Orbiting Carbon Observatory–2 (OCO-2)

Warn Level, Bias Correction, and Lite File Product Description

Version 1
August 31, 2015
Data Releases 7 and 7R
Lite File Version B7.1.01

National Aeronautics and
Space Administration

JPL

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Lite Files, Warn Level and Bias Correction Determination

Lukas Mandrake, Chris O'Dell, Debra Wunch, Paul O. Wennberg, Brendan Fisher, Gregory B. Osterman, Annmarie Eldering

8-30-2015, Second Public Data Release
For use on version B7.1.01 data

Overview

This document provides the details of the content of the B7.1.01A release of the OCO-2 lite files. The document begins with a description of the file content and key fields. This is followed by details of the bias correction and warn levels for the dataset.

Two fundamental data analysis tasks that must be performed before any dataset is ready for scientific use are filtration and bias correction. The first addresses which soundings to include in an analysis, while the second attempts to remove bias in the data associated with error in the retrieval of other parts of the state vector (e.g. error in XCO_2 is associated with error in retrieval of variables associated with meteorological, surface or aerosol properties)

Traditionally, filtration is achieved by a Data Quality Flag, a Boolean measure of data acceptability. This is an easy and well-understood method, and for that reason the "Lite" file product contains such a flag (xco2_quality_flag). However, this quality flag may eliminate more or fewer soundings than desired for a particular analysis. In addition to the quality flag, the light files include a new variable named Warn Levels (WLs). WLs permit users to select the percentage of data to retain, with the "most trusted" soundings offered first. Users can sweep the filtration thresholds used in any analysis finding an appropriate filter for a particular analysis. Warn Level construction and function details are summarized below.

In addition to providing guidance on filtration, the lite files also include suggested bias correction to the raw retrievals of XCO_2. The OCO-2 validation team has evaluated bias in the Build 7.0 (B7) XCO_2 L2 product using data from November 2014 through May 2015. Bias correction construction and function details are summarized below.

Summary advice to data users

- Do not use data above warn level (WL) 15
- Above WL 12, errors well in excess of the stated a posteriori errors should be expected.
- BC was generated using November 2014 to May 2015. Extrapolation to other date ranges will decrease accuracy
- BC improves the XCO_2 estimate in a global-average sense. Specific regions may experience less improvement or perhaps even degradation compared to the raw L2 retrievals.
- Exercise caution for data with airmass > 2.5 due to a high bias (reaching up to several ppm for the ocean glint collection mode.)
- A post bias correction, per-footprint adjustment is recommended for final XCO_2 preparation (page 29.)

Format Specification

OCO-2 B7100A Lite Files

Overview Information

The OCO-2 L2 Lite files contain a subset of the information in the standard OCO-2 L2 product. They are meant to be significantly smaller but still contain all necessary information for typical science analyses. In addition, they have some value added:

- Only contain L2 soundings that converged
- They include a nominal recommended filtering flag (xco2_quality_flag)
- They contain a set of "warn levels" which allows the user to tune the level of data quality desired.
- They include the recommended bias correction already applied to XCO_2 (the XCO_2 without bias correction is also contained in the Lite files)

There is one file per day, for each day that had at least one retrieved sounding.

The OCO-2 Lite files are in the netCDF-4 format (http://www.unidata.ucar.edu/software/netcdf/docs/netcdf/NetCDF_002d4-Format.html). Because netCDF-4 is a subset of HDF-5, the files may be read with both netCDF and HDF software.

Generally speaking, each field in the file is described in the *Attributes* of that field within the file itself. Descriptions for selected fields are given below, but please be sure to read the *Attributes* of each used field within the actual Lite file.

After this, we describe the filtering and bias correction approach.

File Structure & Fields

The primary variables that most users will need exist at the main level. In addition, there are some additional variables that certain users might want, contained in three groups within the file: Preprocessors, Retrieval, and Sounding. Some NetCDF readers may not see these groups; if this happens, please update your NetCDF reader or use an HDF-5 reader.

Main Level Variables

- ***latitude*** The latitude at the center of the sounding field-of-view.
  Parent Field: L2s/RetrievalGeometry/retrieval_latitude

- ***longitude*** The longitude at the center of the sounding field-of-view.

Parent Field: L2s/RetrievalGeometry/retrieval_longitude

- ***time*** The time of the sounding in seconds since 1970-01-01 00:00:00 UTC.
  Parent Field: None (computed from RetrievalHeader/retrieval_time_string)

- ***date*** The full date and time of the sounding in UTC, organized as [year,month,day,hour,minute,second,milliseconds]. This information is redundant with that from the *time* variable.
  Parent Field: None (computed from RetrievalHeader/retrieval_time_string)

- ***solar_zenith_angle*** The solar zenith angle (in degrees) at the target.
  Parent Field: L2s/RetrievalGeometry/retrieval_solar_zenith

- ***sensor_zenith_angle*** The satellite zenith angle (in degrees) at the target.
  Parent Field: L2s/RetrievalGeometry/retrieval_zenith

- ***xco2*** The bias-corrected XCO_2 (in units of ppm). This *should be used for science analysis.* The bias correction formulae are contained in the metadata of the file.
  Parent Field: None (derived using bias-correction formula)

- ***xco2_apriori*** The prior XCO_2 assumed by the L2 retrieval, in ppm.
  *Parent Field: L2s/RetrievalResults/xco2_apriori * 1e6*

- ***xco2_uncertainty*** The posterior uncertainty in XCO_2 calculated by the L2 algorithm, in ppm. This is generally 30-50% smaller than the true retrieval uncertainty.
  *Parent Field: L2s/RetrievalResults/xco2_uncert * 1e6*

- ***xco2_quality_flag*** A simple quality flag denoting science quality data. 0=higher quality, 1=lower quality.
  Parent Field: None

- ***warn_level*** A graduated indicator of data quality. 0=highest quality; 19=lowest quality.
  Parent Field: L2s/RetrievalHeader/warn_level

- ***co2_profile_apriori*** The prior profile of co2 in ppm.
  *Parent Group: L2s/RetrievalResults/co2_profile_apriori * 1e6*

- ***xco2_averaging_kernel*** The normalized column averging kernel for the retrieved XCO_2 (dimensionless).
  Parent Group: L2s/RetrievalResults/xco2_avg_kernel

- ***pressure_levels*** The retrieval pressure level grid for each sounding in hPa. Note that is simply equal to SigmaB multiplied by the surface pressure.
  *Parent Group: L2s/RetrievalResults/vector_pressure_levels * 0.01*

- ***pressure_weight*** The pressure weighting function *on levels* used in the retrieval. It has the same dimensions as both "pressure levels" and "co2_profile_apriori".
  Parent Group: L2s/RetrievalResults/pressure_weighting_function

- ***sounding_id*** The sounding_id of the sounding. For GOSAT, this is a 14-digit number defined as YYYYMMDDhhmmss. {YYYY=year, MM=month 1-12, DD=day 1-31, hh=hour 0-23, mm=minute 0-59, ss=seconds 0-59, m=hundreds of milliseconds 0-9, f=footprint number 1-8}. For OCO, it is a 16-digit number defined as YYYYMMDDhhmmssmf. {YYYY=year, MM=month 1-12, DD=day 1-31, hh=hour 0-23, mm=minute 0-59, ss=seconds 0-59, m=hundreds of milliseconds 0-9, f=footprint number 1-8}.
  Parent Field: L2s/RetrievalHeader/sounding_id

- ***source_files*** The L2Std files used to generate this file.
  Parent Group: None

- ***file_index*** The 1-based index used to identify which source file each sounding was drawn from. Ie, file_index=2 for a particular sounding means that sounding was drawn from the 2nd element of *source_files*.
  Parent Group: None

Main-Level Dimension Variables
(these variables are part of the netcdf-4 definition and will not be needed by most users)

- ***bands:*** The three OCO-*2 bands.*
- ***levels:*** The twenty vertical levels in the OCO-2 level-2 full-physics retrieval.
- ***epoch_dimension:*** Variable used for dimensioning the 7-integer *date* variable.

Preprocessors Group

This group contains information for the two ACOS preprocessor algorithms: The O2A cloud screen algorithm, and the IMAP-DOAS algorithm.

- ***co2_ratio***: Contains the ratio of the retrieved CO2 column from the weak Co2 band relative to that from the strong CO2 band. This ratio should be near unity. Significant departure from unity is currently used as a way to flag bad soundings (usually cloud or aerosol-contaminated).
Parent Field: L2s/PreprocessingResults/co_ratio_idp.

- ***h2o_ratio***: Similar to co2_ratio, but for the water vapor column. Also an indicator of cloud or aerosol contamination.
Parent Field: L2s/PreprocessingResults/h2o_ratio_idp.

- ***dp_abp***: This is the retrieved surface pressure minus the "best-guess" surface pressure from the ECMWF forecast model. This has been adjusted for a clear-sky bias as well as the local surface elevation of the observed footprint. A value of this greater than about 10 hPa absolute value typically indicates cloud or aerosol contamination.
*Parent Field: L2s/PreprocessingResults/surface_pressure_delta_abp * 1e-2*

Retrieval Group

This group contains information from the ACOS level-2 retrieval algorithm. It contains many fields that will only briefly be summarized here. These are other fields that may be useful to users. Some were used for quality filtering, some for bias correction, and others for neither.

- ***xco2_raw*** is the "raw" XCO_2 retrieved by the L2 code *without bias correction*. It should not be used for science analysis.
*Parent Field: L2s/RetrievalResults/xco2*1e6*

- ***surface_type*** Surface type used in the retrieval: 0=ocean and corresponds to a Coxmunk+Lambertian surface; 1=land and corresponds to a pure Lambertian surface.
Parent Group: L2s/RetrievalResults/surface_type (changing Coxmunk,Lambertian->0, Lambertian->1)

- ***psurf*** Retrieved surface pressure (in hPa) from the L2 algorithm. Note: this can be multiplied with the variable *SigmaB* to determine the profile of pressures for any sounding.
Parent Group: L2s/RetrievalResults/surface_pressure_fph

- ***psurf_apriori*** A priori surface pressure (in hPa) assumed by the L2 algorithm, taken originally from the short-term ECWMF forecast. Note: this

should be used in conjunction with the variable *SigmaB_Coefficient* to determine the a priori profile of pressures.
Parent Group: L2s/RetrievalResults/surface_pressure_apriori_fph

- ***dp*** The difference *psurf – psurf_apriori*, in hPa. This variable is used in the XCO_2 bias correction over both land and water surfaces.
  Parent Group: None.

- ***albedo_1*** Retrieved Band 1 albedo.
  Parent Group: L2s/AlbedoResults/albedo_o2_fph

- ***albedo_2*** Retrieved Band 2 albedo.
  Parent Group: L2s/AlbedoResults/albedo_weak_co2_fph

- ***albedo_3*** Retrieved Band 3 albedo.
  Parent Group: L2s/AlbedoResults/albedo_strong_co2_fph

- ***albedo_slope_1*** Retrieved Band 1 albedo slope.
  Parent Group: L2s/AlbedoResults/albedo_slope_o2_fph

- ***albedo_slope_2*** Retrieved Band 2 albedo slope.
  Parent Group: L2s/AlbedoResults/albedo_slope_weak_co2_fph

- ***albedo_slope_3*** Retrieved Band 3 albedo slope.
  Parent Group: L2s/AlbedoResults/albedo_slope_strong_co2_fph

- ***aod_ice*** Retrieved Extinction Optical Depth of cloud ice at 0.755 μm.
  Parent Group: L2s/AerosolResults/aerosol_3_aod

- ***aod_water*** Retrieved Extinction Optical Depth of cloud water at 0.755 μm.
  Parent Group: L2s/AerosolResults/aerosol_4_aod

- ***aod_dust*** Retrieved Extinction Optical Depth of dust aerosol at 0.755 μm.
  Parent Group: L2s/AerosolResults/aerosol_(1 or 2)_aod, depending on aerosol_type.

- ***aod_seasalt*** Retrieved Extinction Optical Depth of sea salt aerosol at 0.755 μm.
  Parent Group: L2s/AerosolResults/aerosol_(1 or 2)_aod, depending on aerosol_type.

- ***aod_sulfate*** Retrieved Extinction Optical Depth of sulfate aerosol at 0.755 μm.
  Parent Group: L2s/AerosolResults/aerosol_(1 or 2)_aod, depending on aerosol_type.

- ***aod_bc*** Retrieved Extinction Optical Depth of black carbon at 0.755 μm.
Parent Group: L2s/AerosolResults/aerosol__(1 or 2)_aod, depending on aerosol_type.

- ***aod_oc*** Retrieved Extinction Optical Depth of organic carbon at 0.755 μm.
Parent Group: L2s/AerosolResults/aerosol__(1 or 2)_aod, depending on aerosol_type.

- ***aod_total*** Retrieved Extinction Optical Depth of cloud+aerosol at 0.755 μm.
Parent Group: L2s/AerosolResults/aerosol_total_aod

- ***logDWS*** Given by *max(-5.0, ln{aod_dust + aod_water + aod_seasalt})*. This is used in the XCO_2 bias correction for soundings over land.

- ***deltaT*** Retrieved offset (in Kelvin) to a priori temperature profile.
Parent Group: L2s/RetrievalResults/temperature_offset_fph

- ***co2_grad_del*** Change (between the retrieved profile and the prior profile) of the co2 dry air mole fraction difference from the surface minus that at level 13, measured in ppm. Level 13 is at a pressure P = 0.631579 Psurf. This variable is used in the XCO_2 bias correction over both land and water surfaces.
Derived from L2s/RetrievalResults/co2_profile and co2_profile_apriori.
If c=co2_profile and a=co2_profile_apriori, then
*delta_grad_co2 = (c[20]-c[13])*1e6 – (a[20]-a[13])*1e6*

- ***fs*** Retrieved fluorescence in units of W m^{-2} μm^{-1} sr^{-1} at 757 nm.
Derived from: L2s/RetrievalResults/fluorescence_at_reference.

- ***grad_co2*** The co2 dry air mole fraction difference from the surface minus that at level 13, measured in ppm. Level 13 is at a pressure P = 0.631579 Psurf. *Derived from L2s/RetrievalResults/co2_profile. If c=co2_profile, then grad_co2 = (c[20]-c[13])*1e6.*

- ***h2o_scale*** Retrieved water vapor scale factor.
Parent Group: L2s/RetrievalResults/h2o_scale_factor

- ***reduced_chi_squared_per_band*** Reduced chi-squared value of the L2 fit residuals for each band.
Parent Group: L2s/SpectralParameters/reduced_chi_squared_{band}_fph
Where {band} = {o2, weak_co2, strong_co2}

- ***windspeed*** Retrieved surface wind speed (in m/s) from the L2 algorithm, over water surfaces only.

Parent Group: L2s/RetrievalResults/wind_speed

- ***windspeed_apriori*** Prior surface wind speed (in m/s) used in the L2 algorithm, over water surfaces only. Taken from the ECMWF forecast.
  Parent Group: L2s/RetrievalResults/wind_speed_apriori

- ***T700*** The estimated temperature at 700 hPa for each sounding, taken from the ECMWF forecast and estimated using linear interpolation. Invalid for soundings in which the surface pressure is lower than 700 hPa.
  Parent Group: None.

- ***s32*** Ratio of continuum signal in the strong co2 band to that of the weak co2 band.
  Derived from: L2s/SpectralParameters/signal_strong_co2_fph
  (divided by)
  L2s/SpectralParameters/signal_weak_co2_fph

- ***SigmaB*** The coefficients by which to multiply *psurf* to determine the pressure at each vertical level in the profile. Note this is a single list of numbers, rather than repeated for each sounding (as all the other quantities are).

Sounding Group

Contains all the geolocation and time information for each sounding. For GOSAT, these contain the corrected geolocation information.

- ***altitude*** The mean surface elevation, in meters, in the target field of view.
  Parent Field: L2s/RetrievalGeometry/retrieval_altitude

- ***orbit*** is the orbit number within the current repeat cycle.
  Parent Field: L2s/Metadata/StartOrbitNumber

- ***path*** The WRS path number of the current orbit. 1-233.
  Parent Field: L2s/Metadata/StartPathNumber

- ***footprint*** The footprint number of each sounding (1-8).
  Parent Field: L2s/RetrievalHeader/sounding_index

- ***land_water_indicator*** determines the land surface properties at the field of view. 0: land; 1: water; 2: inland water; 3: mixed. This field is no longer used in the L2 processing to select the surface type.
  Parent Field: L2s/RetrievalGeometry/retrieval_land_water_indicator

- ***land_fraction*** The fraction of land contained in the field-of-view for each sounding. This field is now used in the L2 processing to select the surface

type. When this fraction is greater than 80 percent, it is assigned land surface. When below 20%, it is assigned water. No retrieval is performed for cases between 20 percent and 80 percent. This field can be used in combination with the operation_mode to find land nadir, land glint, and sea glint data.
Parent Field: L2s/RetrievalGeometry/retrieval_land_fraction.

- ***l1b_type*** gives the version number of the input level-1B data. (e.g. "5000", which means B5.0.00)
  Parent Field: Derived from parent file name.

- ***operation_mode*** Nadir (0), Glint (1), Target (2), or Transition (3). This field can be used in combination with land_fraction to find land nadir, land glint, and sea glint data.
-
  Parent Field: Derived from Metadata/OperationMode

- ***solar_azimuth_angle*** The solar azimuth angle (in degrees) at the target, measured east of north from the point of view of an observer on the ground at the target.
  Parent Field: L2s/RetrievalGeometry/retrieval_solar_azimuth

- ***sensor_azimuth_angle*** The satellite azimuth angle (in degrees) at the target, measured east of north from the point of view of an observer on the ground at the target.
  Parent Field: L2s/RetrievalGeometry/retrieval_sensor_azimuth

- ***glint_angle*** The angle (in degrees) between the local direction to the sensor, and the direction for pure glint observation (i.e., the outgoing direction from the surface for specularly reflected solar rays).

- ***airmass*** The relative airmass of the sounding, defined as 1/cos(solar_zenith_angle) + 1/cos(sensor_zenith_angle).

- ***snr_o2*** The estimated signal-to-noise ratio in the continuum of the O2A-band.
  Parent Field: L2s/L1bScSpectralParameters/snr_o2_l1b

- ***snr_weak_co2*** The estimated signal-to-noise ratio in the continuum of the weak CO_2 band.
  Parent Field: L2s/L1bScSpectralParameters/snr_weak_co2_l1b

- ***snr_strong_co2*** The estimated signal-to-noise ratio in the continuum of the strong CO_2 band.
  Parent Field: L2s/L1bScSpectralParameters/snr_strong_co2_l1b

Bias Correction

The main-level XCO_2 data has also been adjusted using a linear bias correction scheme, similar to the approach described in Wunch et al. (2011). Similar to deriving quality flags, we use different XCO_2 truth proxies to allow us to identify and remove XCO_2 biases. In this data release, a number of truth proxies have been used to derive and validate the XCO_2 bias correction parameters, including

- The Southern Hemisphere approximation (see Wunch et al., 2011).
- TCCON (as seen in nadir, glint and target modes)
- The "Small Area Analysis", in which XCO_2 is assumed to be constant for observations taken over distances <~ 100 km within the same orbit.

Note that the specific bias correction parameters employed in this release are given in the global metadata of each lite file. (For example, Bias_Correction_land_ND, with the corresponding footprint-specific corrections given in Footprint_bias_land_ND).

Quality Filtering

There are two methods one can use to quality filter the soundings contained in the Lite data files. The simplest method is to use **xco2_quality_flag**, which is simply a byte array of 0s and 1s. This filter has been derived by comparing retrieved XCO_2 for a subset of the data to various truth proxies, and identifying thresholds for different variables that correlate with poor data quality. It begins with a basic requirement that the warn level (see below) is less than or equal to 15, and applies a number of additional quality filters on top of this, using variables not contained in the warn level definition. These filters are described in more detail in Table 1 below.

The alternative method is new for OCO-2 and is called the "warn level" approach. This is similar to the above method but is a graduated method that is more nuanced than a simple "yes/no" approach. In this approach, each sounding is given a **warn_level**, which may take a value from 0-19. 0 is considered the highest quality, while 19 is the lowest quality. There are roughly equal numbers of soundings in each warn level, so using warn levels 0-9 would mean keeping half the data, etc. If you wish to use this approach for your application, we recommend you study how your results depend on the warn level you used for quality filtering. For more details on this approach, see Mandrake et al. (2013).

Table 1: Quality Filters Applied to Land Soundings

All Land Soundings		
Field	**Lower Limit** (> or =)	**Upper Limit** (< or =)
Warn level	N/A	15
Outcome flag (not in lite file)	N/A	2
Preprocessors/h2o_ratio	0.700	1.030
Preprocessors/co2_ratio	0.995	1.025
Preprocessors/dp_apb	-15.00	5.00
Retrieval/dp	-5.00	10.0
Retrieval/aod_ice	N/A	0.050
Retrieval/Aod_sulfate	N/A	0.400
Retrieval/Aod_dust*	0.001	0.30
Retrieval/Co2_grad_del	-70.0	70.0
Retrieval/albedo_2	0.10	N/A
Blended albedo (2.4*albedo_3 – 1.13*albedo_1) (both in retrieval group)	N/A	0.8
dof_co2 (not in lite product)	1.8	N/A
Sounding/airmass	N/A	3.6
* or AOD dust = 0.0		

Table 2: Quality Filters Applied to Ocean Glint Soundings

Ocean Glint Soundings		
Field	**Lower Limit** (> or =)	**Upper Limit** (< or =)
Warn level	N/A	15
Outcome flag (not in lite file)	N/A	2
Preprocessors/co2_ratio	0.994	1.020
Preprocessors/dp_apb	N/A	0.00
Retrieval/dp	-3.00	9.0
Retrieval/Co2_grad_del	-30.0	5.0
Retrieval/albedo_slope_3•10^5	1.0	10.0
Retrieval/windspeed	2.0	N/a
Sounding/snr_weak_co2	380	N/A
Sounding/airmass	N/A	3.5

References

Mandrake, L., C. Frankenberg, C.W. O'Dell, G. Osterman, P. Wennberg, and D. Wunch, 2013: Semi-autonomous sounding selection for OCO-2. Atmos. Meas. Tech., 6, 2851-2864.

D. Wunch, P. O. Wennberg, et al., 2011: A method for evaluating bias in global measurements of CO_2 total columns from space. *Atmos. Chem. Phys.,* 11 (23), 12317-12337.

Warn Levels

Warn Levels range from 0 (including only the very best data) to 19 (including all the data), and are formulated using a set of filters that minimize the mean monthly standard deviation (MMS) of XCO_2 in locations where the atmospheric variability of XCO_2 is small. We use two test sets: the Southern Hemisphere Approximation (SHA) set (assumes that the entire region from -25 to -60 latitude has minimal variation in XCO_2 (Wunch et al., 2011)) and the Small Area Analysis (SAA) set (assumes that all soundings within a 0.89 degrees latitude span and identical orbit track have the same XCO_2).

The creation of WLs relies on a hyper-dimensional filter and a genetic algorithm that optimizes filter performance for every 0.1% increment of data accepted (transparency). Mandrake *et al*. (2013) describes this technique in detail. For OCO-2 v7 data, the WLs are calculated separately for each measurement mode (nadir land, glint land, glint water, and land target).

Table 1 reports the features derived from the L2 and preprocessing output that were commonly correlated with XCO_2 variability in the two test sets. Many other features could be used, but we have opted for the simplest filter.

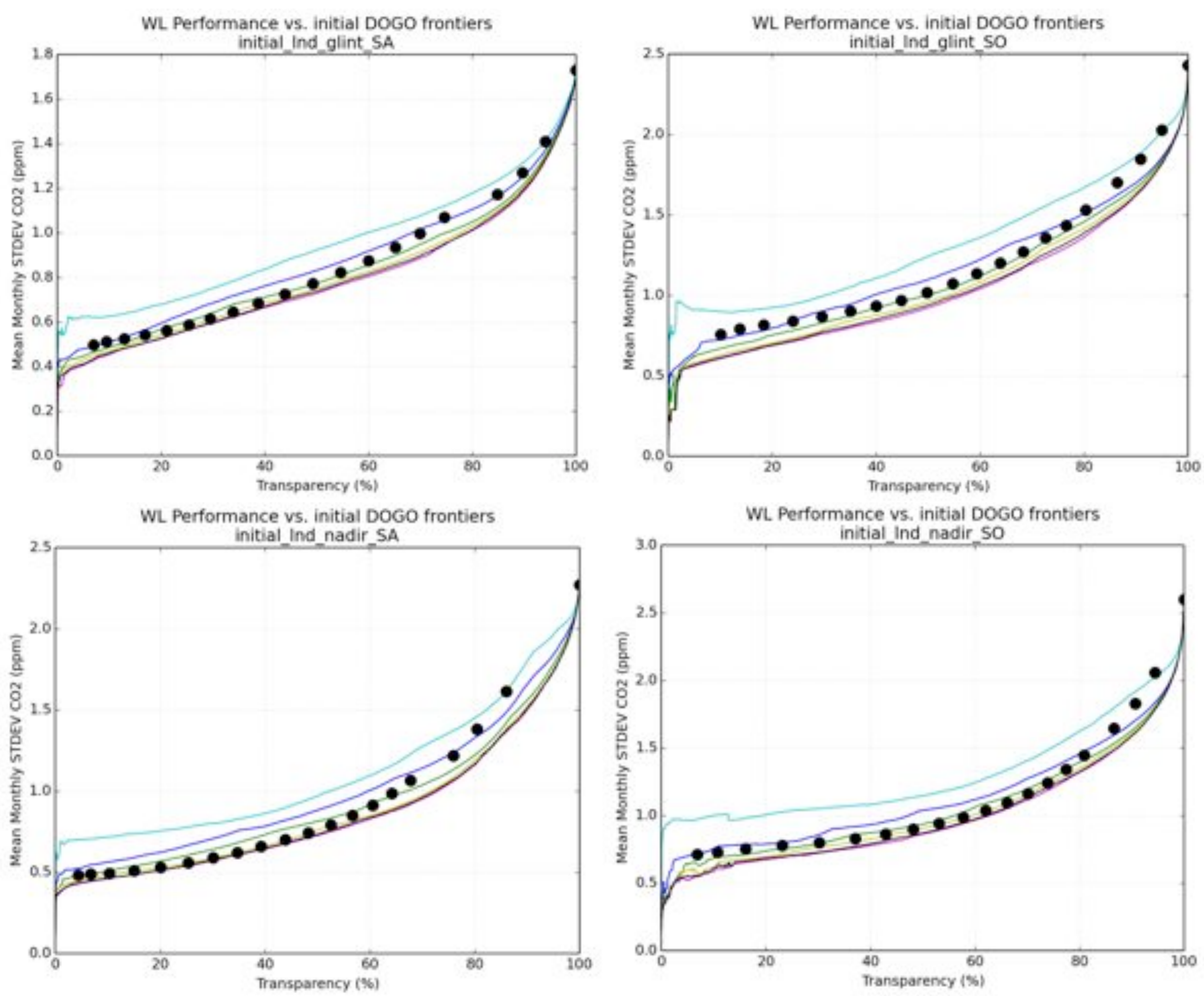
Table 1 – Features selected for Warn Level Development

Feature	Dataset
Aerotol_total_aod	Sea Glint & Land Target
Surface_pressure_delta_abp	Land Nadir, Glint, Target
Retrieval_surface_roughness	Land Nadir, Glint, Target
Relative_residual_mean_square_SCO2	Land Target
Relative_residual_mean_square_WCO2	Land Nadir, Glint
albedo_slope_strong_co2	Sea Glint
co2_vertical_gradient_delta	Sea Glint

These variables were used to design WLs definition. An example of the tradeoff of the merit function (MMS) and transparency is shown below.

In all retrievals in the test sets over land, the same features were selected (dP, surface roughness, rms_sco2, etc.). Glint soundings over water depended on a significantly different feature set. This is not surprising, as the difference in surface type requires different retrieval procedures. Land Target has very small quantities of data available, which prevents a successful definition of WL 0-8. Thus, WL's from 0-8 should be grouped together and treated as an undifferentiated pool of equal quality data.

The final WL definitions sub-select the data into 20 partitions, ordered by increasing MMS in both the SHA and SSA test sets. These numbers are reported in Tables A1-A3 in the appendix. The performance of the WL filters is compared to the unrestricted Pareto Frontiers [Mandrake et al., 2013] and shown to be comparable. Thus, the information of data quality vs transparency has been well-represented and preserved with only this small set of features. Figure 1 shows the performance of each WL (black dots, WL = 0 on left, WL = 19 on right) to admit increasing percentages of data while minimizing the increase of MMS until 100% / the unfiltered MMS is reached. There are many curves because of the mixtures of data mode and analysis type.



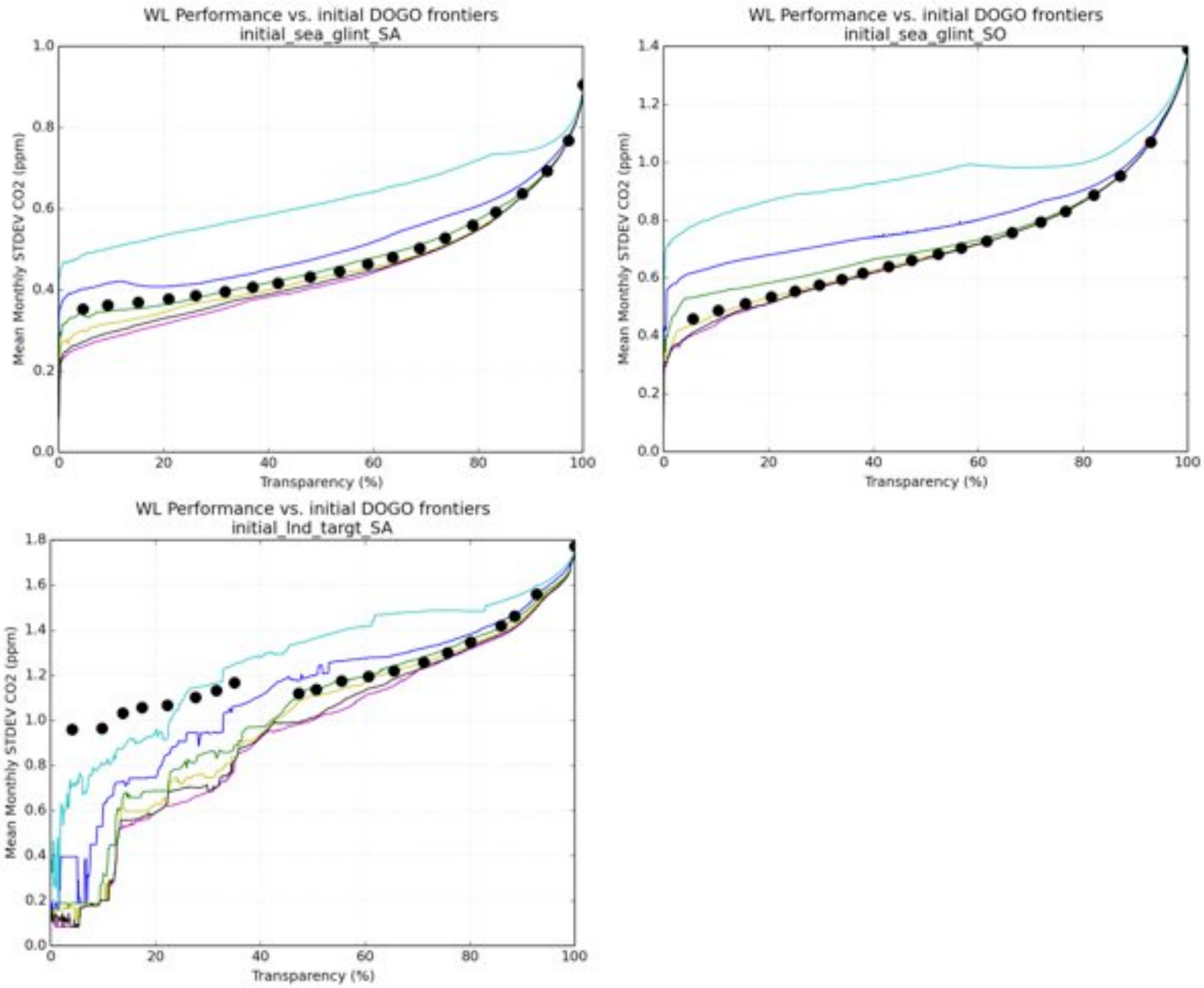


Figure 1. Warn Level performance curves for each data mode. In the title SA corresponds to Small Area Analysis test set while SO corresponds to South Hemisphere Approximation test set. The Land Target WLs below ~8 are ill-defined and should be treated as a single block.

Tables A1-A3 in Appendix A show the specific values that have been used to define the WLs for v7. These WLs have already been calculated and inserted into the v7 lite files; users do not need to calculate WLs directly unless they are using the v7 L2Std files where the WL variable is assigned a fill value.

Characterizing Warn Level Behavior

Warn Levels and XCO_2 characteristics

Warn Levels are designed to minimize the variance of XCO_2 in regions of small atmospheric variability. However, the signal that WL's reduce is so clearly evident that even in locations with significant, physical atmospheric variability or entire global datasets demonstrate the reduction of XCO_2 variance with respect to WL filtration. In figure 2, we show the naive mean and STD of the global and southern hemisphere v7 sea glint dataset as a function of inclusive WL (WL ≤ #). Warn Levels have roughly partitioned the global dataset into three regions, shown in green, orange, and red. The green area, defined by 0 ≤ WL ≤ ~4, demonstrates similar (and small) relative bias and low global scatter (STD). The orange area, defined by ~5 ≤ WL ≤ ~15, demonstrates a higher and less stable bias, with SH and global means diverging in behavior, and a strongly increasing trend in overall XCO_2 scatter for both. Finally, the red area, defined by ~16 ≤ WL ≤ 19, shows complex bias behavior and relatively large XCO_2 scatter in both Global and SH XCO_2. This regime is populated by the hopelessly confounded soundings such as clouds that escaped the cloud filter, very rough ground with complex reflectance, very dim scenes with low signal-to-noise-ratio, high aerosol optical depths, and other scenes not well modeled by the retrieval code.

These three regions are simply hand-drawn for example and discussion; an individual user should create and examine these and other statistics to decide which WLs are most appropriate for their own investigation.

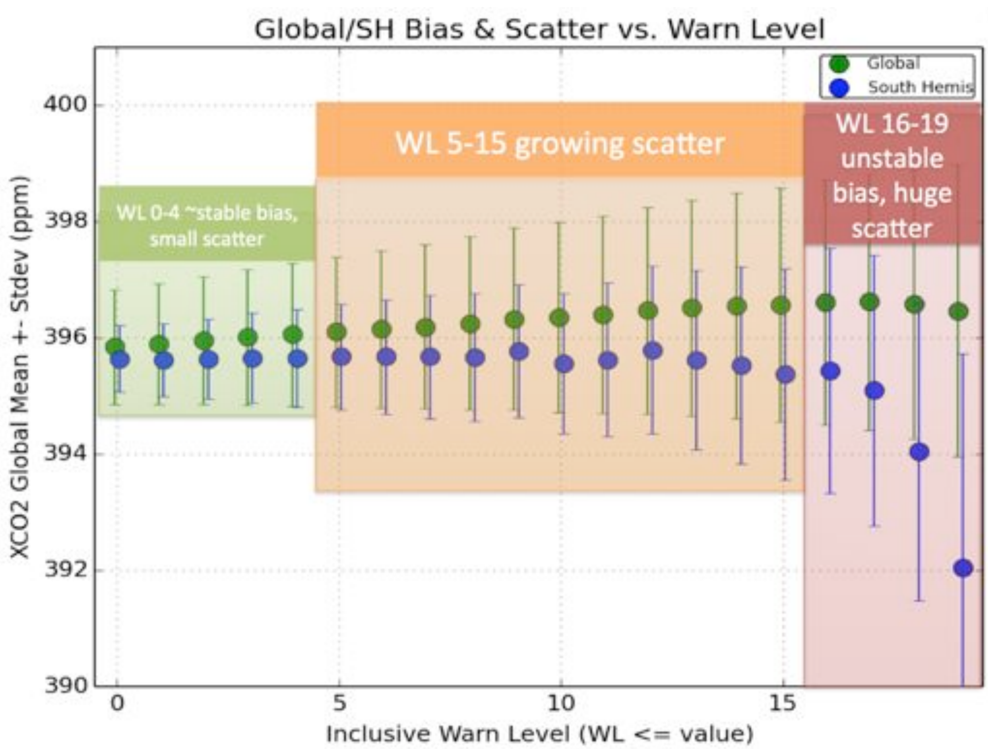


Figure 2 – XCO_2 Bias and Scatter as a function of Warn Level

Warn Level Normality & Minority Outliers

Each WL is designed to reduce the STD(XCO_2) of the included dataset and retain roughly 5% of the total data record per WL included. The resulting distribution of data within each WL bin is found to be that of a normal distribution plus a fat tail at high XCO_2 for land modes and at low XCO_2 for sea modes. Figure 3 shows the histogram of XCO_2 vs WL filtration for the three data modes and several levels of WL filtration (<=5, 10, 15, 19).

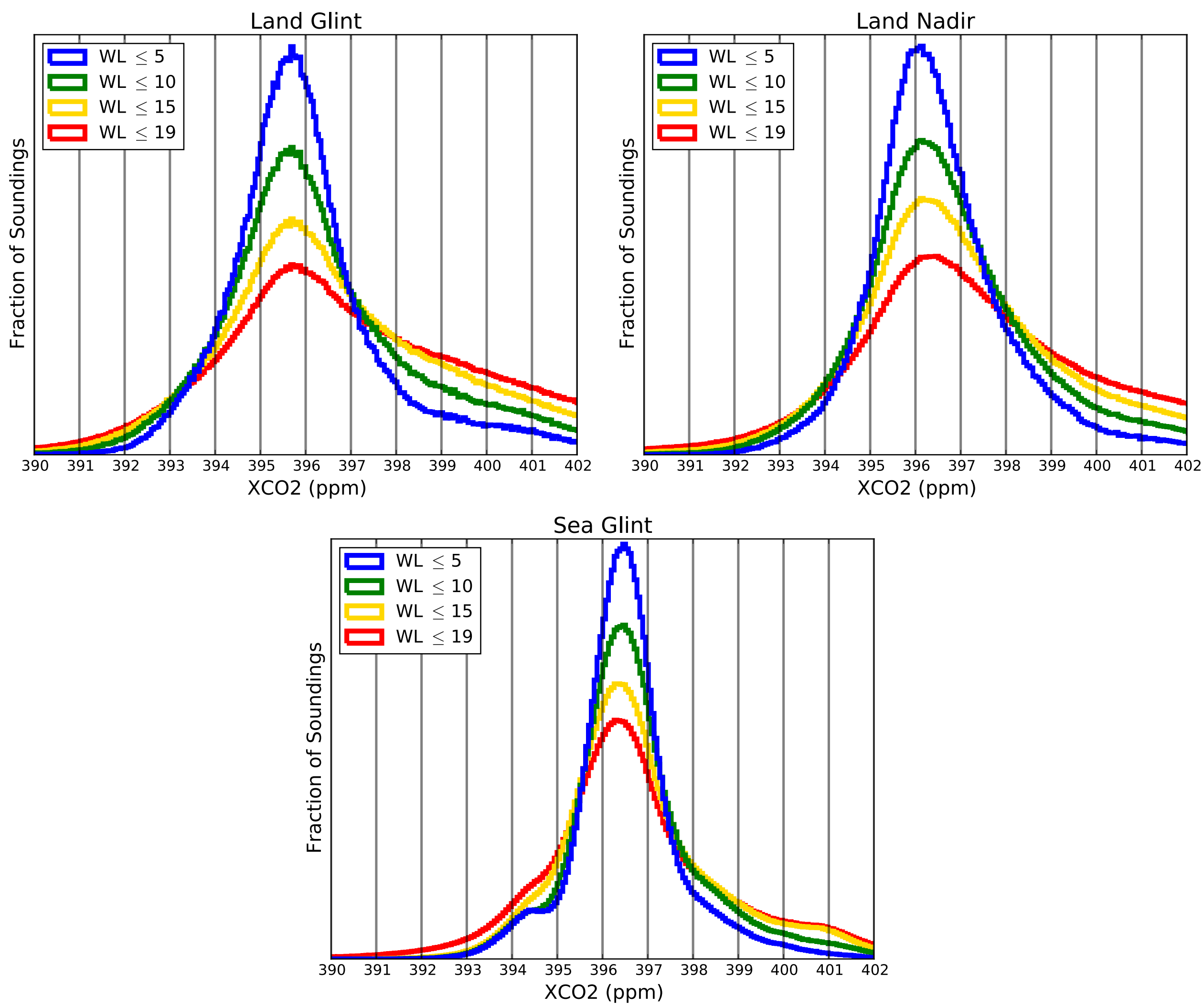


Figure 3 – Histograms of XCO2 vs. WL filtration. Note that higher WL's include steadily more weight in the tails.

Warn Level Spatial Characteristics

The spatial characteristics of the data mapped by WL are shown in figure 4 for 4 WLs selections (inclusive). The geographic patterns of the data change somewhat with WLs. Data-rich regions generally occur over clear, bright locations and data poor regions in areas of intense, persistent cloudiness or near poorly lit poles. There are seasonal patterns to the distribution, and in this figure we show an aggregation of data from November 2014 to May 2015. Because we processed additional data for this analysis south of - 25 degrees latitude, the density is higher in that region.

Figure 5 shows a map of the average WL per region. The bright desert regions (in Africa, Australia) typically have low WL due to brightness and low cloudiness, as do the tropical ocean where there is favorable glint viewing geometry with lower airmass yet high signal.

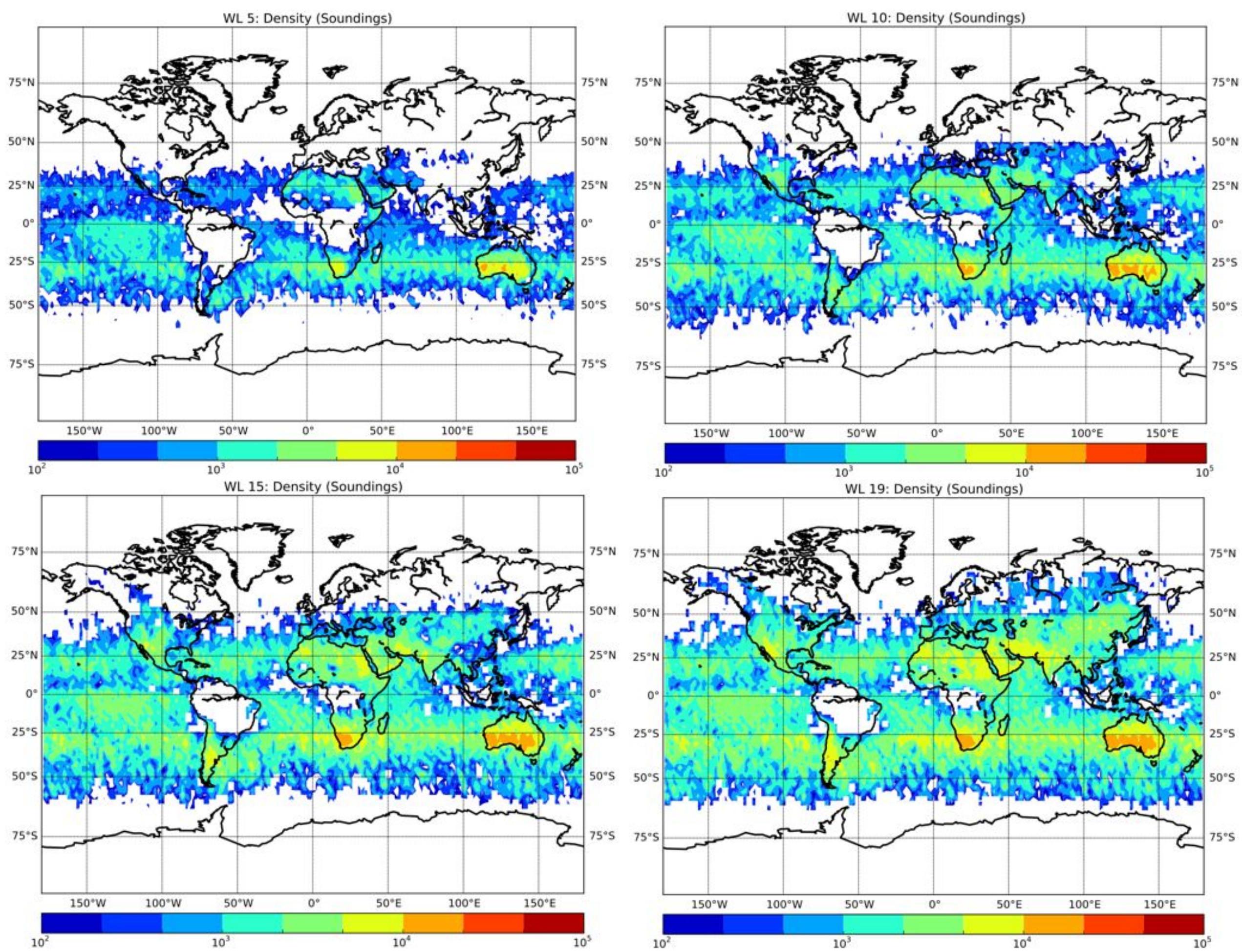


Figure 4 – Data Density as a function of Warn Level

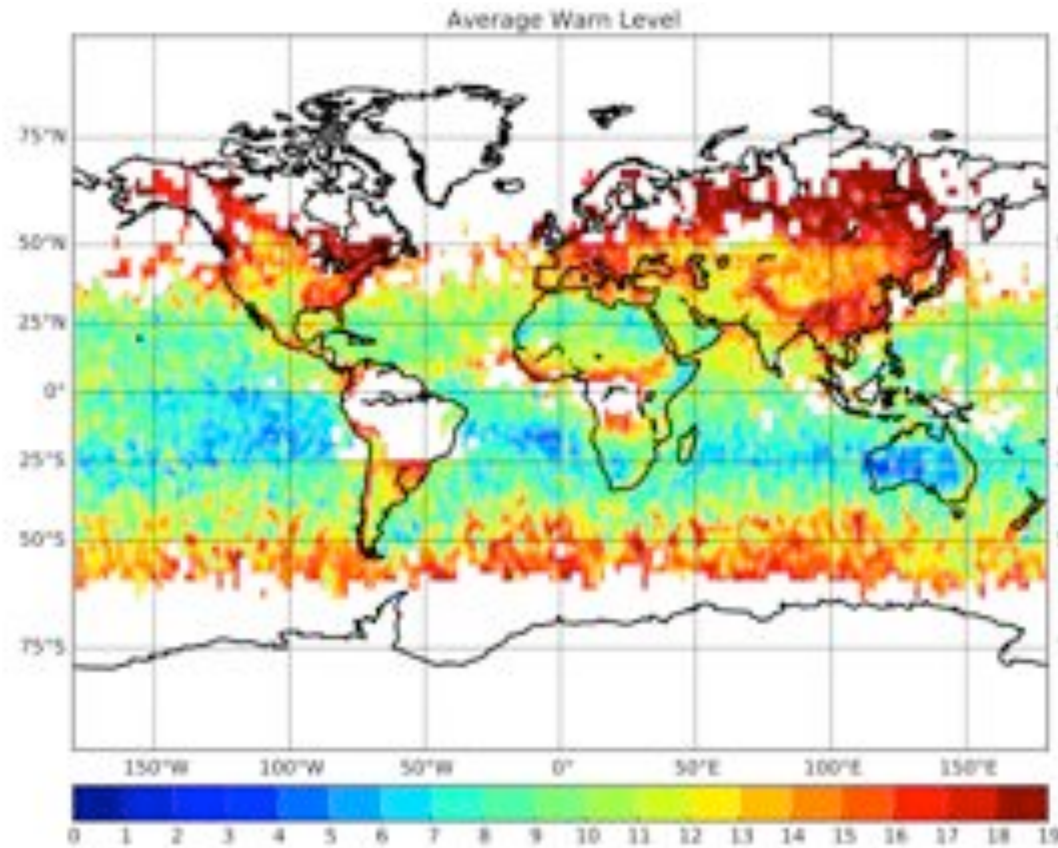


Figure 5 – Average Warn Level Map

How to Use Warn Levels

The basic procedure for use of WLs is outlined in Table 2. A user first decides how much data volume or regional/global coverage they require. Alternately, she can decide how much of an error metric is tolerable (trusted XCO_2 comparison, regional XCO_2 scatter, bias deviation, extrema population presence, or presence of other anomalous sounding behaviors). The WL is then selected that balances the transparency with the error tolerance. The user has then tuned a personal filter for the investigation at hand.

Table 2: **Basic Procedure for Warn Level Usage**

1	Decide requirements beforehand: how much data volume / coverage or scatter / error is needed / tolerable?
2	Begin admitting WL=0, 1, 2, … into project. Monitor above statistics.
3	Stop when data volume / coverage are acceptable, or when scatter / error become intolerable (then back off).

Note that WL's are almost always used Inclusively (meaning that using WL 10 is defined as using all data specified by WL <= 10). This is the proper usage for data filtration. Using a specific WL by itself (data for which WL == 5 alone) is called an Exclusive WL and is sometimes useful for finding the WL at which certain confounding forces appear. By convention, such Exclusive WL's should always be explicitly named to avoid confusion.

WL General Recommendations for v7

For Land Nadir/Glint/Target, WLs above 15 are heavily contaminated and likely not useful. Likewise, for Sea Glint, WLs >= 18 are similarly not likely to be useful. For the sparse dataset of Land Target, WLs <= 8 are poorly defined due to very low N soundings for training. WL <= 8 should be considered a single undifferentiated sounding group of equivalently useful data. Note, final recommendations for warn level use, once bias correction is considered, are reported later in this document.

WL Case 1: Regional Study

Here, a user investigates a difficult environment (e.g. a plume emitted from a polluted city). We wish to harness the WLs to remove the truly useless soundings but still keep interesting yet difficult soundings. The standard procedure of Table 2 is used in the restricted region, watching carefully for the appearance of the plume and associated city pollution. Once the observations to study are clearly present, further WLs may be seen to start including clouds and other non-plume-like soundings. The user therefore backs off the included WLs until a happy medium between plume and unwanted contamination is reached. This method is similar to tuning the contrast in an image, ensuring clear viewing of the desired pattern but not over-emphasizing noise.

Case 2: Uniform Global Study

For this analysis, a user requires uniform global coverage rather than dense coverage of easy regions (e.g. deserts) and sparse coverage of difficult regions (e.g. tropics). The user would define a global grid of bins to fill with the highest quality data until each bin reaches its designated data volume goal. As in Table 2, the co-located data added to each bin is always added in order of WL from lowest to highest. At the end of the process, the soundings selected by the grid represent a spatially uniform collection of the most trusted soundings per region. Some troublesome regions will therefore consist of higher WL soundings or none at all where no alternatives were available, while easier regions will only select the most pristine soundings. This is a powerful selection method that allows the user to dial-in the precise spatial coverage desired for their application with the assurance of using the best quality data available.

Bias Correction

The bias correction (BC) maps the original XCO_2 retrievals of the OCO-2 L2 algorithm to our best estimate of XCO_2. We used v7 data spanning from November of 2014 through May of 2015. The BC approach implemented in the lite files uses regional analysis to identify, within the OCO-2 XCO_2 data records themselves, which terms best predict bias (or Unexplained Variance (%UV)) in the retrievals. Our definition of %UV always uses a null-hypothesis denominator for ease of comparison.

We have constructed several training datasets for our bias correction, such that the set provides an approximation that permits us to define the XCO_2 truth against which the retrievals are evaluated while attempting to insure that real XCO_2 atmospheric gradients are not removed by in the BC.

We use the same two regionally-based training sets used to create the WLs to search for features that predict anomalous variations in XCO_2. As a reminder, the first is the Southern Hemisphere Approximation (SHA), as described in Wunch et al. (2011), in which we assume quasi-constant XCO_2 between 25°S and 60°S latitude during the seven month training period. This method incorporates heterogeneous environments across three continents and surrounding ocean into a single training set that can be used to study what features of the retrieval produce variation in the measured XCO_2. We used several methods to evaluate the sensitivity of the bias evaluation with respect to error in the assumption of invariant XCO_2 in the training set. While the SH XCO_2 data from TCCON provides some guidance (figure 6), several other methods were explored including use of model fields. We find that both the bias features selected and the sensitivity of XCO_2 with respect to these features is relatively insensitive to how the small temporal and latitudinal gradients between 25°S and 60°S latitude are parameterized.

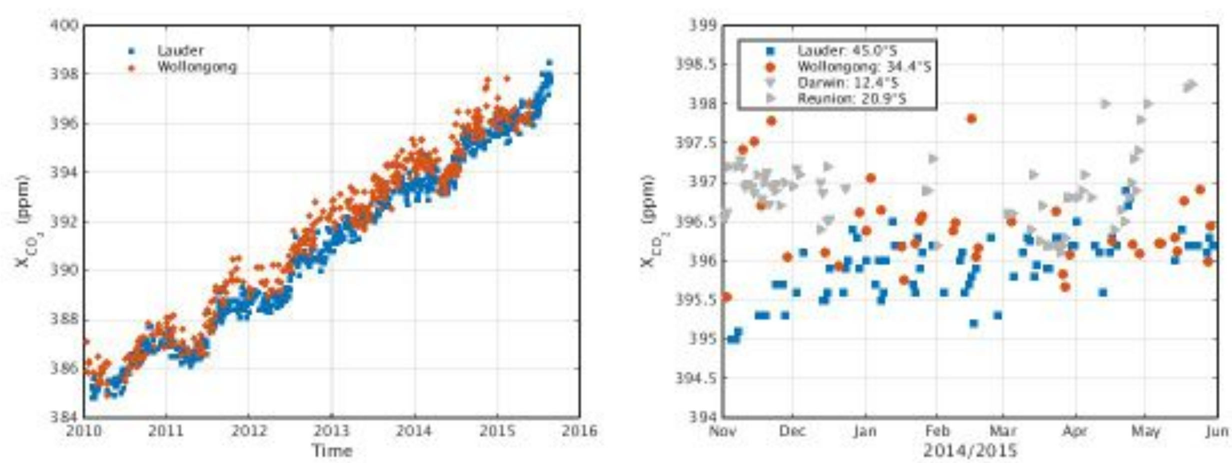


Figure 6: TCCON data in the southern hemisphere show small latitude gradients and seasonal cycles south of 25°S. In the left panel, the Lauder (45°S) and Wollongong

(34.4°S) TCCON time series are shown from 2010 to the present. There is a small (~1 ppm amplitude) seasonal cycle, and an overall secular increase (~2.2 ppm/yr). The Wollongong data are higher than the Lauder data by ~0. 5 ppm, leading to a latitudinal gradient of ~0.05 ppm/degree. The right panel shows the data during November 2014 through May 2015 and includes other southern hemisphere TCCON stations (Darwin and Reunion Island) that are north of 25°S. We limit the SHA to south of 25°S, as the Darwin and Reunion data show significantly more seasonal variations.

The second training set is created using Small Area analysis in which we identify groups of dense soundings (20-80) obtained in spatially-small neighborhoods (<0.9° latitude) obtained within the same orbit (e.g within ~30 seconds). We assume that XCO_2 is constant over these very small areas, and that any XCO_2 variation is an artifact of the retrieval. We expect that except in regions of intense sources or sinks of CO2 (e.g. urban regions) or along frontal systems true variation in XCO2 is much smaller than the variance produced by the L2 algorithm. As shown in figure 7, these groupings are globally distributed, sampling a wider heterogeneity than the southern hemisphere.

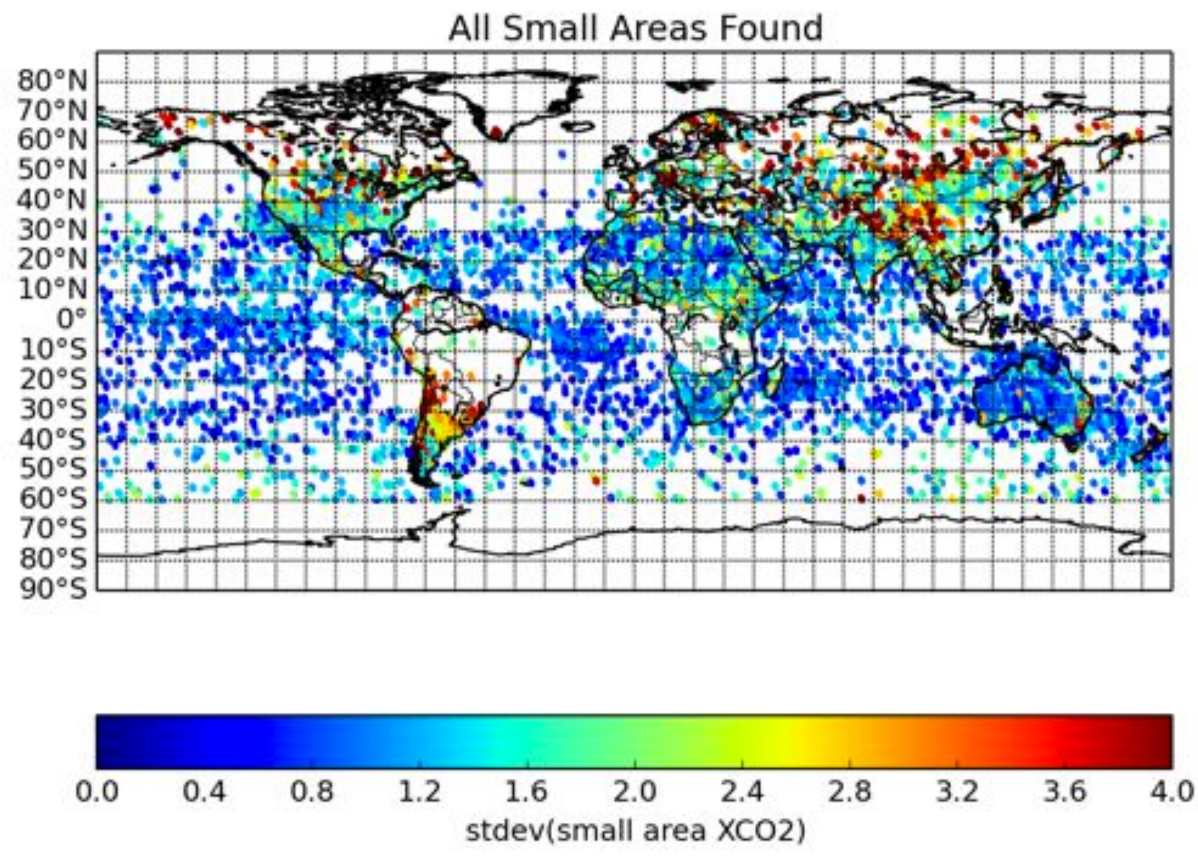


Figure 7 – Location of small areas used in Bias Correction

Note that in the bias correction, TCCON data are used only to define a global offset between the OCO-2 data and the WMO in situ scale as a final step.

The three steps of BC determination are as follows, made concrete by highlighting terms in the final bias correction equation:

$$XCO_{2\,Corrected} = \frac{XCO_{2\,Raw} - FOOT[fp, mode] - FEATS[mode]}{TCCON\_ADJUST[mode]}$$

Step 1) Removing Footprint Bias (term FOOT)

OCO-2 obtains 8 side-by-side, simultaneous, nearly-colocated measurement scenes called footprints. Although the XCO_2 retrievals made from these soundings should, on average, be identical, there are small and highly statistically significant differences. These undoubtedly arise from imprecision in the L1 calibration.

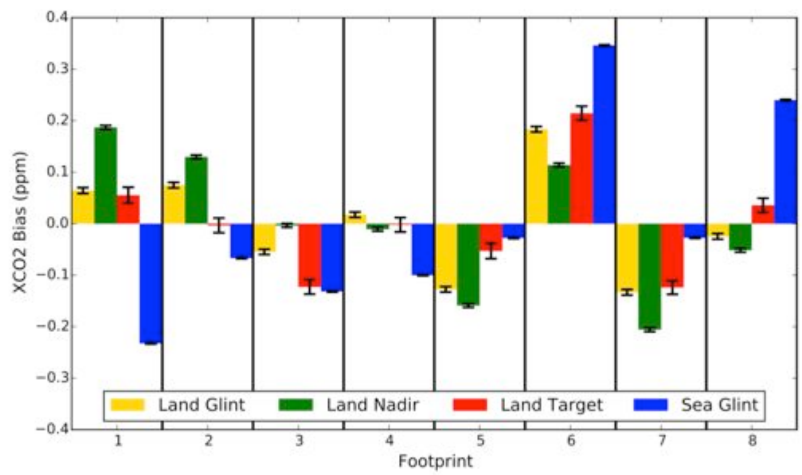


Figure 8 – Footprint Bias vs. Mode

For the v7 data set, we examined highly filtered data grouped into small areas. For most modes, the data set contained millions of soundings, so they were further down selected to only include the best data. Data were selected for the analysis when all 8 footprints in one sounding frame (ID) converged. Small Areas were kept that had at least 100 soundings and STD(XCO_2) < a threshold defined by percentile (10%, or 75% for target). For each remaining small area, we computed the median XCO_2 as the "ground truth" value, and subtracted this from the observed XCO_2 to calculate the deviations for each footprint. The reported values are the average of the differences across all soundings per footprint. The filtration resulted in land selections near bright, clear regions such as deserts. For Sea Glint, the dataset was about 68K soundings per footprint, and 16K and 28K for Land Glint and Land Nadir, respectively. The footprint offsets are listed in Table 4 under the term FOOT and shown in Figure 8. Once obtained, they may be subtracted from all sounding XCO_2 in preparation for the next steps of BC.

$$XCO_{2\,Corrected} = \frac{XCO_{2\,Raw} - FOOT[fp, mode] - FEATS[mode]}{TCCON\_ADJUST[mode]}$$

Step 2) Identify unphysical XCO_2 variability (term FEATS)

In this second step, we identify variables retrieved simultaneously with XCO_2 that are correlated with spurious XCO_2 variability using a multivariate linear regression. This is the procedure followed by nearly all GOSAT XCO_2 bias corrections to date (e.g. Wunch et al. 2011, Guerlet et al. 2013).

Several approaches were used to explore this large space of features in the L2 and preprocessing fields: 1) a genetic algorithm that preserves pairwise and higher relationships between features vs. all possible warn level filtration options (Mandrake et al., 2013); 2) a simpler search that sequentially adds the single most %UV-reducing feature from a smaller, expert-curated feature set; and 3) a traditional linear regression sensitivity analysis, again upon a smaller expert-chosen feature set. Using these approaches, we identified four variables that appear to drive approximately half of the %UV in both the SHA and SAA training sets. The different analysis regions and modes have minor differences in the order of features chosen, but the overall features selected and their slopes are remarkably consistent. Table 3 shows for the Land Nadir small area and Sea Glint small area set, the reduction in unexplained variance (UV in %) as features are added to the bias correction fit using three different WL filtration levels:

Table 3: Features used in bias correction with reduction in unexplained variance

Land Features Selected	UV WL05	ΔUV	UV WL10	ΔUV	UV WL15	ΔUV
dP_fph	83.8	16.2	77.9	22.1	80.5	19.5
co2_grad_del	60.3	23.5	58.9	19.0	68.7	11.8
log(DWS)	50.0	10.4	49.8	9.1	60.0	8.7

Sea Features Selected	UV WL05	ΔUV	UV WL10	ΔUV	UV WL15	ΔUV
co2_grad_del	78.8	21.2	81.4	18.6	84.0	16.0
dP_fph	73.9	4.9	75.8	5.6	75.9	8.1

These terms are the differences between the retrieved and the *a priori* surface pressure (dP), the retrieved abundance of coarse aerosol (e.g. dust, sea salt, or water clouds), and (very large and unphysical) variation in the retrieved vertical profile of CO_2 (parameter co2_grad_del) from that assumed in the prior.

There is a fundamentally-different bias behavior over land vs. water surfaces requiring separate bias correction terms. Over land, there is no significant difference between the nadir- and glint feature-dependent biases.

$$XCO_{2\ Corrected} = \frac{XCO_{2\ Raw} - FOOT[fp, mode] - FEATS[mode]}{TCCON\_ADJUST[mode]}$$

Step 3) Determine global offset from TCCON (term TCCON_ADJUST)

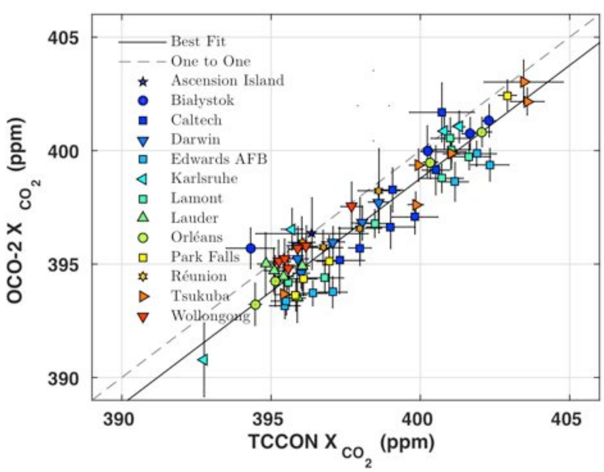


Figure 9 – OCO-2 vs TCCON Alignment Plot

The analyses described in Steps 1 & 2 provide no estimate of the overall global bias in XCO_2 To determine the offset, we used target mode observations obtained over the TCCON sites. A linear regression between TCCON XCO_2 and the bias-corrected OCO-2 target mode XCO_2 was performed, with the intercept forced to zero (Figure 9). Overall, this analysis results in the following equation for land target: OCO-2 = $(0.99694 \pm 0.00102) \times$ TCCON. OCO-2 retrievals are consistently lower than TCCON (about 1 ppm).

The data used to develop Figure 9 were limited to WL 15, with the outlier filters applied as described in the lite file section. In addition, the solar zenith angle and retrieval zenith angles are restricted to less than 40 degrees. Regarding TCCON data, sites are included only if there are 5 or more TCCON retrievals within one hour of the overpass, or more than 10 TCCON retrievals available within 2 hours of the overpass. After filtering, 64 targets are included out of the 121 that were collected. They were collected between September 7th, 2014 and August 11th, 2015. Most of the targets not included in Figure 9 suffered from significant cloud contamination.

The TCCON adjustment was determined relative to the OCO-2 Land Target observations. We assume that with the restricted retrieval zenith, the land Target and Land Glint have identical bias. Future work will examine the veracity of this assumption and explore alternative methods of spreading the divisor between modes.

Land Nadir and Sea Glint are matched in time and space to Land Glint in order to propagate the TCCON alignment divisor. Coastline measurements are used to match the Land Nadir/Sea Glint datasets (Figure 10). Land Nadir and Sea Glint both produced different divisors by the equation $\frac{<XCO_2>_{mode1}}{divisor_1} = \frac{<XCO_2>_{mode2}}{divisor_2}$. Results can be found in the final Bias Correction, Table 4.

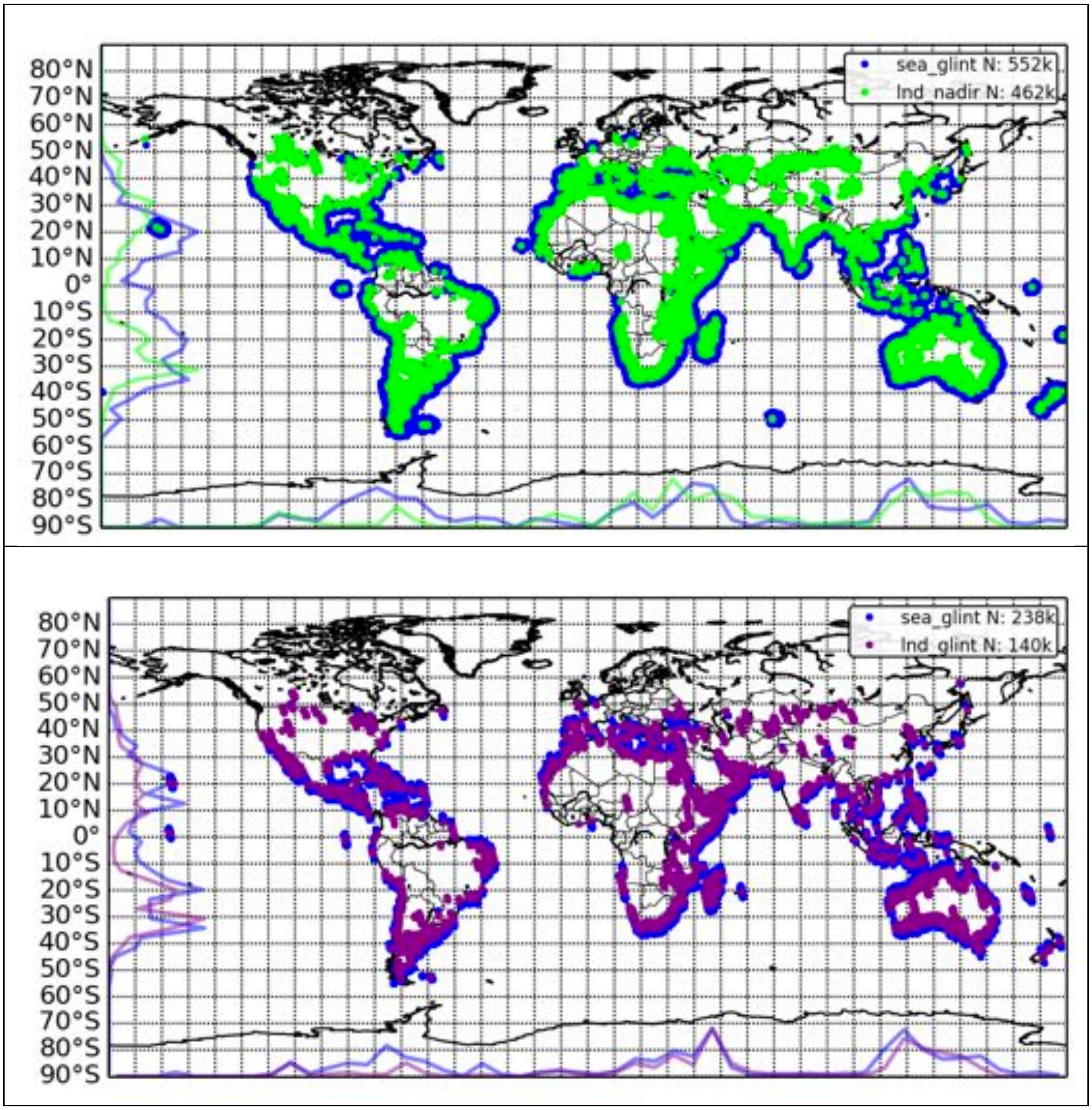


Figure 10 – Land Nadir/Glint vs. Sea Glint matching

Table 4 Bias Correction Formula (for use on v7.0)

$$XCO_{2\,Corrected} = \frac{XCO_{2\,Raw} - FOOT[fp, mode] - FEATS[mode]}{TCCON\_ADJUST[mode]}$$

where:

	FOOTPRINT BIAS (FOOT) (ppm)							
Footprint (fp)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
LAND GLINT	0.06	0.07	-0.05	0.02	-0.13	0.18	-0.13	-0.02
LAND NADIR	0.19	0.13	0.00	-0.01	-0.16	0.11	-0.21	-0.05
LAND TARGET	0.06	0.00	-0.12	0.00	-0.05	0.21	-0.12	0.04
SEA GLINT	-0.23	-0.07	-0.13	-0.10	-0.03	0.35	-0.03	0.24

	FEATURE BIAS (FEATS) (ppm)
LAND (ALL)	– 0.3*(**dP** - 1.4) – 0.6*(**log_DWS** + 2.9) – 0.028*(**co2_grad_del** – 8.4)
SEA GLINT	– 0.08*(**dP** – 3.1) + 0.077*(**co2_grad_del** + 7.7)

	OVERALL DIVISOR (TCCON_ADJUST)	**Method**
LAND GLINT	0.9970	Assigned (similar to Target)
LAND NADIR	0.9955	Propagated via mean(XCO_2)
LAND TARGET	0.9970	Derived using TCCON + Target Mode
SEA GLINT	0.9990	Propagated via mean(XCO_2)

NOTE! This constant is needed for the L2 standard but not Lite products

	Variable definitions using full/HDF/path
co2_grad_del	1e6 * RetrievalResults/co2_vertical_gradient_delta
dP	0.01 * (RetrievalResults/surface_pressure_fph – RetrievalResults/surface_pressure_apriori_fph)
log_DWS	Max (-5 , log (dust_aod + water_aod + salt_aod))
dust_aod (1-based)	$\sum_{k=1}^{2}$ (AerosolResults/aerosol_types[k] == 'DU') * (AerosolResults/aerosol_k_aod)
water_aod (1-based)	$\sum_{k=1}^{2}$ (AerosolResults/aerosol_types[k] == 'Water') * (AerosolResults/aerosol_k_aod)
salt_aod (1-based)	$\sum_{k=1}^{2}$ (AerosolResults/aerosol_types[k] == 'SS') * (AerosolResults/aerosol_k_aod)

Recommendation for XCO_2 WL/BC use, subject to limitations:

Considering the warn level and bias correction characteristics together, we recommend the following for initial investigations:

- Do not use data above warn level (WL) 15
- Above WL 12, errors well in excess of the stated a posteriori errors should be expected.
- BC was generated using November 2014 to May 2015. Extrapolation to other date ranges will decrease accuracy
- BC improves the XCO_2 estimate in a global-average sense. Specific regions may experience less improvement or perhaps even degradation compared to the raw L2 retrievals.
- As discussed below, extreme caution is recommended in the use ocean glint data obtained at high airmass (> 2.5). These are known to suffer from a high bias (several ppm at airmass > 3).

Residual Footprint Bias

Post application of the bias correction as described above, it was determined that a footprint-dependence XCO_2 bias had been re-introduced via the FEATS term log(DWS) (figure 11). For the purposes of offering users the best XCO_2 for analysis purposes, it is recommended that users perform a final footprint bias correction as documented in Table 5 and the following equation. Future Lite product releases will include this step in the bias corrected XCO_2.

$$XCO_{2\,Final} = XCO_{2\,Corrected} - FOOT2[fp, mode]$$

Table 5: Residual Footprint Bias

	RESIDUAL FOOTPRINT BIAS (FOOT2) (ppm)							
Footprint	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
LAND GLINT	-0.13	-0.09	-0.01	-0.14	0.00	0.06	0.09	0.21
LAND NADIR	-0.17	-0.10	-0.02	-0.09	0.04	0.03	0.13	0.18
LAND TARGET	-0.17	-0.04	0.02	-0.10	0.05	0.01	0.07	0.16
SEA GLINT	0.03	0.00	0.02	0.00	0.00	-0.08	0.02	0.01

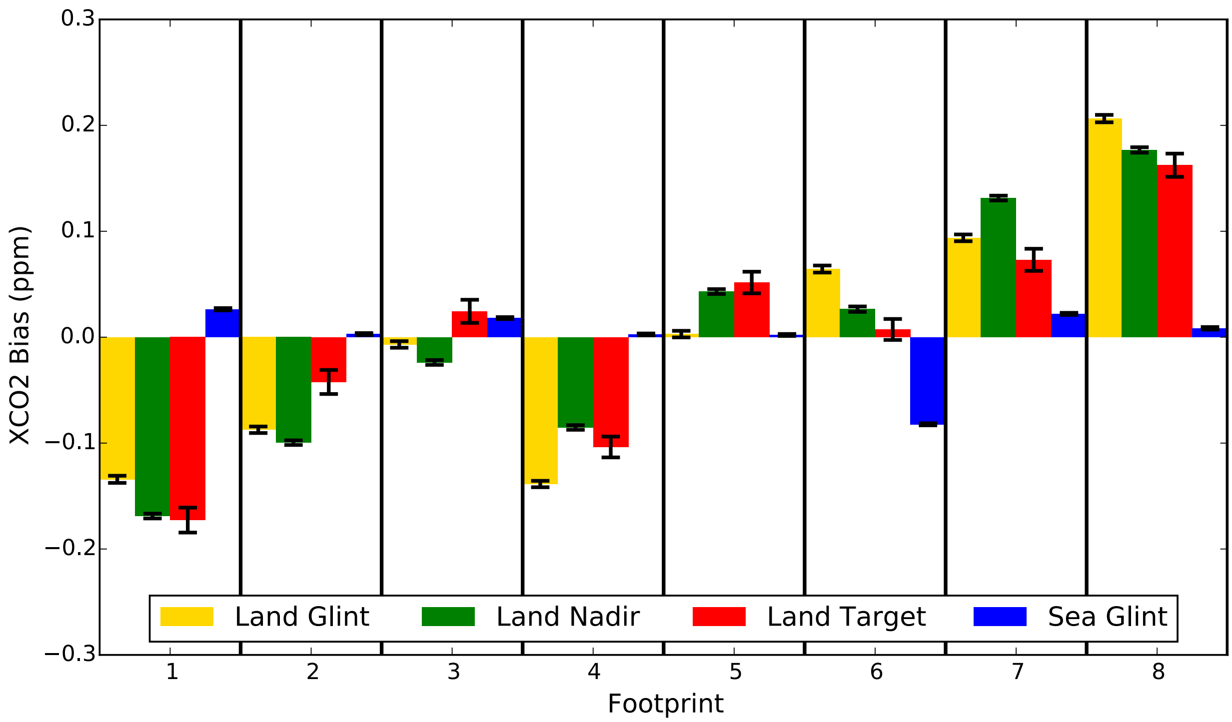


Figure 11: Residual Footprint Bias induced by log(DWS) term in bias FEATS term

Bias Correction Observations and Remaining Issues

Outlier Removal

The Warn Levels were designed, as described above, to reduce spurious XCO_2 variance. However, they still permit low-density outliers (points with anomalous XCO_2, for example soundings over snow) that do not affect the overall metrics used to derive the WLs. These outliers may still pose a threat to certain analysis such as linear regressions, where far-flung points with low density can still strongly perturb slope calculations. For the purposes of Bias Correction calculation, we developed an additional outlier filter (see "Lite" file Data Quality Flag documentation) to flag such problematic minority populations. Future versions of the WLs will include outlier handling automatically.

Figure 12 shows the bias and standard deviation of XCO_2 error as a function of selected filtering variables over land. XCO_2 has already been bias-corrected in these plots. After filtering to WL<=15, the standard deviation falls to 1.54 ppm. However, it is clear that a number of variables can lead to significant bias in the *bias-corrected* XCO_2. For instance, over snow and ice surfaces, where the blended albedo > 0.8, a strong negative bias (and very large scatter) can result. Similarly for the error in the surface pressure retrieval and a number of other variables. Figure 13 shows the same for ocean glint observations. Only ~13% additional observations are removed with the outlier filtering, but some variable exhibit strong XCO_2 bias in the tails of their distributions, most notably airmass and Band 3 albedo slope.

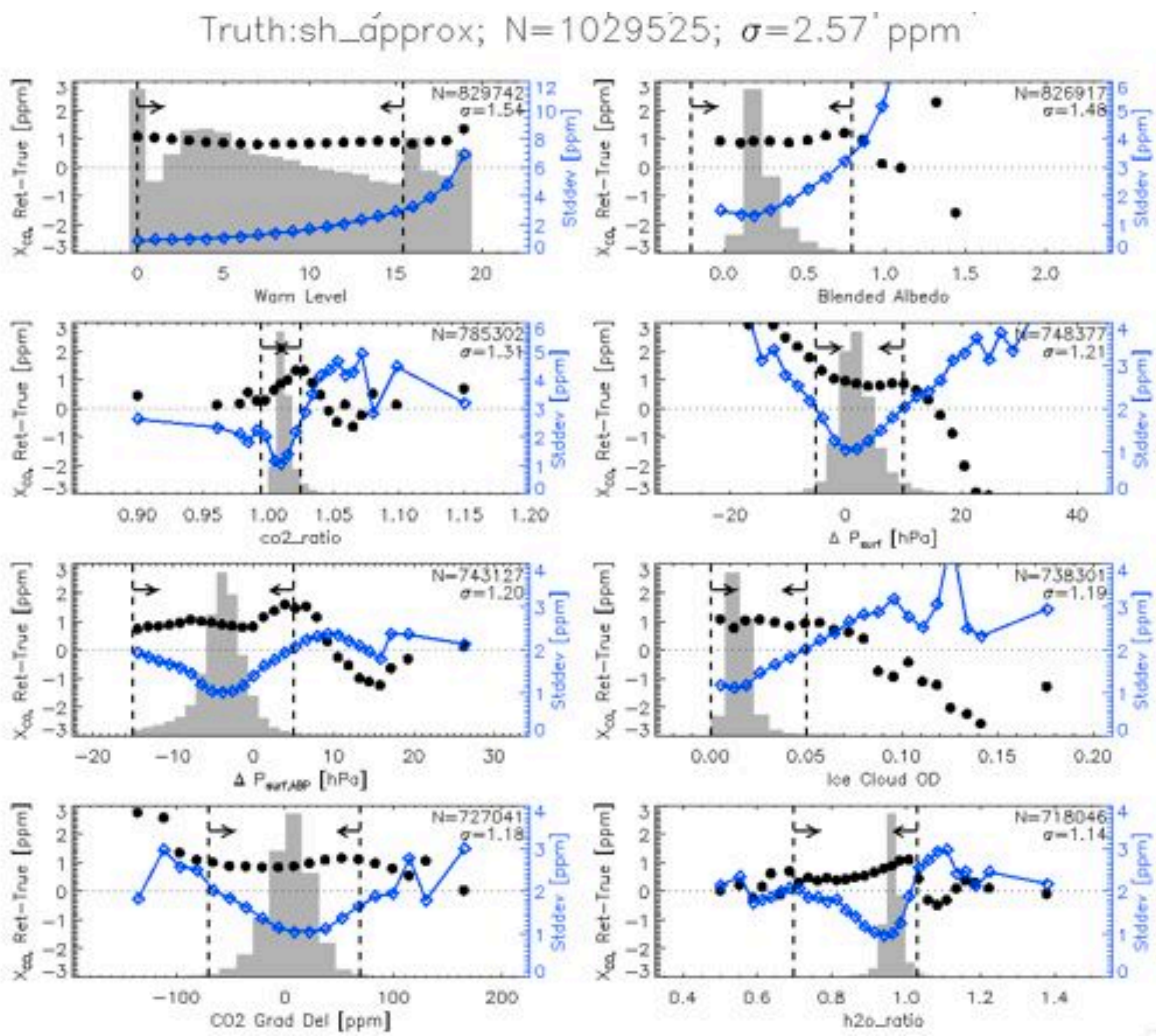


Figure 12 – Bias & Stddev of XCO_2 error as a function of selected outlier filtering variables for land observations. Here, the error is derived using the Southern Hemisphere approximation. Even though only the tails of the distributions are being removed, the XCO_2 error variance is significantly reduced relative to WL<=15 alone.

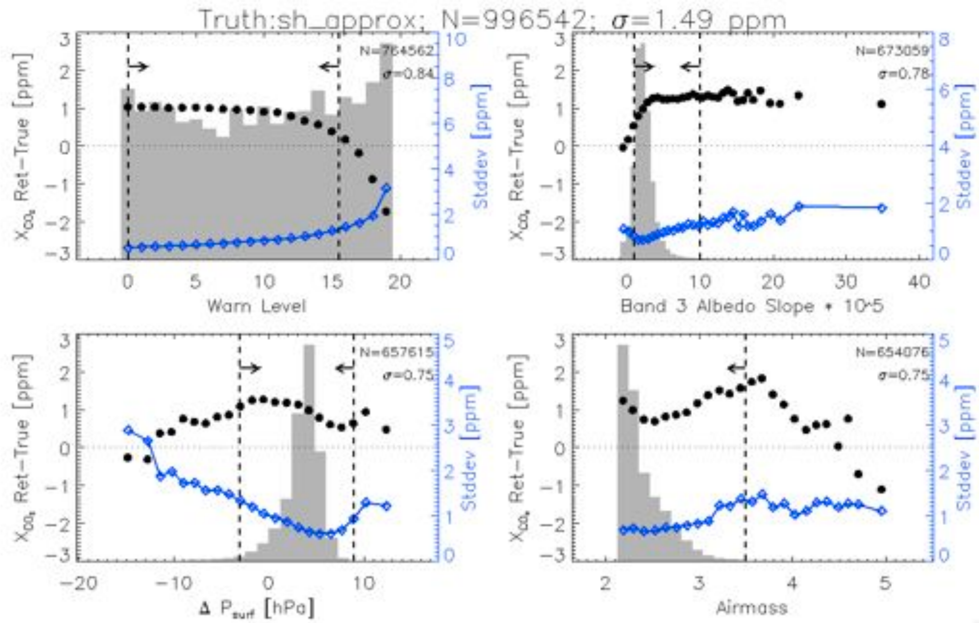


Figure 13 – Same as Figure 12, but for ocean glint observations.

Regional Features

The selected features that form the unphysical bias trends from Step 2 were selected by minimizing the global %UV. Small regions of the world may manifest very different predictive features representative of their specific environment.

Airmass dependence

Airmass may contribute or be associated with anomalous XCO_2 variance. However, the influence of airmass is often masked by other, more predictive features (such as CO2_grad_del) that combine airmass bias with other terms.

Bias in the retrieved XCO_2 with airmass was explored in the southern hemisphere and in target data. While only minimal airmass dependence was observed in Target after bias correction (Figure 14), significant XCO_2 error is associated with airmass in the ocean glint data in the Southern Hemisphere (Figure 15). A clear trend with airmass is observed. This trend is present with and without bias correction and represents an as-yet unexplained behavior of the retrieval. Because airmass and season are correlated, this bias also produces a time variance in the retrievals (Figure 16). Research is continuing on how to isolate and remove this bias. Such behavior was not recognized in GOSAT retrievals likely because they extend only to modest airmass.

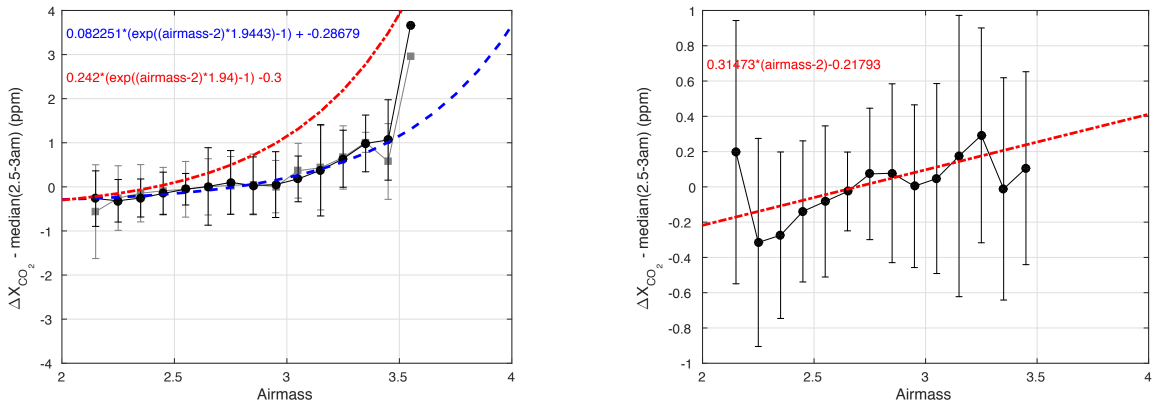


Figure 14: Airmass dependence in uncorrected target data, with exploratory expressions (lower left), airmass dependence in bias corrected target data with a linear fit.

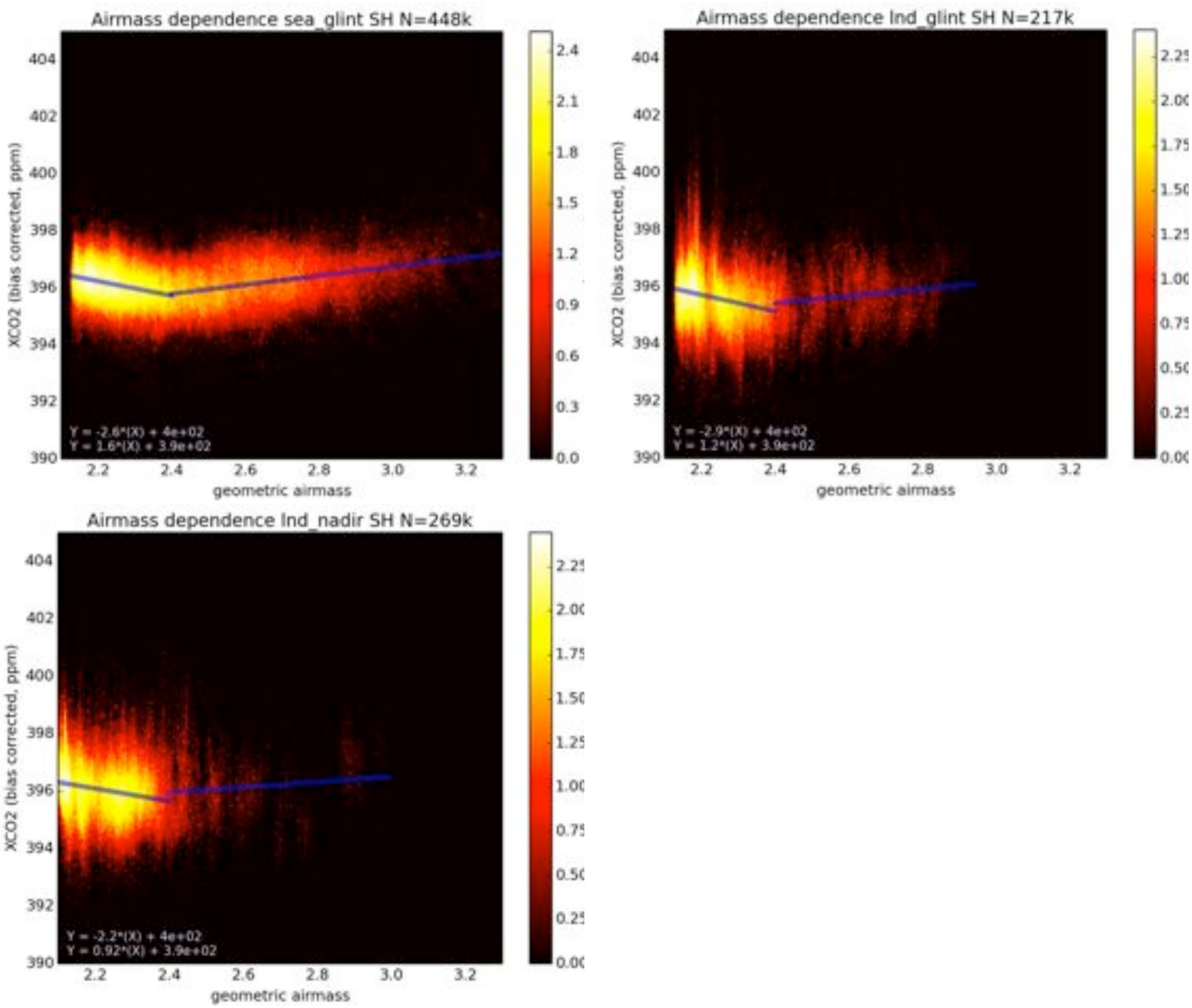


Figure 15: Airmass dependence in bias corrected sea glint data (upper left), land glint (upper right), and land nadir (bottom).

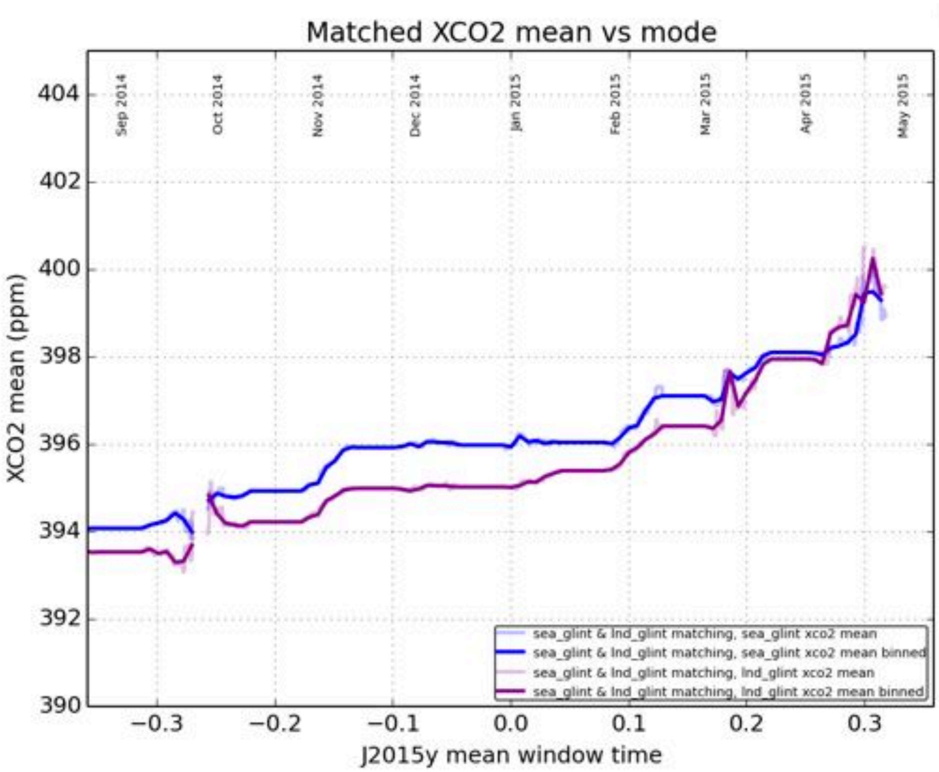


Figure 16: Mean XCO_2 (ppm) vs. time (decimal year, 2015=0). Blue shows Sea Glint, while purple shows Land Glint. The right of the graph shows a change in relationship.

Bias (in)Stability vs. Filtering

The bias correction defined in Table 4 was developed from the filtered training set. We used a filter of WL ≤ 10 for Land and Sea and applied a series of outlier-removing filters to remove rare but troublesome points, as defined in the "Lite" file documentation for the quality flag. It is important to recognize that the coefficients listed in Table 4 are not independent of the choice of filtering. For example, as illustrated in the left panel of Figure 17, the two terms in the Sea Glint bias solution depend on the WL filter chosen. These plots have normalized the slope coefficients of each term by dividing by the std of the term and multiplying by the std of the XCO_2 values. Thus, they represent sensitivities, a measure of relative significance. For Sea Glint, we observe that below WL 10, there is relatively little variation in the determined coefficients of the bias terms. However, above WL 10 the solutions diverge, decreasing the accuracy of the BC formula. Similarly, as illustrated on the right of Figure 17, CO2_grad_del is a non-linearly rising function of WL over its entire span, while dP first increases then decreases in its influence. More advanced formulations of bias, involving non-linear terms and potentially filtration-dependent terms, are being considered for the next release of data to address these problems.

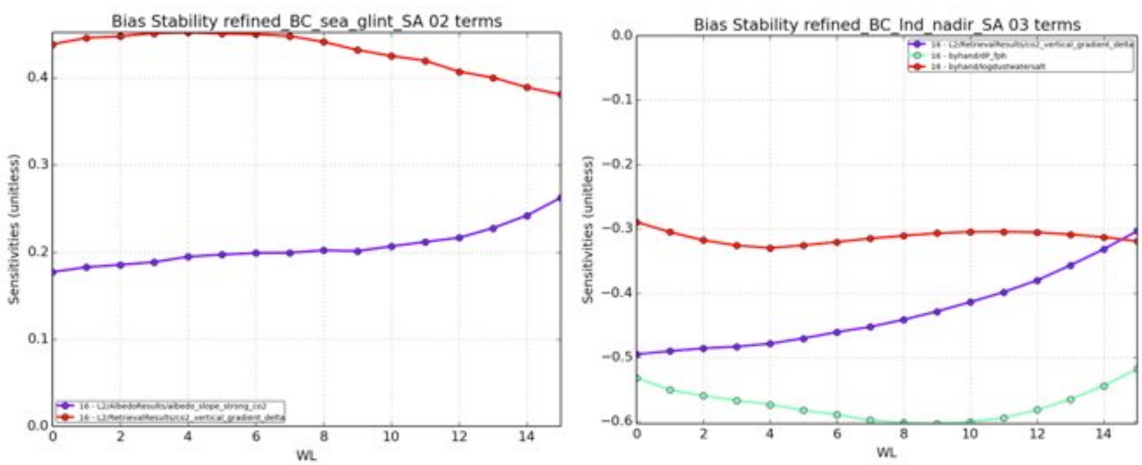


Figure 17 – Stability Curves for Bias Correction Terms, Sea & Land

Histogram of Bias Terms

Figure 18 shows the histograms of bias contributions from Step 2 as well as the overall Bias Correction. The logDWS (blue) parameter's spike relates to the cutoff in its definition (see final bias correction formula). 1 ppm is subtracted from the overall bias determination just as it is for TCCON to match the World Meteorological Organization (WMO) *in situ* trace gas measurement scales.

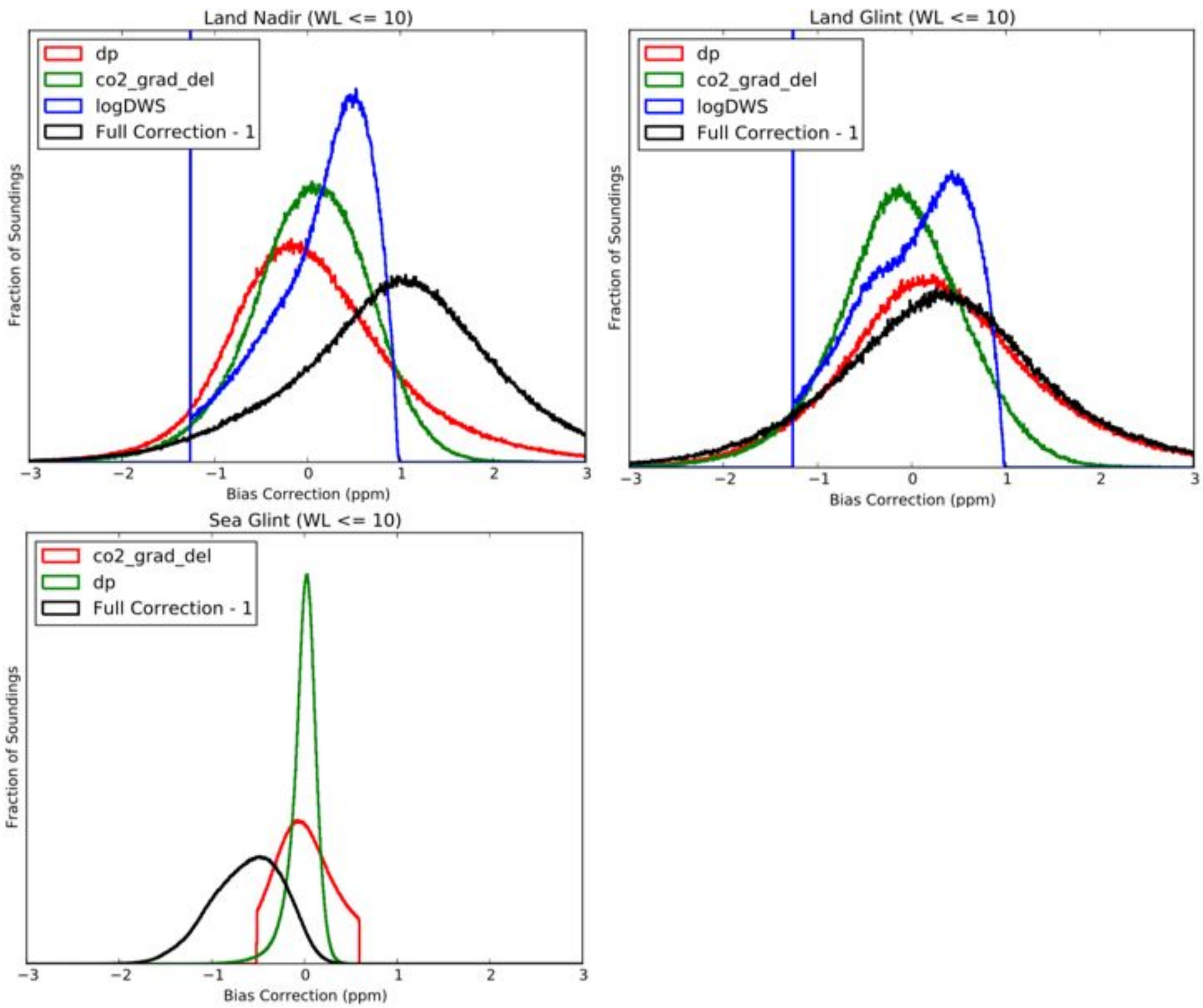


Figure 18 – Histogram of Bias Terms

Global Bias Magnitude

Figures 19 to 21 show Land Nadir, Land Glint, & Sea Glint regional biases ($XCO2_{corrected}$ – XCO_2). White indicates a zero mean effect, while gray regions have no data. Extreme bias correction effects (dark red/blue, ~ 3.5 ppm) are chiefly observed in bins with low N.

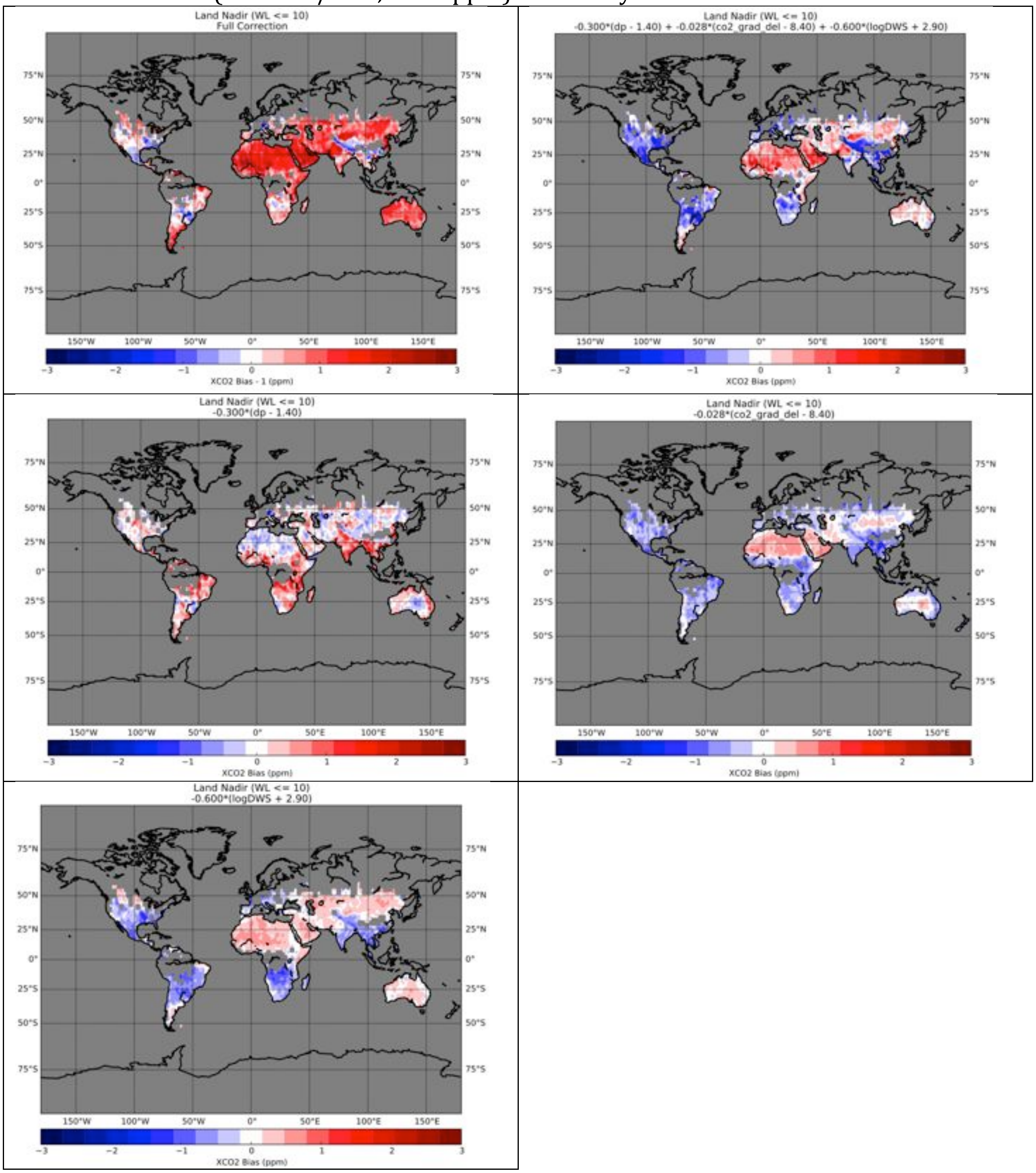


Figure 19 - Land Nadir Bias Correction Terms, Geographic Distribution

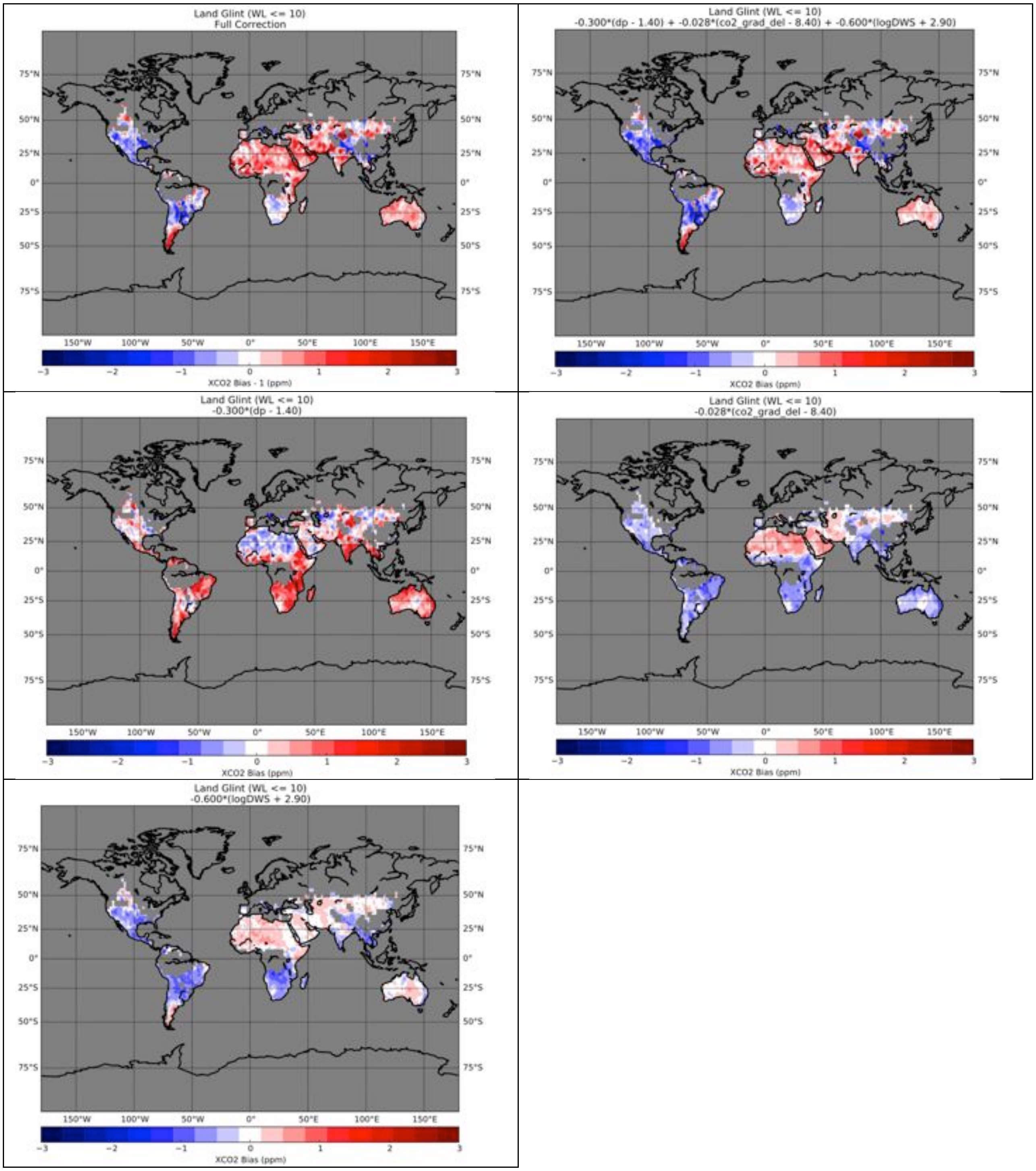


Figure 20 - Land Glint Bias Correction Terms, Geographic Distribution

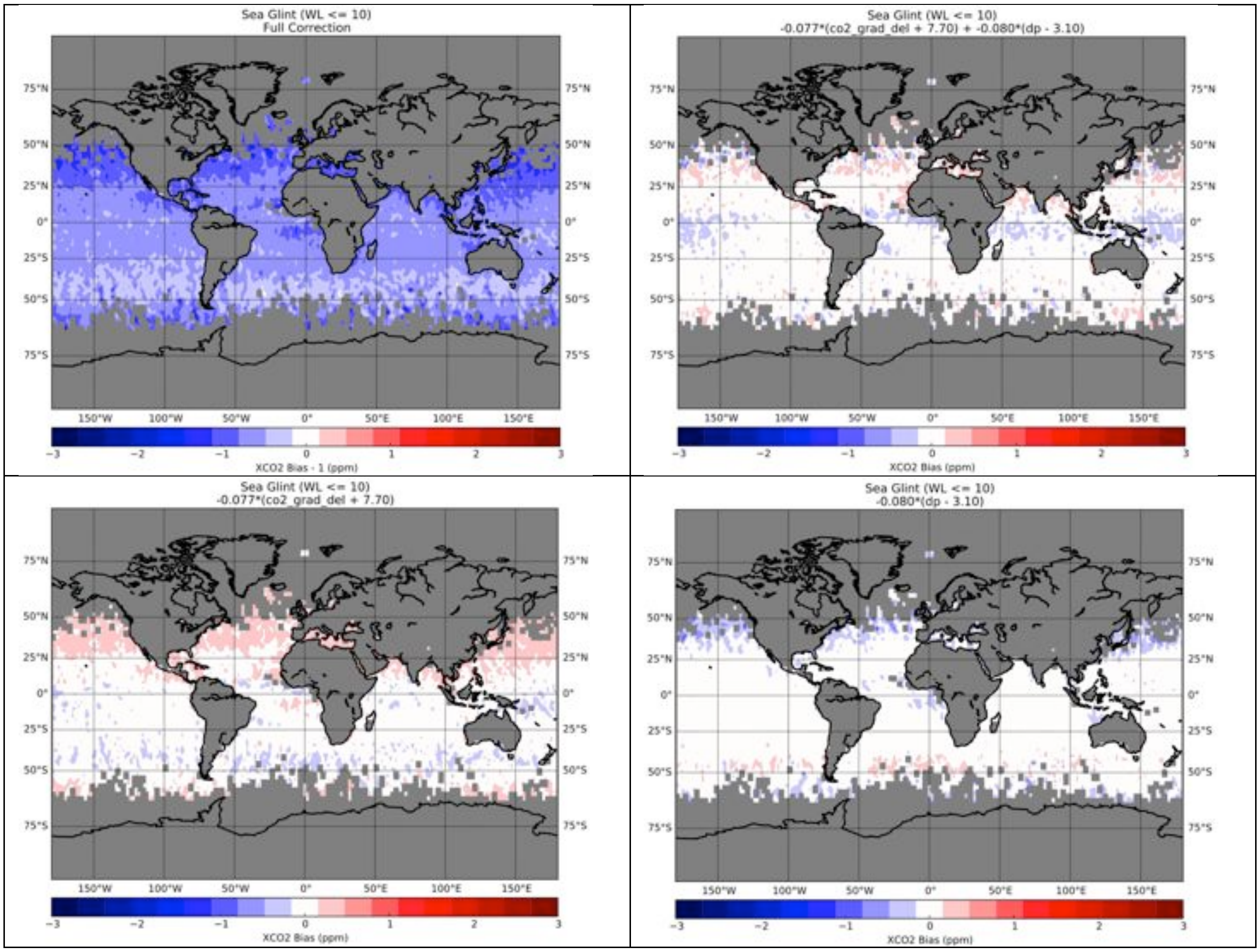


Figure 21 - Sea Glint Bias Correction Terms, Geographic Distribution

Feedback

Please provide feedback and comments to us at oco2_feedback@list.jpl.nasa.gov. You can receive updates and news by subscribing to our list. Email a blank message to sympa@list.jpl.nasa.gov with the subject: subscribe oco2_updates.

References

Guerlet, S., et al: Impact of aerosol and thin cirrus on retrieving and validating XCO_2 from GOSAT shortwave infrared measurements, *J. Geophys. Res. Atmos.*, **118**, 4887–4905, doi:10.1002/jgrd.50332, 2013

Mandrake, L., Frankenberg, C., O'Dell, C. W., Osterman, G., Wennberg, P., and Wunch, D.: Semi-autonomous sounding selection for OCO-2, *Atmos. Meas. Tech.*, **6**, 2851-2864, doi:10.5194/amt-6-2851-2013, 2013.

Wunch, D., et al: A method for evaluating bias in global measurements of CO_2 total columns from space, *Atmos. Chem. Phys.*, **11**, 12317-12337, doi:10.5194/acp-11-12317-2011, 2011

Appendix A:

Here we report the exact values used to define the warn levels for v7. Note that there are three tables, one each for land (nadir and glint modes treated the same way), glint water, and land target.

The variables in the tables that follow are used exactly as read from the L2 data files. No scaling is applied (this is especially important for co2_vertical_gradient_delta, which is used in the unitless form for warn levels and in ppm for the bias correction.

Variables preceded by a '>' symbol indicate that values ABOVE the listed threshold are rejected, while '<' indicate values BELOW the listed threshold are rejected.

Table A1: Warn level definitions for Land Nadir and Land Glint

WL_defs	>aerosol total aod	<surface pressure delta abp	>retrieval surface roughness	>relative residual mean square strong co2
5%	0.090	-675	7.00	0.00142
10%	0.106	-686	7.35	0.00145
15%	0.123	-696	8.25	0.00149
20%	0.140	-707	9.70	0.00155
25%	0.159	-718	11.70	0.00161
30%	0.179	-729	14.25	0.00169
35%	0.200	-739	17.35	0.00177
40%	0.222	-750	21.00	0.00187
45%	0.245	-789	25.20	0.00198
50%	0.269	-844	29.95	0.00210
55%	0.294	-914	35.25	0.00223
60%	0.320	-1000	41.10	0.00237
65%	0.347	-1101	47.50	0.00252
70%	0.375	-1218	54.45	0.00269
75%	0.404	-1350	61.95	0.00286
80%	0.435	-1507	70.00	0.00305
85%	1.400	-1725	78.60	0.00324
90%	2.300	-2005	90.00	0.00380
95%	2.800	-2347	100.00	0.00550

Table A2: Warn level definitions for Sea Glint

WL defs	>aerosol total aod	<albedo slope strong co2	>co2 vertical gradient delta	<co2 vertical gradient delta
5%	0.046	0.0000140	-0.00000663	-0.00000996
10%	0.051	0.0000133	-0.00000604	-0.00001041
15%	0.056	0.0000127	-0.00000543	-0.00001087
20%	0.061	0.0000121	-0.00000481	-0.00001132
25%	0.065	0.0000114	-0.00000418	-0.00001178
30%	0.070	0.0000108	-0.00000352	-0.00001223
35%	0.075	0.0000102	-0.00000285	-0.00001269
40%	0.080	0.0000095	-0.00000217	-0.00001314
45%	0.090	0.0000089	-0.00000147	-0.00001360
50%	0.105	0.0000083	-0.00000075	-0.00001370
55%	0.120	0.0000076	-0.00000002	-0.00001440
60%	0.135	0.0000070	0.00000090	-0.00001510
65%	0.155	0.0000051	0.00000250	-0.00001550
70%	0.175	0.0000019	0.00000420	-0.00001620
75%	0.200	-0.0000026	0.00000660	-0.00001740
80%	0.240	-0.0000085	0.00000880	-0.00001850
85%	0.320	-0.0000156	0.00001350	-0.00002000
90%	0.500	-0.0000239	0.00002100	-0.00002250
95%	0.700	-0.0000500	0.00003400	-0.00002800

Table A3: Warn level definitions for Land Target

WL defs	<surface pressure delta abp	>surface pressure delta abp	>retrieval surface roughness	>relative residual mean square WCO2
5%	-200	36	2.50	0.00135
10%	-650	75	5.00	0.00140
15%	-700	114	8.00	0.00145
20%	-720	153	8.25	0.00150
25%	-760	192	14.00	0.00155
30%	-1125	231	14.10	0.00160
35%	-1130	270	14.20	0.00165
40%	-1135	309	14.30	0.00170
45%	-1140	348	14.40	0.00200
50%	-1200	388	16.50	0.00205
55%	-1350	427	16.75	0.00210
60%	-1400	466	17.00	0.00225
65%	-1500	505	17.25	0.00240
70%	-1775	544	17.50	0.00245
75%	-1850	583	26.50	0.00250
80%	-2100	622	28.00	0.00255
85%	-2475	661	30.00	0.00265
90%	-2600	700	35.00	0.00280
95%	-2700	1100	70.00	0.00290