Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2020-177-RC1, 2020

Atmospheric Measurement Techniques Discussions
Open Access

Interactive comment on "XCO2 estimates from the OCO-2 measurements using a neural network approach" *by* Leslie David et al.

Anonymous Referee #1

Received and published: 7 July 2020

The authors describe in their manuscript the usage of an artificial neural network (ANN) to retrieve XCO2 and surface pressure from OCO-2 radiances. The topic fits well to the aims and scopes of AMT and is highly relevant because of the immense computational resources needed to process current and moreover future satellite data with state of the art full physics algorithms which are still prone to biases that require empirical corrections and have unknown origin. However, concluding my general and specific comments below, the presented material does not provide enough evidence to support the main conclusions namely that the results indicate that the ANN approach outperforms the operational NASA full physics algorithm and that it can be used to improve our knowledge of CO2 fluxes. It is understandable to limit the ANN development to simple cases at first (e.g., nadir only) and it is also not to be expected that the ANN will produce perfect results from the first try on. However, the functionality must be provable. This means it must become clear that the ANN has indeed learned and generalized primarily from the spectral information so that it is able to follow also unexpected CO2 features such as plumes. The presented material seems not sufficient to prove or disprove this. I'm sorry that I cannot give a more positive feedback, but because of this and the many open questions I cannot recommend a publication at AMT. However, due to the relevance of the topic, I encourage the authors to continue their work and, if the results allow it, to resubmit a revised manuscript.

General comments:

I have some concerns about the suitability of the input parameters to the ANN. As described in my specific comment L86-L88, potentially important information is rejected from the spectra.

The training data set spans over the same four-year (2015-2018) time period as the test data set and the authors emphasize that this allows XCO2 variations of about 2%. If it is important, that the training covers all possible CO2 concentrations, it is questionable, if a representative training data set can be found suitable also to analyze future OCO-2 data because CO2 is continuously increasing from year to year. What happens when applying the ANN to data from April 2020 including unprecedented large CO2 concentrations due to the continuous year-to-year increase of atmospheric CO2? Would the ANN still give reasonable results when training only with data from 2015?

ANNs are more or less black boxes in the sense, that it is not easily possible to find out which physical relationships they have learned. As CO2 is well mixed and long-lived, it is easy to make a relative good guess of its concentration without using any measurements. As an example, by estimating XCO2 only from latitude and time one can achieve already a good agreement with TCCON. Nevertheless, nothing new can be learnt about CO2 from such estimates. Therefore, it is crucial to prove that the ANN's XCO2 is primarily coming from the absorption depth within the spectra. This,

Printer-friendly version

Discussion paper

however, has not been done. One could confront the ANN with simulated spectra and show that it is indeed able to follow the simulated variations of the CO2. Such simulations could also be used to derive an estimate for the column averaging kernels, e.g., for different solar zenith angles. Additionally, one could apply the ANN to small scenes (with basically constant observation geometry) from which it is known that they include isolated CO2 plumes (see Nassar et al. (2017), or Reuter et al. (2019) for examples). Than it could be analyzed in how far the ANN is capable to follow the XCO2 enhancements and in how far the results agree with those of ACOS.

There are indications for over fitting: as the authors state, the maps in Fig.5 show biases in the test data set which do not exist for the training data set.

Conventional full physics algorithms allow (and usually require) post filtering by analyzing, e.g., the spectral fit quality but also quantities such as the posteriori XCO2 error estimate. Is this also possible for an ANN approach?

The TCCON validation seems to be not state of the art. For example, it does not consider the averaging kernels. For the ANN this is because they have not been computed, but for the TCCON and ACOS they are available. Which metrics have been used to quantify the performance? Usually, the average single sounding precision and the station-to-station biases are computed as a minimum set of parameters describing the quality. Please note that a low scatter could potentially also be observed when simply ignoring the satellite data which always add instrumental noise. This should also be considered/discussed when comparing CAMS (including no instrumental noise) with TCCON.

Please specify what is meant with the different measures of agreement that are used throughout the manuscript (trueness, accuracy, precision, skill, quality).

Specific comments:

L19: Please define what exactly is meant with precision (at least within the main text).

Is it, e.g., the standard deviation of the retrieved values or the standard deviation of the difference to a truth?

L42-L44: "Similarly..." This paragraph reads like a description of the light path proxy method, e.g., used by Schneising et al. (2008). However, the idea of full physics algorithms is fundamentally different to this. Because of wavelength dependencies of the surface reflection and of the scattering properties (e.g., optical thickness), the light path is different in the O2 and CO2 bands. Therefore, full physics algorithms use the O2 band to infer knowledge on the scattering particles or processes which allow to estimate the light path in the CO2 band from measurements in the O2 and CO2 band (see, e.g., Butz et al. (2011), Cogan et al. (2012), O'Dell et al. (2018), Reuter et al. (2017), Yoshida et al. (2013)). However, it is correct, that some (not all) full physics algorithms also retrieve the surface pressure.

L45: Molecular (Rayleigh) scattering is not a main difficulty as it is well known. The main problems are aerosols and clouds (see publications cited in my last comment).

L49: Please describe what is meant with "have been optimized".

L53: Why do you consider the signal to be the deviation from the prior? Shouldn't the signal be rather the actual variability.

L65-L67: RT models can simulate the radiance usually extremely accurate. However, the input to the RT models (e.g., unknown scattering phase functions, surface BRDF) and approximations needed to meet the requirements on the computational efficiency are the problems. Additionally, there may be unknown instrumental effects (uncertainties in the instrumental line shape function, polarization sensitivity, stray light, etc.).

L71: "The evaluation results show ..." should be moved to the discussion.

L76: Why footprint #4 not #7 or #3? Do the results critically depend on the used footprint?

L77: Please discuss if this issue principally will render ANN approaches impossible for glint observations and if not, outline potential solutions.

L80: Have you attempted to remove/mask the most affected pixels?

Sec.2: Please describe which OCO-2 data exactly has been used. Which version and where can be obtained from?

L86-L88: Dividing by the maximum is potentially not ideal because this maximizes the influence of instrumental noise or outliers due to cosmic rays and, additionally, it does not account for slopes in the spectrum. Such effects could be reduced by, e.g., dividing by the 90% percentile of the, e.g., 100 left-most spectral points. However, my main concern here is another: Dividing by the maximum radiance removes important information from the spectrum. Namely the information on albedo (as mentioned in the manuscript). As discussed in the literature (see provided references of full physics algorithms), unknown scattering properties introduce among the largest uncertainties in XCO2 retrievals. Knowledge of the albedo is important to infer knowledge of scattering properties. Consider an atmosphere with a surface pressure of 1000hPa and a scattering layer at 500hPa, reflecting 1% or the incoming radiation. Let CO2 absorb 80% of the radiance along the light path (sun-surface-satellite). This means about 40% will be absorbed along the light path sun-scattering layer-satellite. In the case of an albedo of 100%, the average absorption would be only slightly less than 80%. In the case of an albedo of 0%, the average absorption would amount 40%. This means, the relative depth is not a good measure for the number of particles in the total column. If you would normalize by the solar incoming radiance instead of the maximum radiance you would retain information on albedo, and therefore, also on scattering. Additionally, it shall be noted that, the light path in the O2 band can significantly differ from that in the CO2 band because of differences in the albedo and scattering properties.

L91: The influence of the azimuth should be discussed in the results section.

L95: 2557 input neurons are quite a lot and results in a rather complex ANN. Often one tries to reduce the dimensionality of the input data by performing, e.g., a PCA. This can probably also help the ANN to generalize instead of memorize. Why have you decided to use the full dimensionality of the input?

L96: Why 500 hidden neurons? Is there a rule of thumbs to select a suitable number of hidden neurons? Please discuss how your results depend on the complexity of the network topology. Do you use a so called bias neuron?

L100: Preventing the ANN from over training is certainly important. However, I'm not sure if it is a good idea to stop iterating before convergence is reached. Could the fact that overtraining happens for more than 100 iterations hint at a too complex network topology. Would it be an option to prevent from over fitting by choosing a less complex network with fewer hidden neurons? Additionally, a plot showing the convergence behavior would be nice to have (e.g., RMSD performance of the training and test data vs. number of iterations).

L103-L104: It is an important point whether the ANN uses information of time and position of the observation or not. Therefore, please discuss, in how far the observation geometry can provide the ANN indirectly with information on the position and/or time of the observations. Which parameters do you mean with "observation geometry" (is, e.g., surface elevation included)?

L104-L106: CO2 does not only have a seasonal variation, but it is also continuously increasing from year to year. Therefore, when having in mind a potential application to future data, the ANN will usually have to deal with concentrations larger than used for the training. How would the ANN behave in such cases? This could, e.g., be answered by confronting the ANN with RT simulated radiances.

L113: Why do you use ppm as unit for surface pressure?

L123: As mentioned in L121, surface based measurements have been used.

L125: Model pressure usually includes water vapor. However, XCO2 is the dry-air column-average. Therefore, strictly speaking, you would have to compute the weights

Printer-friendly version

Discussion paper

according to the pressure difference corrected for the water vapor content.

L128-L137: Interpolating only within the sorted spectrum is a good idea. However, the surface reflectance can introduce significant slopes within the spectra which may significantly change the rank of the spectral pixels. How large is the impact of this effect?

L166-L171, Fig.3: Why do you use for Fig.3 (left) only soundings where an ACOS retrieval has been made? This drastically reduces the number of cloudy soundings. From Fig.3 (left), I would estimate, that the ANN is capable to derive the surface pressure for definitively cloudy cases nearly bias free with a standard deviation of better than 5hPa. In cases of clouds that are not optically extremely thin, the spectra should not include significant information on the surface pressure. Additionally, the light path is shortened which can usually be interpreted as low surface pressure. Please discuss why the ANN is still capable to derive the surface pressure so well.

L175: I would suggest to also show the corresponding figure for XCO2. If it turns out that the ANN is also capable, to derive XCO2 in definitively cloud contaminated scenes, I would also suggest to add a discussion where this information is coming from.

L181-L190: How have you accounted for the column averaging kernels? At least for ACOS and TCCON, this information is available and should be used. Additionally, the usage of TCCON data should be described in Sec.2 and it should be mentioned in the main text, where the TCCON data can be obtained from and when they were downloaded.

L206-L210: I have some concerns with this arguing. If you chose a very complex network topology, the ANN might be well able to reproduce CAMS (including systematic persistent biases). If the ANN is too simple, it may not be able to follow the actual CO2 variability. I have the impression that this paragraph implicitly assumes that the network complexity is “just right” so that the ANN was not able to learn biases of CAMS but still generalized enough to follow the actual CO2.

L214: Please discuss why you observe a more or less persistent bias pattern in the even months but not significant differences in the odd months (used for the training). The fact that the training data set performs significantly better than the test data set usually hints at over fitting.

L215: In L206-L210 you suggest that XCO2 difference may come from model deficiencies. Why do you interpret the surface pressure differences as ANN biases?

Fig.5, 7, A1, A2: Please also show the even month used for the training because significant differences between even and odd months, can hint at a potential over fitting. The shown differences are in the order of 0.5%. What is the expected impact of neglecting the averaging kernels?

L237: The presented material does not allow this conclusion. Please, particularly, see my specific comments related to the validation method, the used input data for the ANN, and the lack of a prove that the ANN's XCO2 variability is indeed primarily coming from the spectral information.

L238-L239: Please define precision. For which product there is no independent truth (surface pressure or XCO2)?

L260-L263: I would have concerns with both options: i) If the OCO-2 spectra do not include information on, e.g., the upper most CO2, the ANN's AKs will have no sensitivity here independently from the used training data. ii) AKs usually differ from L2 algorithm to algorithm. I would suggest to compute typical AKs by confronting the ANN with simulated spectra for different observation geometries.

Technical corrections:

L11,43: column integrated CO2 dry air mole fraction -> column-average dry-air mole fraction of CO2

L13: uses a full -> is a so called full

L38: During -> Along

L63: Comprehensive -> Representative

L124: observations -> observation

Fig1: Please add a legend for the dashed/dotted yellow and red lines. The visibility of these lines is poor. The caption or the axis should include the information which models have been used? rˆ2 -> r2

L182, L184: TCCON network -> TCCON

L183: Fourier Transform Infrared -> Fourier transform infrared

L183: tuned against -> calibrated with

L184: "tuning" do you mean "bias correction"?

L186: "neither. . .nor" do you mean "either . . . or"?

Fig.3: Please increase the font size.

Fig.4: Please increase the font size. TCCON station names should start with upper case letters.

Fig.5, 7, A1, A2: The font size is too small, white is ambiguous (snow/ice or delta_p = 0), green is not explained.

References:

Butz et al. (2011): Butz, A., Guerlet, S., Hasekamp, O., Schepers, D., Galli, A.,Aben, I., Frankenberg, C., Hartmann, J.-M., Tran, H., Kuze,A., Keppel-Aleks, G., Toon, G., Wunch, D., Wennberg, P.,Deutscher, N., Griffith, D., Macatangay, R., Messerschmidt, J.,Notholt, J., and Warneke, T.: Toward accurate CO2and CH4observations from GOSAT, Geophys. Res. Lett., 38, L14812,https://doi.org/10.1029/2011GL047888, 2011

Cogan et al. (2012): Cogan, A. J., Boesch, H., Parker, R. J., Feng, L., Palmer, P. I.,Blavier, J.-F. L., Deutscher, N. M., Macatangay, R., Notholt, J.,Roehl, C., Warneke, T., and Wunsch, D.: Atmospheric carbondioxide retrieved from the Green-house gases Observing SATel-lite (GOSAT): Comparison with ground-based TCCON obser-vations and GEOS-Chem model calculations, J. Geophys. Res.,117, D21301, https://doi.org/10.1029/2012JD018087, 2012

O'Dell et al. (2018): O'Dell, C. W., Eldering, A., Wennberg, P. O., Crisp, D., Gunson,M. R., Fisher, B., Frankenberg, C., Kiel, M., Lindqvist, H., Man-drake, L., Merrelli, A., Na-traj, V., Nelson, R. R., Osterman, G. B.,Payne, V. H., Taylor, T. E., Wunch, D., Drouin, B. J., Oyafuso,F., Chang, A., McDuffie, J., Smyth, M., Baker, D. F., Basu, S.,Chevallier, F., Crowell, S. M. R., Feng, L., Palmer, P. I., Dubey,M., García, O. E., Griffith, D. W. T., Hase, F., Iraci, L. T., Kivi,R., Morino, I., Notholt, J., Ohyama, H., Petri, C., Roehl, C. M.,Sha, M. K., Strong, K., Sussmann, R., Te, Y., Uchino, O., and Ve-lazco, V. A.: Im-proved retrievals of carbon dioxide from OrbitingCarbon Observatory-2 with the version 8 ACOS algorithm, At-mos. Meas. Tech., 11, 6539–6576, https://doi.org/10.5194/amt-11-6539-2018, 2018

Nassar et al. (2017): Nassar, R., Hill, T. G., McLinden, C. A., Wunch, D., Jones, D., and Crisp, D.: Quantifying CO2 emissions from individual power plants from space, Geophys. Res. Lett., 44, 10045–10053, https://doi.org/10.1002/2017GL074702, 2017

Reuter et al. (2017): Reuter, M., Buchwitz, M., Schneising, O., Noël, S., Rozanov,V., Bovensmann, H., and Burrows, J. P.: A Fast AtmosphericTrace Gas Retrieval for Hyperspectral Instruments Approximat-ing Multiple Scattering – Part 1: Ra-diative Transfer and a Po-tential OCO-2 XCO2Retrieval Setup, Remote Sens., 9, 1159,https://doi.org/10.3390/rs9111159, 2017

Reuter et al. (2019): M. Reuter, M. Buchwitz, O. Schneising, S. Krautwurst, C.W. O'Dell, A. Richter, H. Bovensmann, and J.P. Burrows: Computation and analysis of at-mospheric carbon dioxide annual mean growth rates from satellite observations during

2003-2016, Atmos. Chem. Phys., https://www.atmos-chem-phys.net/19/9371/2019, 2019

Schneising et al. (2008): O. Schneising, M. Buchwitz, J. P. Burrows, H. Bovensmann, M. Reuter, J. Notholt, R. Macatangay, T. Warneke: Three years of greenhouse gas column-averaged dry air mole fractions retrieved from satellite - Part 1: Carbon dioxide. Atmospheric Chemistry and Physics, doi:10.5194/acp-8-3827-2008, 8, 3827-3853, 2008

Yoshida et al. (2013): Yoshida, Y., Kikuchi, N., Morino, I., Uchino, O., Oshchepkov, S.,Bril, A., Saeki, T., Schutgens, N., Toon, G. C., Wunch, D., Roehl,C. M., Wennberg, P. O., Griffith, D. W. T., Deutscher, N. M.,Warneke, T., Notholt, J., Robinson, J., Sherlock, V., Connor, B.,Rettinger, M., Sussmann, R., Ahonen, P., Heikkinen, P., Kyrö,E., Mendonca, J., Strong, K., Hase, F., Dohe, S., and Yokota,T.: Improvement of the retrieval algorithm for GOSAT SWIRXCO2and XCH4and their validation using TCCON data, At-mos. Meas. Tech., 6, 1533–1547, https://doi.org/10.5194/amt-6-1533-2013, 2013

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2020-177, 2020.