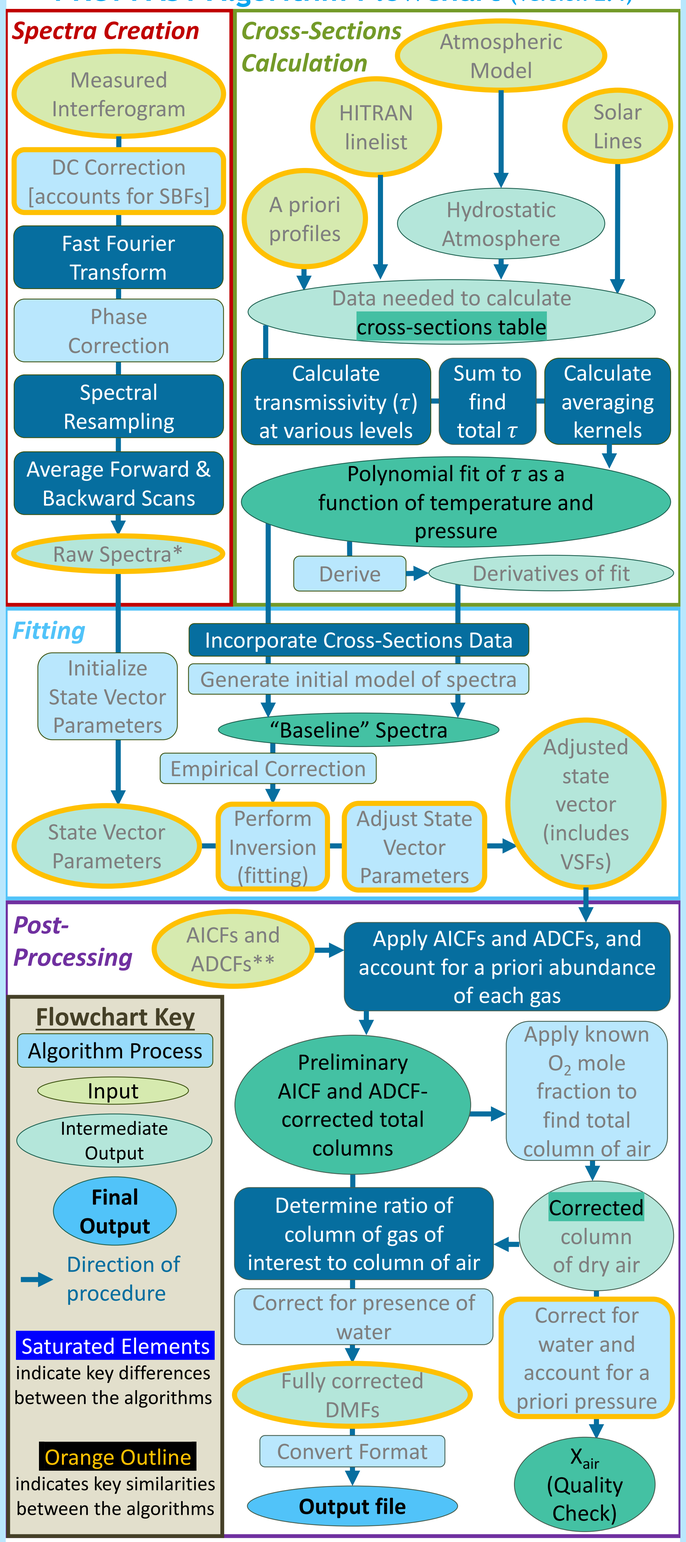
Comparison of Retrieval Algorithms for EM27/SUN Spectrometer Data

Noah A. Schneiderman[1]*, Elizabeth Spicer[1], Wesley T. Honeycutt[2]

1: University of Oklahoma, School of Meteorology; 2: University of Oklahoma, Gallogly College of Engineering
*Presenting author; contact: noahschneiderman@ou.edu

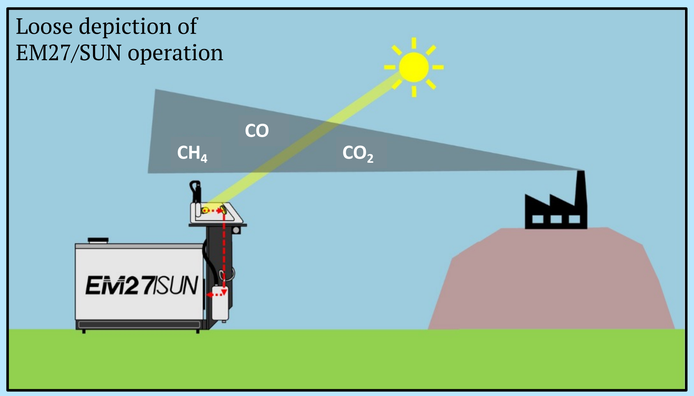
PROFFAST Algorithm Flowchart (Version 2.4)



*PROFFAST identifies flawed interferograms (called "laser sampling quality check" in GGG) and removes them as part of the process of creating spectra.
**Includes an additional ADCF exclusively applied to CO_2.

Key to Acronyms

DC: Direct current
GEOS-FP-IT: (NASA weather database)
VSF: Volume Mixing Ratio Scaling Factor
DMF: Dry-air Mole Fraction
ADCF: Airmass-Dependent Correction Factor
SBF: Source Brightness Fluctuation
HITRAN: High-resolution Transmission Molecular Absorption Database
X_{luft}, X_{air}: Dry-air Mole Fraction of Dry Air
AICF: Airmass-Independent Correction Factor



Abstract

There exist two networks that use ground-based infrared spectrometers to determine the abundance of trace gases in the atmosphere at various locations worldwide, each of which developed a separate retrieval algorithm for converting raw spectra data into processed dry-air mole fractions. Although these two algorithms should produce identical results, discrepancies have been observed between them. In this project, I conducted a deep analysis of the structure of each retrieval algorithm by performing an extensive literature review and confirming my findings with the code itself. I created a detailed flowchart depicting the operation of each algorithm, noting both major differences and key similarities between them. I also used each algorithm to analyze sample data collected with the University of Oklahoma's EM27/SUN instrument. In doing so, I observed a slight disparity between retrievals from the two algorithms, which worsened under nonideal conditions.

Introduction and Methods

Broad scientific consensus holds that anthropogenic greenhouse gas (GHG) emissions are causing global climate change at an unprecedented rate (Intergovernmental Panel On Climate Change 2023). As such, it has become more important than ever for the scientific community to have accurate measures of atmospheric GHG concentrations and distribution. Two networks, known as the Total Carbon Column Observing Network (TCCON) and the COllaborative Carbon Column Observing Network (COCCON), composed of Bruker 125HR FTS spectrometers and Bruker EM27/SUN spectrometers, respectively, monitor trace gases—principally carbon dioxide (CO_2), methane (CH_4), and carbon monoxide (CO)—measured as dry-air mole fractions (Wunch et al. 2011) (Frey et al. 2015). The retrieval algorithms used by each network to produce this information from raw data, GGG2020 and PROFFAST, respectively, differ slightly, causing inconsistencies in final results. Although relatively minor (Sha et al. 2020), these discrepancies are nonetheless problematic, as even a small error can hinder analysis of the minute spatial and temporal variations of GHG concentrations.

The literature review entailed mainly searching databases, and the flowcharts were created based on detailed notes on each section of the algorithm. When collecting sample data, we used a Bruker EM27/SUN spectrometer and Vaisala PTB330 with Campbell-Scientific CR6 data logger software.

Results and Discussion

My analysis of the algorithms yielded the flowcharts shown on either side. Some of the most notable differences between the algorithms include the order of steps in the interferogram-to-spectra conversion process, the differing methods of incorporating *a priori* data (GGG2020's forward model vs. PROFFAST's cross-sections table), and the order in which the post-processing steps are carried out.

In the sample data, I observed that retrievals taken with GGG2020 tended to be mildly higher than those taken with PROFFAST. We deployed the EM27/SUN on two days, March 25th and 26th, and the mean difference (GGG – PROFFAST) in the retrieved dry-air mole fraction of carbon dioxide (XCO_2) was 0.613 ppm on the 25th and 0.102 ppm on the 26th.

All in all, especially for the 26th, this is a remarkably small discrepancy considering the number of differences between the algorithms. The difference was likely worse on the 25th due to clouds hindering measurements more on that day; the variations in solar intensity could exacerbate the differences between the algorithms' methods of converting interferograms to spectra.

Conclusions

Although their structural differences can cause minor inconsistencies between them, GGG2020 and PROFFAST are both effective tools for the processing of EM27/SUN data. Further analysis of the code and more rigorous empirical testing is needed to definitively relate algorithm differences to observed discrepancies.

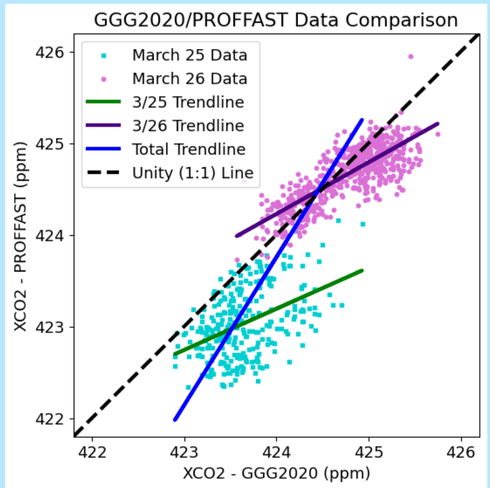
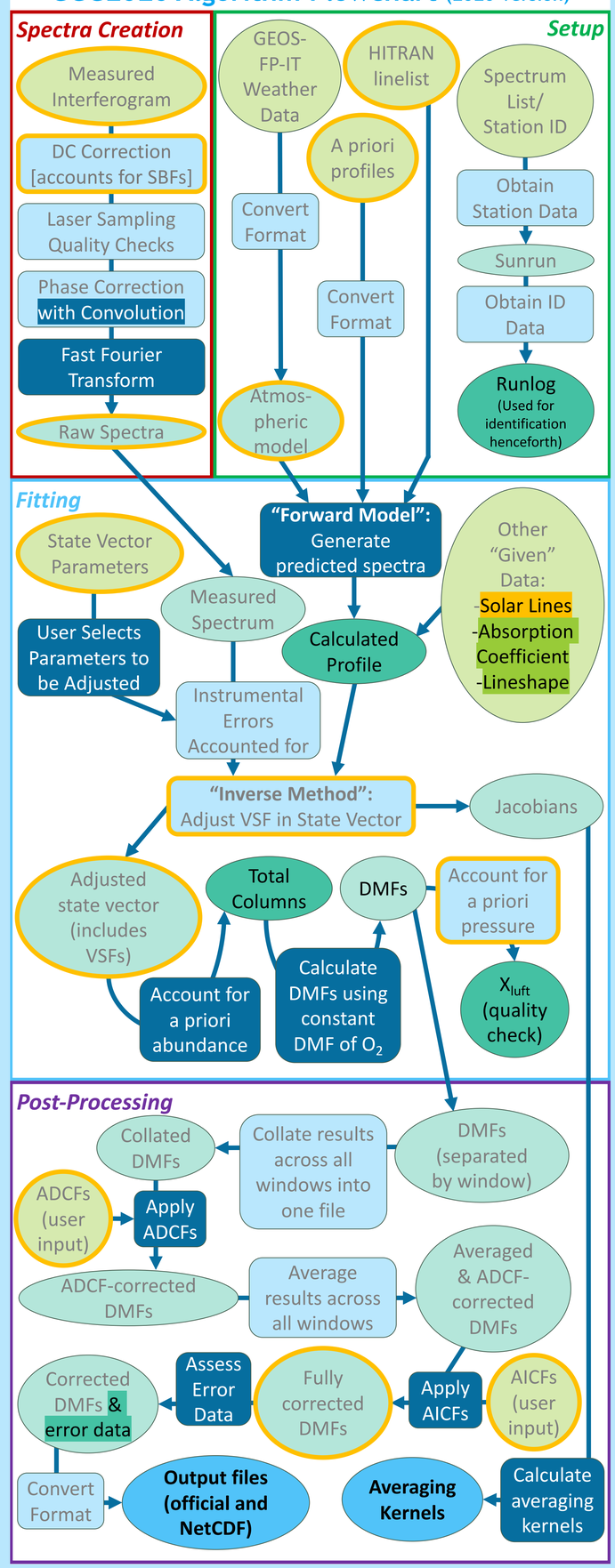


Figure: Comparison of the dry-air mole fraction of carbon dioxide (XCO_2) measured during the EM27/SUN deployment on March 25-26 as calculated with GGG2020 (x-axis) and with PROFFAST (y-axis). Identical retrievals would fall on the black dashed line.

GGG2020 Algorithm Flowchart (2020 Version)



References

All references can be viewed using this QR code.

Acknowledgments

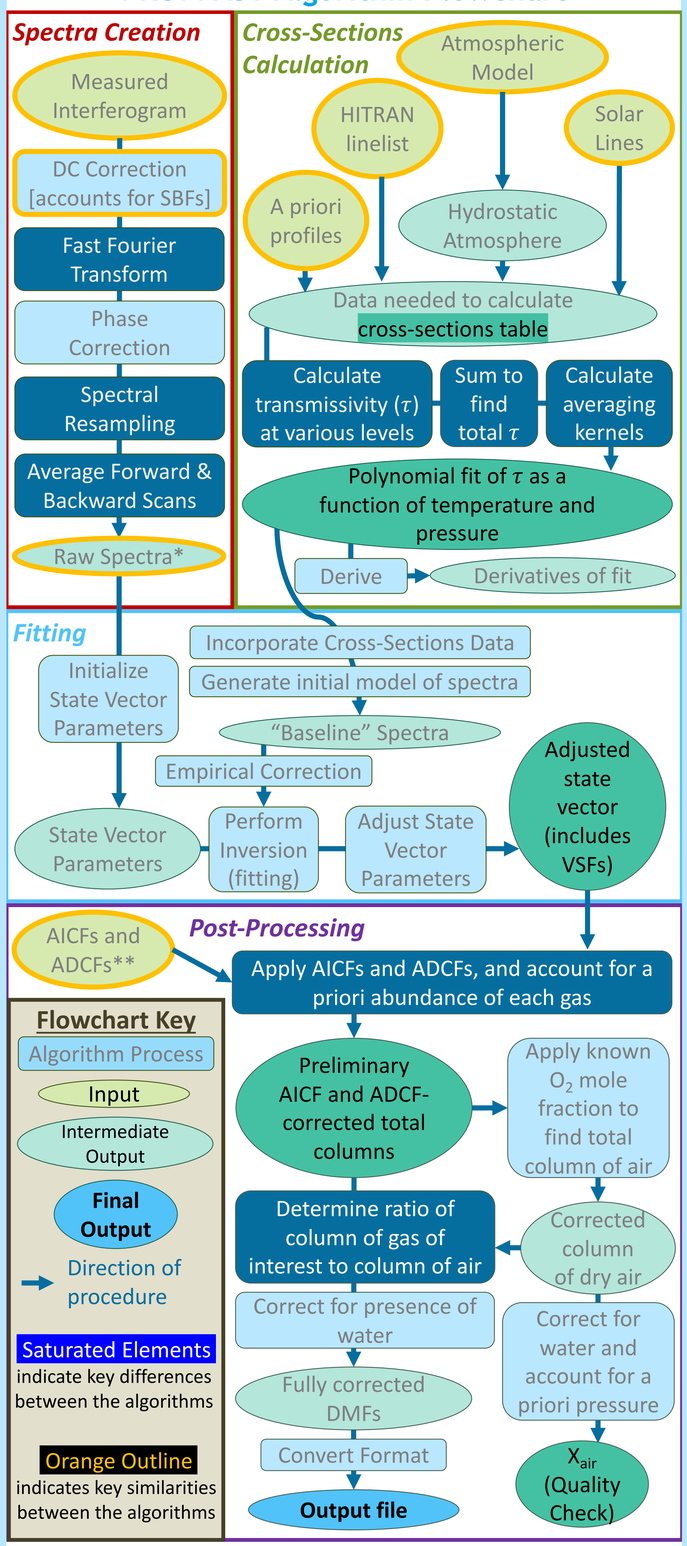
I would like to thank the University of Oklahoma's Honors College for organizing and funding the First-Year Research Experience (FYRE) program. I would also like to thank Annmarie Eldering for her role in guiding this project, and Nicole Jacobs for processing the data with GGG2020 on our behalf.

Comparison of Retrieval Algorithms for EM27/SUN Spectrometer Data

Noah A. Schneiderman[1]*,
Elizabeth Spicer[1], Wesley T. Honeycutt[2]

1: University of Oklahoma, School of Meteorology; 2: University of Oklahoma, Gallogly College of Engineering
*Presenting author; contact: noahschneiderman@ou.edu

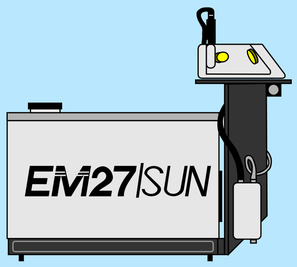
PROFFAST Algorithm Flowchart



*PROFFAST identifies flawed interferograms (called "laser sampling quality check" in GGG) and removes them as part of the process of creating spectra.
**Includes an additional ADCF exclusively applied to CO_2.

Key to Acronyms

DC: Direct current
GEOS-FP-IT: (NASA weather database)
VSF: Volume Mixing Ratio Scaling Factor
DMF: Dry-air Mole Fraction
ADCF: Airmass-Dependent Correction Factor
SBF: Source Brightness Fluctuation
HITRAN: High-resolution Transmission Molecular Absorption Database
X_{luft}, X_{air}: Dry-air Mole Fraction of Dry Air
AICF: Airmass-Independent Correction Factor



Left: The University of Oklahoma's EM27/SUN, shown operating in its enclosure. Taken March 25, 2024, from the roof of the National Weather Center.

Abstract

There exist two networks that use ground-based infrared spectrometers determine the abundance of trace gases in the atmosphere at various locations worldwide. Each developed a separate retrieval algorithm for converting raw spectra data into processed dry-air mole fractions. Although these two algorithms should produce identical results, discrepancies have been observed between them. In this project, I conducted a deep analysis of the structure of each retrieval algorithm by performing an extensive literature review and confirming my findings with the code itself. Using a detailed flowchart explaining the operation of each algorithm, I compared common elements between them. I also collected sample data with the University of Oklahoma's EM27/SUN instrument, and analyzed the raw data with each algorithm. I noted the bias of each method which **[insert result here].**

Introduction and Methods

Broad scientific consensus holds that anthropogenic greenhouse gas emissions are causing global climate change at an unprecedented rate (Intergovernmental Panel On Climate Change 2023). As such, it has become more important than ever for the scientific community to have accurate measures of atmospheric greenhouse gases' concentrations and distribution. Two networks, known as the Total Carbon Column Observing Network (TCCON) and the COllaborative Carbon Column Observing Network (COCCON), composed of Bruker 125HR FTS spectrometers and Bruker EM27/SUN spectrometers, respectively, monitor trace gases measured as dry-air mole fractions (Wunch et al. 2011) (Frey et al. 2015). The retrieval algorithms which produce this information from the raw data for each network, GGG2020 and PROFFAST, respectively, differ slightly. While the errors are relatively minor (Sha et al. 2020), they are nonetheless problematic, as even a small error can hinder analysis of the relatively minute spatial and temporal variations of greenhouse gas concentrations.

The literature review entailed mainly searching databases, and the flowcharts were created based on detailed notes on each section of the algorithm. When collecting sample data, we used a Bruker EM27/SUN spectrometer and Vaisala PTB330 with Campbell-Scientific CR6 data logger software.

Results and Discussion

My analysis of the algorithms yielded the flowcharts seen at the right. Some of the most notable differences between the algorithms include the order of steps in the interferogram-to-spectra conversion process, the differing methods of incorporating a priori data into the fit (GGG2020's forward model vs. PROFFAST's cross-sections table), and the order in which the post-processing steps are carried out. As for the sample data, I observed that **[difference between GGG2020 data and PROFFAST data].** Possible causes of this difference include **[potential causes of data discrepancies that may be observed when I process the data next week].**

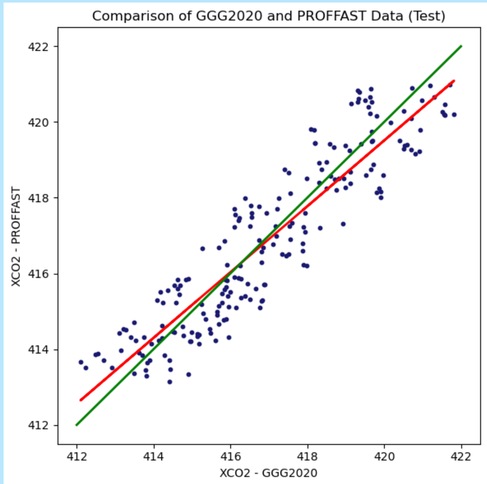
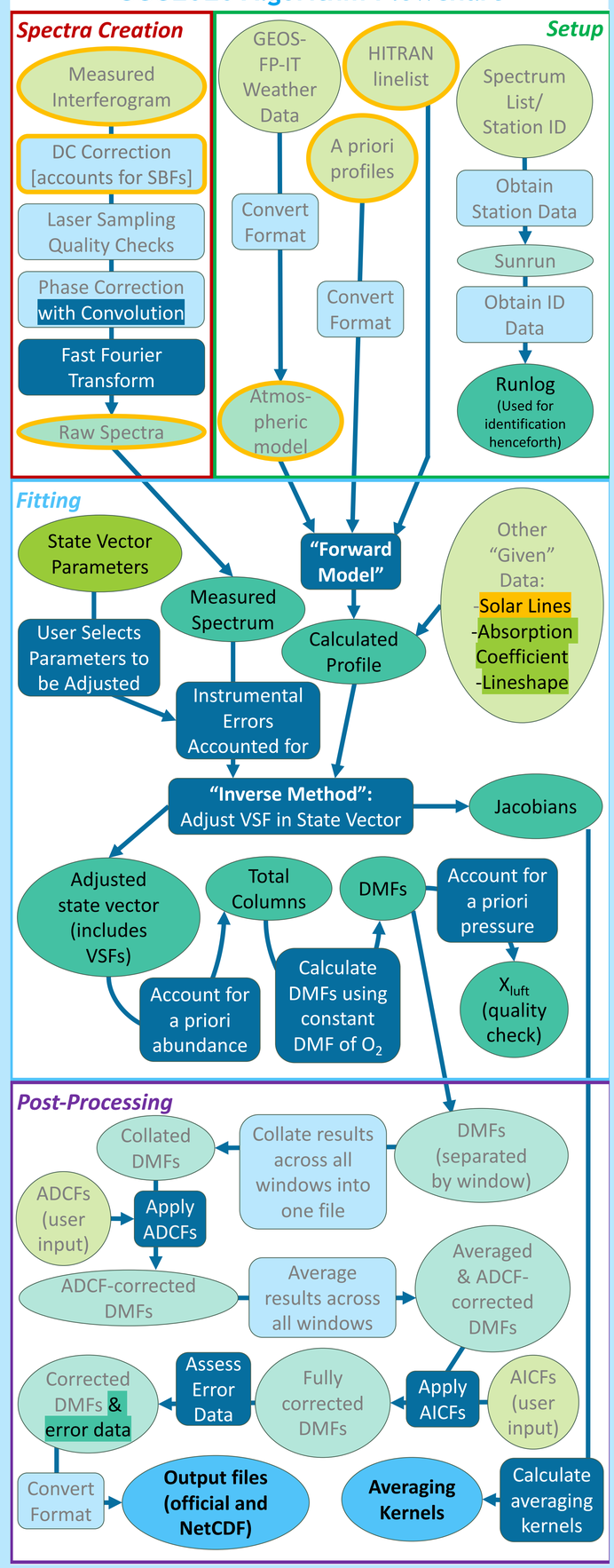


Figure 1: Comparison of the dry-air mole fraction of carbon dioxide (XCO_2) measured during the EM27/SUN deployment on March 25-26 as calculated with GGG2020 (x-axis) and with PROFFAST (y-axis). The red line represents the best fit between the datasets, while the green line represents the best-fit line for identical data (i.e., what the red line would look like if the algorithms produced identical data).

Conclusions

Although their structural differences can cause mild discrepancies between them, GGG2020 and PROFFAST are both effective tools for the processing of EM27/SUN data. [brief summary of results/discussion, suggestions for future work if applicable]

GGG2020 Algorithm Flowchart



Acknowledgments

I would like to thank the University of Oklahoma's Honors College for organizing and funding the FYRE program. I would also like to thank Annmarie Eldering for her role in guiding the project.

References

All references can be viewed here:

Comparison of Retrieval Algorithms for EM27/SUN Spectrometer Data

Noah Schneiderman*, Elizabeth Spicer, Wesley Honeycutt

Insert affilaitons Here

Abstract

There exist two networks that use ground-based infrared spectrometers determine the abundance of trace gases in the atmosphere at various locations worldwide. Each developed a separate retrieval algorithm for converting raw spectra data into processed dry-air mole fractions. Although these two algorithms should produce identical results, discrepancies have been observed between them. In this project, I conducted a deep analysis of the structure of each retrieval algorithm by performing an extensive literature review and confirming my findings with the code itself. Using a detailed flowchart explaining the operation of each algorithm, I compared common elements between them. I also collected sample data with the University of Oklahoma's EM27/SUN instrument, and analyzed the raw data with each algorithm. I noted the bias of each method which **[insert result here].**

Introduction and Methods

Broad scientific consensus holds that anthropogenic greenhouse gas emissions are causing global climate change at an unprecedented rate (Intergovernmental Panel On Climate Change 2023). As such, it has become more important than ever for the scientific community to have accurate measures of atmospheric greenhouse gases' concentrations and distribution. Two networks, known as the Total Carbon Column Observing Network (TCCON) and the COllaborative Carbon Column Observing Network (COCCON), composed of Bruker 125HR FTS spectrometers and Bruker EM27/SUN spectrometers, respectively, monitor trace gases measured as dry-air mole fractions (Wunch et al. 2011) (Frey et al. 2015). The retrieval algorithms which produce this information from the raw data for each network, GGG2020 and PROFFAST, respectively, differ slightly. While the errors are relatively minor (Sha et al. 2020), they are nonetheless problematic, as even a small error can hinder analysis of the relatively minute spatial and temporal variations of greenhouse gas concentrations.

The literature review entailed mainly searching databases, and the flowcharts were created based on detailed notes on each section of the algorithm. When collecting sample data, we used a Bruker EM27/SUN spectrometer and Vaisala PTB330 with Cambell-Scientific CRR6 data logger software.

Results and Discussion

My analysis of the algorithms yielded the flowcharts seen at the right. Some of the most notable differences between the algorithms include the order of steps in the interferogram-to-spectra conversion process, the differing methods of incorporating a priori data into the fit (GGG2020's forward model vs. PROFFAST's cross-sections table), and the order in which the post-processing steps are carried out.

As for the sample data, I observed that **[difference between GGG2020 data and PROFFAST data].** Possible causes of this difference include **[potential causes of data discrepancies that may be observed when I process the data next week].**

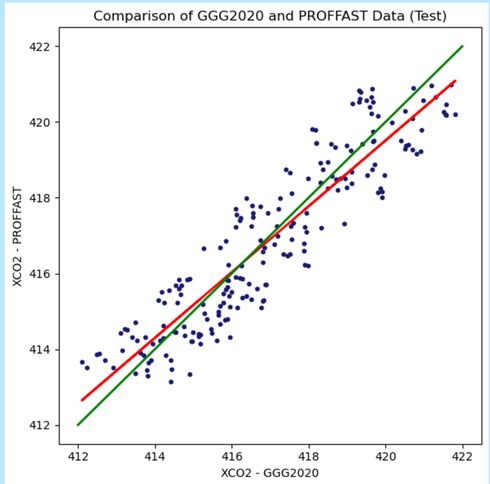
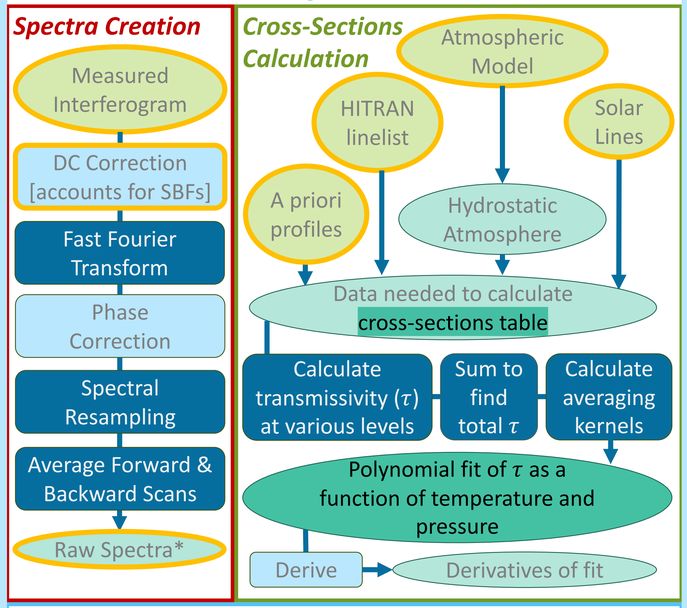


Figure 1: Comparison of the dry-air mole fraction of carbon dioxide (XCO_2) measured during the EM27/SUN deployment on March 25-26 as calculated with GGG2020 (x-axis) and with PROFFAST (y-axis). The red line represents the best fit between the datasets, while the green line represents the best-fit line for identical data (i.e., what the red line would look like if the algorithms produced identical data).

PROFFAST Algorithm Flowchart



*PROFFAST identifies flawed interferograms (called "laser sampling quality check" in GGG) and removes them as part of the process of creating spectra.

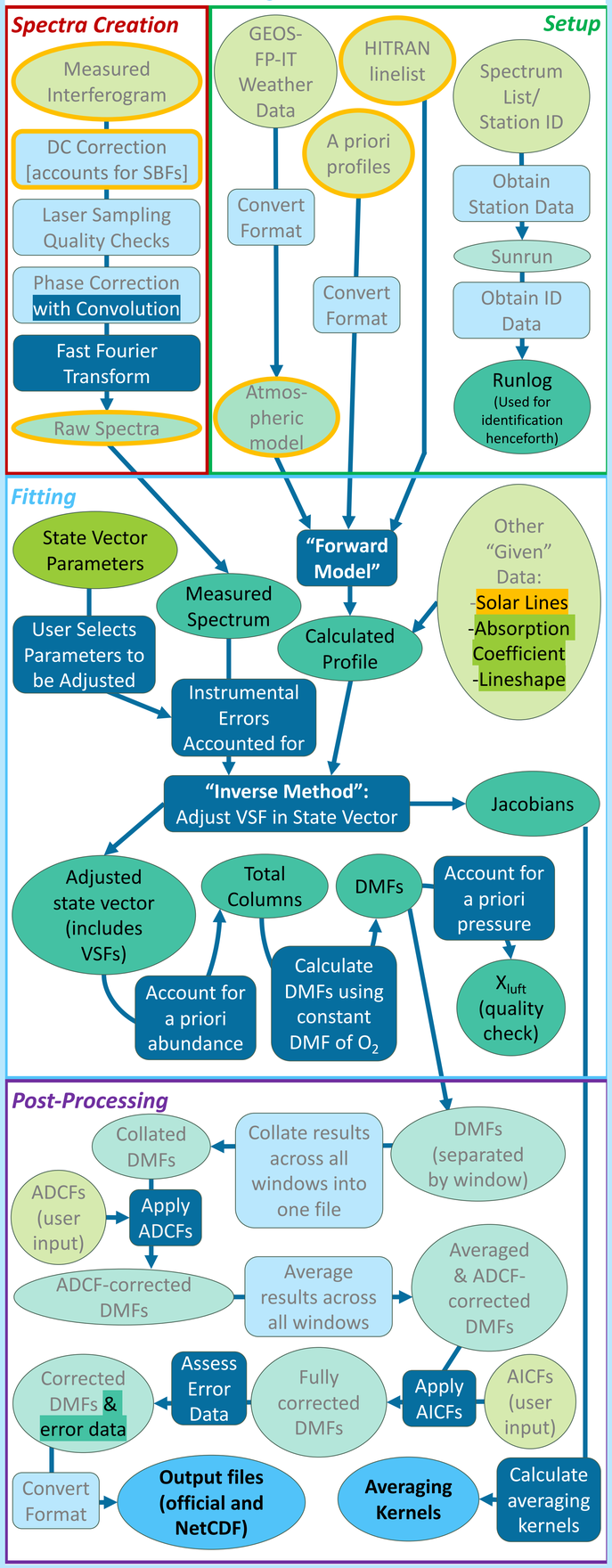
**Includes an additional ADCF exclusively applied to CO_2.

Key to Acronyms

DC: Direct current
GEOS-FP-IT: (NASA weather database)
VSF: Volume Mixing Ratio Scaling Factor
DMF: Dry-air Mole Fraction
ADCF: Airmass-Dependent Correction Factor
SBF: Source Brightness Fluctuation
HITRAN: High-resolution Transmission Molecular Absorption Database
X_{luft}, X_{air}: Dry-air Mole Fraction of Dry Air
AICF: Airmass-Independent Correction Factor

Picture of EM27/SUN deployment, if space permits

GGG2020 Algorithm Flowchart



References

Frey, M., and Coauthors, 2019: Building the COllaborative Carbon Column Observing Network (COCCON): long-term stability and ensemble performance of the EM27/SUN Fourier transform spectrometer. *Atmos. Meas. Tech.*, **12**, 1513–1530, https://doi.org/10.5194/amt-12-1513-2019.

Intergovernmental Panel On Climate Change, 2023: *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. 1st ed. Cambridge University Press,.

Sha, M. K., and Coauthors, 2020: Intercomparison of low- and high-resolution infrared spectrometers for ground-based solar remote sensing measurements of total column concentrations of CO2, CH4, and CO. *Atmos. Meas. Tech.*, **13**, 4791–4839, https://doi.org/10.5194/amt-13-4791-2020.

Wunch, D., and Coauthors, 2011: The Total Carbon Column Observing Network. *Phil. Trans. R. Soc. A.*, **369**, 2087–2112, https://doi.org/10.1098/rsta.2010.0240.

(Wunch et al. 2011)

Acknowledgments

I would like to thank my mentors, Dr. Wesley Honeycutt and Elizabeth Spicer, for all their help and support throughout this project. I would also like to thank the University of Oklahoma's Honors College for organizing and funding the FYRE program.