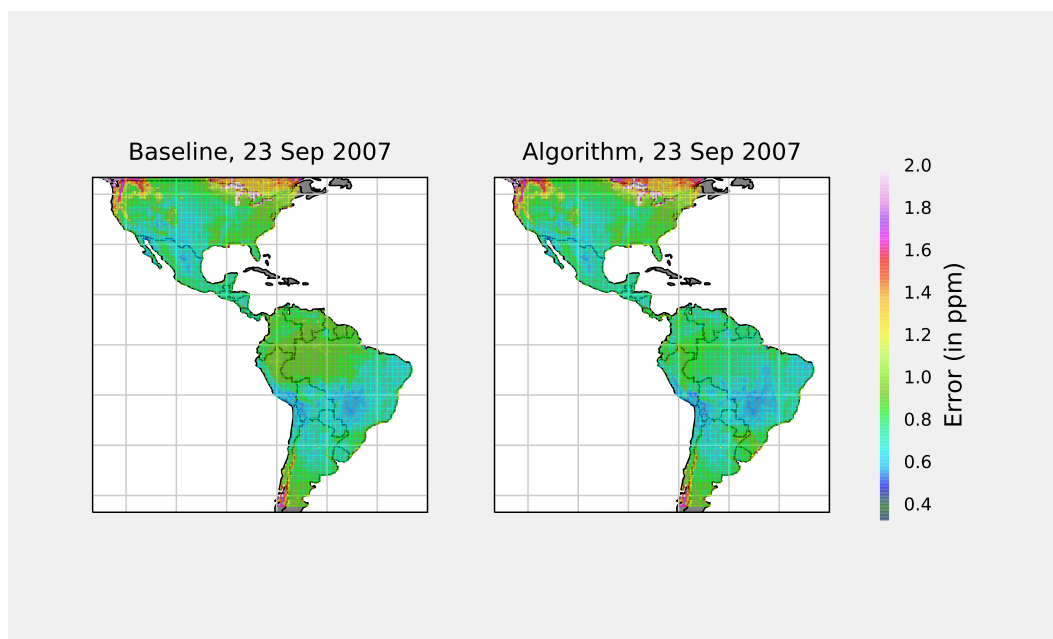
Atmospheric
Measurement
Techniques
Discussions

Figure 8. Comparing the algorithm-selected strategy (right) to the baseline strategy (left).
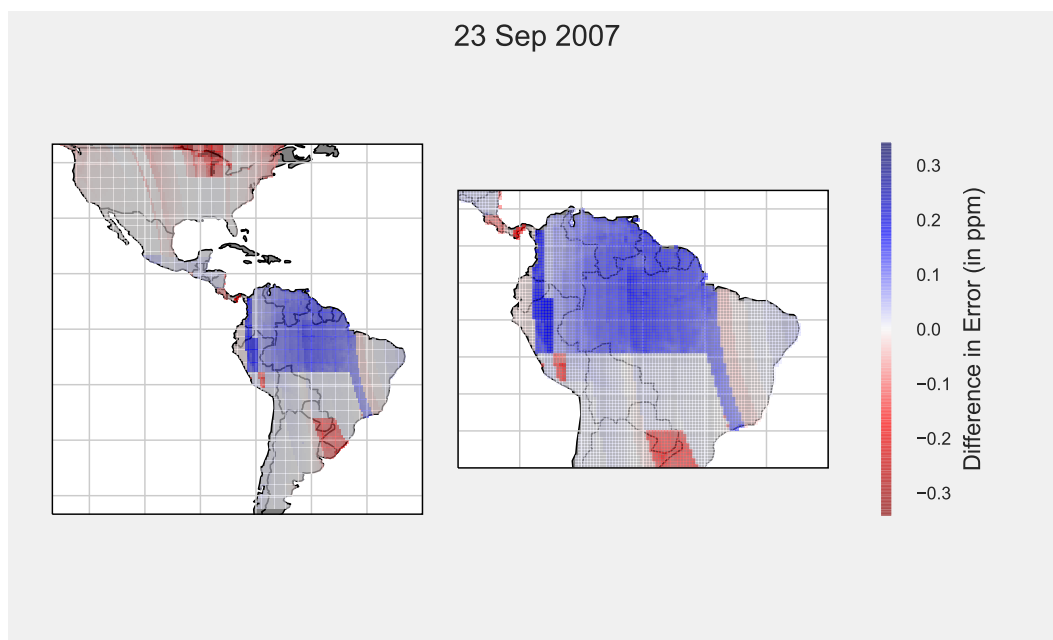

Figure 9. This plot shows the spatial distribution of baseline error minus the algorithm error for the Autumn Equinox. The blue highlights areas of improvement and the red shows areas of decreased performance. There is significant improvement in predicted errors over the Amazon.
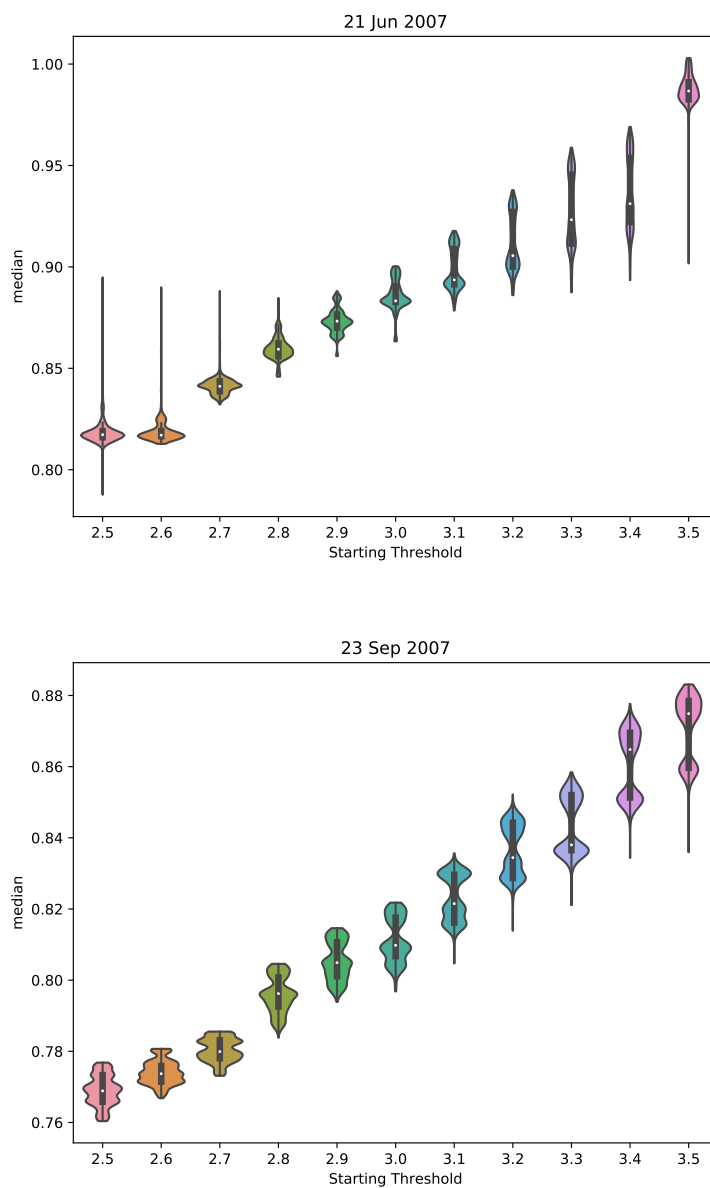
Figure 10. Violin plots show the effect of starting threshold on error distribution medians: Summer Solstice (top) and Autumn Equinox (bottom).