Dear Piet

We understand it is sometime difficult to obtain answers from the reviewers, and we appreciate your work.
As an answer to your message, we have added in our answers to the reviews below a clear identification of the changes that have been made in the manuscript to answer their comments. The manuscript in track-change mode is very similar to the one we send last August, although we have added a few sentences on lines 313-327.

We feel that we have accounted for most of the reviewer recommendations and clearly justified why we did not for the few other ones, and we hope you agree so that the paper can now be published

With best regards

Leslie David, Françius-Marie Bréon, Frédéric Chevallier

Answer to reviews

We warmly thanks both reviewers for their work, their careful analysis, and their suggestions that, we hope, led to an improve version of the manuscript.
We agree with most comments, but not all of them as detailed in the following.

In the following, the reviews are in Italic whereas our answers are in plain text

Reviewer 1

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 un-expected 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.

We have added an analysis (in the first two paragraphs of section 4 along with the figure 8) to demonstrate that the Neural Network retrieves the surface pressure and XCO2 from the spectral information and not by simply reproducing the training material.

Conversely, we do not follow the reviewer suggestion to analyze fine scale structures such as plume. First, our first NN attempt only uses a single FOV rather than the 8 cross-track FOV, so that we lack the imagery capabilities that are most useful to identify plumes. Second, this would be a full additional study and is clearly out of the scope of this first paper on the subject.

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.

See our answer to the specific comment

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 largeCO2 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.

The NN approach must use a training dataset that is representative of the observations to which it is applied. Thus, the NN that we have trained with 2015-2018 data would generate poor results when applied to 2020 observations. We acknowledge that this is a limitation of the approach. However, there has not been real time application of OCO-2 observations so far, so that we do not see that as a strong limitation. We have added a brief paragraph in this sense in section 4 (lines 324 to 327)

We certainly agree that the approach is a black-box so that its capabilities can only be judged empirically. This is what we attempt in the paper.

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, however, has not been done.

We certainly agree that a good a priori can be obtained from the latitude and time. We are well aware of this fact and this is why we do not provide any information on the location and time as input to the ANN. We agree that some indirect information about the latitude and day-in-the-year may be extracted from the observation geometry (We clarified this point in lines 112 to 115 of the revised manuscript). Conversely, there is no such indirect information on the longitude and the year of observation (i.e. the viewing geometry is the same from one year to the next, and for all longitudes at a given time). Thus, the ability of the NN to accurately retrieve the ≈2 ppm increment from year to year as well as the zonal gradients is necessarily derived from the spectra and not from the other input data. We explained this in details in the second paragraph of section 4 (starting at line 278).

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 theXCO2 enhancements and in how far the results agree with those of ACOS.
We have added a section to demonstrate that the NN does not only mimic the training data and provides additional information (see the first two paragraphs in section 4 and the figure 8). Conversely, we have not followed the reviewer suggestion to analyze XCO2 plumes, as discussed above.

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.
There is little doubt that there is over fitting of the training dataset (as we mentioned in line 106) and this is why we pay attention to have independent training and evaluation data.

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?
No. The NN approach is a black box and it does not allow, to our knowledge, a posterior error estimate. This is made clear on lines 322-323

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).
We agree that we do not use the averaging kernel as this information is not directly available for the NN. We have attempted to improve the manuscript and make clear the various comments of the reviewer stated here (this is discussed in lines 313-321 of the revised manuscript)

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?

We feel that there is a clear definition to precision that refers to the standard deviation to some other measure. Here, we do not want to use the term accuracy as we recognize there may be some bias in the validation dataset that we use. We have made a few changes in the text at line 294 to remove ambiguity

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 theO2 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.*

We certainly agree with this description of the XCO2 retrieval algorithm. However, our text only gives a rough introduction and provides a reference for a more detailed description for interested readers. We do not feel there is anything misleading in the two sentences

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).

We agree. The use of "molecules" was inappropriate in the sentence and has been changed to "particles" (see line 45 of the revised manuscript)

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

We meant "estimated for a best fit between the measured and modeled spectra". The sentence has been changed to "The radiative transfer models that are used for the retrieval leave significant residuals between the measured and modelled spectra, even after the XCO2 and aerosol amount have been inverted for a best fit" (see line 49)

L53: Why do you consider the signal to be the deviation from the prior? Shouldn't the signal be rather the actual variability.
The deviation from the prior is the innovation, i.e. the information that is really brought by the observation. If the variability is already known (as for the growth rate or even a large fraction of the seasonal cycle), it cannot be considered as a signal brought by the observation.

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.).
Agreed. We added "In addition, there may be some wrong assumptions and unknown instrumental defect that are not accounted for in the forward modeling." (lines 67-69)

L71: "The evaluation results show…" should be moved to the discussion.
Agreed (done)

L76: Why footprint #4 not #7 or #3? Do the results critically depend on the used footprint?
There is no reason to select one rather than the others. We have not analyzed whether the results depend on the footprint but see no reason why it would.

L77: Please discuss if this issue principally will render ANN approaches impossible for glint observations and if not, outline potential solutions.
The text clearly says that the doppler effect that is significant for the glint -but not so for nadir-observation introduce a complication (see first paragraph of section 2). Whether this makes the ANN approach impossible for the glint observation cannot be answered before it is attempted.

Since then, we have had excellent results with the glint observation, but this is not in the scope of the present paper.

L80: Have you attempted to remove/mask the most affected pixels?
Yes. We have removed the pixels on the edge of the spectra, as mentioned in the text (see line 86 and following)

Sec.2: Please describe which OCO-2 data exactly has been used. Which version and where can be obtained from?
Quality flags and XCO2 estimates are from Lite V9r whereas the spectra are obtained from product L1B v8r. This information is added to the manuscript (lines: 52, 116, 122 and 160)

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.

We agree that the results presented in the paper are for a first attempt that is open to improvements. In practice, we do not divide by the maximum, as incorrectly mentioned in the text, but by something that is similar to the reviewer suggestion: The normalization is based on the mean radiance of the 90-95 percentile range. This is now properly mentioned in the text at line 90.

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 1000 hPa 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.

Although we agree with the reviewer comments, NOT dividing by the albedo also causes issues. We felt that it would help the NN training to use input radiances that vary little outside of the absorption band. Our results indicate that fairly accurate results are obtained with such choice. We certainly agree that the NN approach can be further improved in the future, by us or others. Still, the results obtained with our configuration are, we feel, sufficiently interesting and novel to deserve publication

L91: The influence of the azimuth should be discussed in the results section.
We felt it fits better as it is. We have not moved this very short discussion

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?

There is no indication that the PCA does a better job than the ANN itself to retain all the information that is available. Use the full dimensionality of the input, which retain all potential information, seems like a natural choice to us

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?

We made a few attempts with a different number of neurons. With 50 neurons, the results were clearly of lower quality. With a larger number of neurons, the training time became significantly larger. The number of neurons could be optimized in the future. We added two sentences starting at line 101 in the manuscript.

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).

Note that there is never a full convergence of the network so that a subjective choice of iteration stop is necessary. We do make such convergence plots (see below). We did not feel it provide significant information. To follow the reviewer suggestion, the convergence plot was added to the appendix (figure A1) and mentioned in the text at line 104 and following.

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.

The observation geometry varies with the latitude and the season so that the NN may infer some location information from this input. Conversely, it is the same for one year to the next and, at a given date, for all longitudes. Thus, there is no information on the longitude or the year of observation in the geometry parameters that are provided to the network. We clarified this point at the lines 110 to 115 of the revised manuscript.

Which parameters do you mean with "observation geometry" (is, e.g., surface elevation included)?

The observation geometry is the sun zenith angle and the relative azimuth. The surface elevation is not included (this has also been clarified at line 1110)

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.

We certainly agree that the NN approach cannot be used to process observations that have been obtained later (or earlier) in time than the training dataset. This point is added in the discussion section around line 325.

L113: Why do you use ppm as unit for surface pressure?
Typo corrected. Thanks

L123: As mentioned in L121, surface based measurements have been used.
This was a mistake. We meant that no TCCON (surface based remote sensing) observations have been used. It is corrected. Thanks

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 according to the pressure difference corrected for the water vapor content.
Although it was not explicit in the manuscript, all pressures in the atmospheric transport model that our CAMS product uses (LMDz) are dry air pressures. Thus, there is no need for a correction for the water vapor. We have made it clear in the revised version by adding the sentence "Note that the model layers use "dry" pressure coordinates so that there is no need for a water vapor correction in the vertical integration" (line 143 and following).

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?
We have not attempted to quantify 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.
We have acknowledged that our current version of the NN is not independent from the ACOS product. In particular, it uses the results from the Cloud detection algorithm. Only the observation that have been through this detection are used. The "Definitely Cloudy" soundings that are used here have been declared "clear" by the first cloud detection process. There is little doubt that the results would not be the same for the observations that are truly cloudy. Note also that the "cloudy" observations in Figure 3 are rare.
We have added a clarification in the text: "Note also that the observations used hare have all been classified as "clear" by the ACOS pre-processing. Thus, most OCO-2 observations are not used here and Figure 3 should not be interpreted as the ability to retrieve the surface pressure in cloudy conditions"

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.
See comment above that the "cloudy" observations here is only a very small fraction -probably very specific- of the cloudy observations

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.
We have not used the averaging kernels.

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.
This information was added in the legend of Table 1

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.
We certainly do not claim that the NN complexity is "just right". We only give the results for the configuration that we have chosen. As for the argument, it seems clear that if the difference between CAMS and the NN have a random structure, they cannot be used to improve the flux estimates. If they have a spatio-temporal structure, then there is some hope towards that objective. Two sentences have been added to manuscript at line 238 to clear up.
Note also the discussion in the next paragraph that the structures in the surface pressure are a bad signal toward that objective

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.
There is no doubt that there is over-fitting of the data that are used for the training. This is inevitable and this is the reason why we attempt to clearly distinguish the training and evaluation dataset.

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?
Although we do expect biases in the XCO2 modeling (because our knowledge of the surface fluxes is far from perfect), we do not expect such biases in the numerical weather modeling of the surface pressure. Numerous studies have shown that the surface pressure accuracy is better than 1 hPa.

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.
There is over-fitting of the training dataset. This is inherent to the method. As a consequence, the even month differences are very small and provide no relevant information

The shown differences are in the order of 0.5%. What is the expected impact of neglecting the averaging kernels?
Unfortunately, we do not have this information

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.

We have added some material to show that the information comes primarily from the spectral information. We hope this material is sufficient to convince the reviewer that, indeed, we have enough information for the conclusion that the NN approach allows a high precision. Also note the use of the verb "indicate" and not "demonstrate"

L238-L239: Please define precision. For which product there is no independent truth (surface pressure or XCO2)?
This comment applies to both, at the scale of the FOV footprint. The relative accuracy of the "truth" surface pressure (numerical weather mode) is certainly better than that of the XCO2. We have changed the sentences to
However, there are indications that the accuracy on the surface pressure is better than 3 hPa RMS, while the precision (StdDev) of XCO2 is better than 0.8 ppm. Indeed, the data used for the product evaluation has its own error that is difficult to disentangle from that of the estimate based on the satellite observation.

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.
We have added a sentence line 320 as per the reviewer suggestion. The various options must be analyzed and tested, but it is clearly out of the scope of the present paper. We agree the reviewer suggestion has potential, but may lead to significant computer requirement.

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
Done (all of the above)

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? r^2 -> r2
Corrected

L182, L184: TCCON network -> TCCON
L183: Fourier Transform Infrared -> Fourier transform infrared
Corrected (both)

L183: tuned against -> calibrated with
"Calibration" assumes that experimental conditions are controlled ("under specified conditions" see https://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2012.pdf), which is not possible with remote sensing. "Tuning" is appropriate here.

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

L186: "neither...nor" do you mean "either...or"?
No. we mean the target mode that is neither nadir nor glint (and that is difficult to handle with the NN approach)

Fig.3: Please increase the font size.

Fig.4: Please increase the font size. TCCON station names should start with uppercase 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.
All corrected

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 fromGOSAT, Geophys. Res. Lett., 38, L14812,https://doi.org/10.1029/2011GL047888,2011C9

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 TCCONobser-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 version8 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 Retrievalfor 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. Bovens-mann, M. Reuter, J. Notholt, R. Macatangay, T. Warneke: Three years of greenhousegas column-averaged dry air mole fractions retrieved from satellite - Part 1: Carbondioxide. 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., Sher-lock, 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 re-trieval algorithm for GOSAT SWIRXCO2and XCH4and their validation using TCCONdata, At-mos. Meas. Tech., 6, 1533–1547, https://doi.org/10.5194/amt-6-1533-2013,2013Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2020-177, 2020.

Reviewer 2 (Chriss O'Dell)

Review of David et al., "XCO2 estimates from the OCO-2 measurements using a neural network approach.", by Chris O'Dell.

This work details a fast, artificial neural network (NN) approach to retrieving surface pressure and the column-mean dry air mole fraction of CO2(XCO2) from high-spectral resolution measurements in the near-infrared from the Orbiting Carbon Observatory-2 (OCO-2). Traditionally, the most accurate XCO2 retrievals have been from semi-physical("Full-physics") retrievals. These are typically iterative, and typically include accurate calculations of multiple-scattering from thin layers of clouds and aerosols, which makes them exceedingly slow. They also tend to be subject to importance biases (of order 1ppm) due to forward model errors(such as spectroscopic or instrument calibration). A neural-network approach is extremely appealing because it automatically solves the speed problem (NN's are very fast, likely tens of milliseconds per retrieval, vs. minutes for a typical FP approach), and may solve some of the bias problem as well, because they simply train on the "right answer", and do not have to know the details of spectroscopy or instrument models.

We fully agree with this summary and comment

General Comments:
This work is the first serious effort to use NN's as applied to the XCO2 retrieval problem, and the author's mostly do a good job. However, there are a number of weaknesses and methodological problems in this work that need to be strengthened before I can recommend publication. I see this is a "major revision", but ultimately I believe this work can and should be published, as I believe (for the reasons above) that NN's hold great promise as applied to the XCO2-from-satellites problem.
We hope the reviewer can be convinced by our additions and corrections in the paper, together with some of our answers below

Most significantly, as the first reviewer pointed out, it is difficult to ascertain from the manuscript alone exactly what the NN algorithm has learned. The authors trained it on a model (CAMS), and as a first validation, tested it against the same model. Using alternating months as training vs. testing is helpful, but the model certainly has deficiencies that are persistent longer than a month in certain regions, so testing well vs. that same model simply is not a validation. The only other real validation given is against TCCON, which seemed to perform well but because of TCCON's lack of good global coverage, again it is very hard to tell how well the model performs globally in any real sense.
We have added a discussion and further analysis (beginning of the discussion section) to answer the reviewer comment

I am also worried about having to re-train the model every year to deal with the ~2 ppm secular increase, which over a mere 4 years is roughly equal to the entire global variability of XCO2.
This is definitely a weakness of the NN approach that we do not try to hide. Unfortunately, we see no way to fix this weakness. Note however that most scientific studies achieved with the OCO-2 dataset apply to "old" observations so that the ability to process the data in near real-time does not appear essential. We have added a paragraph in the discussion section on this comment.

I believe it would make their argument much stronger if they ran their algorithm on a few select powerplant cases where the enhancement in the plume is reasonable well-constrained. This is possible for power plants with good bottom-up emissions estimates, and cases where the wind speed is reasonably well-known. See for instance Nassar et al. (2017) for some sample cases. If the NN doesn't see a plume at all, we know that it hasn't been properly trained; if it does, it will dramatically bolster the arguments in this work. Arguing that just because the NN doesn't have direct access to location or time information does not mean it cannot indirectly learn other relationships that allow it to appear to learn well. This is a hypothesis, not a proof that it has learned what you think it did.

Although we agree that the analysis over plant plumes is something that must be done eventually, it is clearly out of the scope of the paper that is a first try at using the NN approach. As described in the manuscript, we use a single footprint so that our output dataset does not have the "imagery" information that is most useful for the plume. In addition, the identification of plumes in the observation dataset, the use of an atmospheric transport model, and the analysis of each case would be a full study in itself.

Related to this, I'm somewhat concerned by training on CAMS and then considering to use the resulting XCO2 to correct CAMS. To show that your method works, you almost need to run a full (and fairly complicated) OSSE where you have a true world, a CAMS-like model world with some CO2 errors (spatially and temporally correlated), train on the latter, and then see if you can recover the former with the NN. I don't see how this is guaranteed to work, honestly. How do you know that you won't somehow reproduce systematic errors in CAMS by using the NN approach? You state in the text that you tacitly assume that CAMS errors are not correlated OCO-2 spectra in given areas and for given months. But because the CAMS errors (likely of order 1 ppm) are of a similar magnitude as the XCO2 signal, it is important to point out that this is merely an assumption, and more extensive validation (or a detailed OSSE study) is necessary to prove it.

To answer this reviewer concern (that are somewhat similar to those of the first reviewer), we have added a full section in the paper (lines 278-290, and figure 8). We hope it can convince the reviewers

Also, you claim to use the "ACOS cloud flag", which you say has values 0,1,2, or 3, as a way to define both your training and testing data sets. I think you mean "Preprocessing Results/cloud_flag_idp" in the L2Std file. If this is correct, please know that this flag is little used by the community. In fact, I've never heard of anyone using it, actually. It was defined about a decade ago for GOSAT and not really touched since then (I verified this with the author of the code that defines it). It has never been carefully validated and it appears to be extremely restrictive ("co2_ratio_idp"must be between 0.99 and 1.01 to pass, which is extremely restrictive and appears to cutout entire regions of the globe). Further, using outcome_flag=1 is also quite restrictive. Can you please comment on these flags, and why you didn't use the far simpler ACOS xco2_quality_flag, which is widely used by the community and is the generally adopted quality flag to use? In the plot below, I have attempted to show the differences between the two approaches for May 2016. I had to match the L2Std files(v8r) to the Lite files(v8), so there may be some differences to what you used in your work, but the general conclusion is that you miss a great deal of data with the highly restrictive data set you are working with. Thus, because it is so restrictive, it may be a far easier task than what ACOS tries to do, which is get the best error possible for the xco2_quality_flag = 0 dataset, which is roughly 6 times bigger.

We agree that our description of the cloud flags and quality flags used in our study was unclear. We have attempted to correct that.

In practice, we do use only observations with xco2_quality_flag=0. This was not mentioned in the manuscript which is a clear oversight (corrected)
The selection Outcome_flag=1 is used for the training. We have analyzed the Psurf precision statistics as a function of the outcome flag (Figure 3). Since there are no significant difference, the evaluation data and the maps are based on the data with no restriction on the outcome flag. As for the cloud flag, we indeed refer to the Preprocessing Results/cloud_flag_idp. We were not aware that this flag was not used by the community. For the training, we only use cloud_flag=3 (absolutely clear) which is roughly half of the dataset. The analysis of the result precision as a function of this flag (Figure 3) indicates that the results obtained for cloud flag=2 are not significantly different than those with cloud flag=3. On the other hand, those with cloud flag=0 or 1 are significantly poorer. As a consequence, we have retained the observations with cloud flag of 2 or 3, which are 96% of the dataset. Although the cloud flag is not used in the community (as explained to us by the reviewer) it seems that it has some value (cloud flag =0 or 1 is of lesser quality). In the previous version of the manuscript, we used, for the evaluation, only data with cloud flag =3. As a consequence, the number of soundings is significantly increased (a factor of ≈2) and the statistics are slightly changed.
This is described in lines 121-126

Finally, in section 2 please give the sources of ACOS/OCO-2 data you used with more specificity. What specific versions and datasets of OCO-2 did you use? V8r, or just V8? Did you use L2Std files, L2Diafiles, Lite files, etc?
This information is now provided line 120 (L2 Lite V9r and the associated warn level and flags from v8r L2lite, and L1b from v8r)

What you train on is pretty critical. I think you should at the very least show a sounding density map of your training (and testing) set.
The map requested by the reviewer has been added in the supplementary (figure A2)

Further, I think you should carefully explain your reasoning on how you choose the filtering. You must at least mention the xco2_quality_flag, and ideally you would retrain (or at least test) using this, if you aren't going to define your own quality flags. If you choose to train using very restrictive (clear-sky conservative) filters, please explain this is more detail. Also, both outcome_flag and warn_level (which you use for filtering) come from the ACOS L2FP product (cloud_flag_idp comes from a fast, preprocessor code, the IMAP-DOASPreprocessor, or IDP). It would be much better if you could avoid this entirely, because currently you are throwing away all the soundings that didn't converge or were skipped by the ACOS team, which relies on all the peculiarities of our specific algorithm. To make a useful NN algorithm, it ultimately must be independent of any full physics algorithm, unless you want to train on soundings that pass some smaller subset of data that includes L2FP quality flags, but test on a more complete set of soundings that doesn't use any L2FP quality flags. But you do not appear to do this.
We certainly agree that, eventually, it will be necessary to become independent of the ACOS preprocessing. This is mentioned explicitly in the discussion section. Conversely, at this stage, we feel it is desirable to process the exact same dataset as that successfully processed by ACOS. This allows a meaningful comparison of the product precision and accuracy.
We have added a few sentences to explain which flags are used, and why we selected those at the beginning of page 4 of the revised manuscript.

Specific Comments:

L20, Abstract: I don't think TCCON is a "sunphotometer". That kind of implies more moderate resolution measurements. How about "reference ground-based spectrometer" or something similar?
We changed photometer to spectrophotometer. We do not think there is any need to be more specific in the abstract. Most readers will know exactly what we are referring to and those who do not can get the information in the paper.

L80: Please clarify "a limited set of spectral elements". Make clear these are the solar (Fraunhofer) lines you're talking about.
Done

L85(near there): Do you try to mask deep solar lines as well (to remove their Doppler effects)? Please clarify, with a why or why not.
We have added the sentence : "Conversely, we do not remove the spectra that are affected by the deep solar lines, and let the NN handle these specific features".

L90 (near there): Might you include the polarization angle directly to the NN, in addition to / instead of the relative azimuth? That might work even better.
Thanks. We take this suggestion for a future evolution of the algorithm.

L96: As the readers of this article are likely not NN experts, please discuss the pros and cons of # of hidden layers vs. # of neurons. Also please discuss how was the number 500 chosen or optimized.
An earlier version of the NN approach used only 50 hidden neurons and led to less accurate results. A higher number or neurons led to much higher training time. We made a few tests, but there was no attempt for an exhaustive analysis and optimization. This could be done for the future although we are satisfied by the current version. We have added two sentences on this subject (line 122)

L113: I think you mean "1 hPa", not "1ppm".
Yes. The typo is corrected

*L125: XCO2 is defined as weighted by the number of dry air molecules per square meter in each layer, not the pressure width. This can be shown to be roughly proportional to dP * (1-q) for a given layer (e.g., O'Dell et al., 2012), where dP is the pressure width and q is the specific humidity in kg/kg. Please recalculate your model XCO2 using this more standard formulation, if possible, or defend your non-traditional XCO2 definition. The differences are generally small (tenths of a ppm), so it is defensible, but if you can be correct, it is best to do so.*

Although it was not explicit in the manuscript, all pressures in the atmospheric transport model that our CAMS product uses (LMDz) are dry air pressures. Thus, there is no need for a correction for the water vapor. We have made it clear in the revised version (line 143).

L138/Figure 1: This is supposedly for the evaluation dataset, but includes N=381k soundings? In section 2, you say the evaluation dataset only includes 155k soundings. So something is wrong –please explain or fix.

For the training, we use only the highest quality data as indicated by the outcome flag and cloud detection flag. We then apply the trained NN to a larger fraction of data, with no restriction on

the outcome flag and a less strict restriction on the cloud flag (2 or 3). The difference between 155k and 381k is entirely explained by these criteria selection.
We certainly agree that it was poorly explained and we have modified the text for better clarity.

L138-152: Based on Fig 5, there appears to be some problem in the surface pressure retrieval over mountains, specifically a high bias generally in these regions (visible over the Tibetan Plateau and the U.S. Rocky Mountains). Please discuss.
In the original version of the paper (second paragraph of section 3), we wrote "…although there is some indication of biases for the lowest pressures that are under-represented in the training dataset." We have added "These biases affect the observations over high elevation surfaces such as the Tibetan Plateau or the US Rocky Mountains"
I suggest that including the surface elevation in the NN may be a good idea, though technically it shouldn't be necessary.
Indeed, it does not appear to be necessary, although it would be easy to do so

Regarding TCCON comparison: It would be useful to include the following statistics for ACOS, NN, and CAMS vs. TCCON: Overall Mean, Overall StdDev, and Stddev of Station mean biases. These are useful to evaluate accurate vs. TCCON in simple statistics. See for example Fig 18b in O'Dell et al (2018). It shows a mean bias of Nadir Land observations vs. TCCON of 0.30ppm, and a stddev of 1.04 ppm (it has not calculated the stddev of the station-level mean biases; some groups do this, others not). Finally, it doesn't look like you're applying the averaging kernel (AK) correction when comparing ACOS to TCCON. This typically makes the stddev about 0.1 ppm better. If you do not make this correction, please point this out in the text.
We have added the overall statistics in the text, and the station by station statistics in Table 1. Indeed, we do not use the AK for the validation. We make that clear in the revised version of the manuscript.

Fig4: Please include a horizontal dashed line so we can see the zero-level. Also, please be clear in the caption or the text if the ACOS and NN are sounding-matched. Typically, when we compare ACOS to TCCON, which use all xco2_quality_flag=0 data. If you were to do this, it may change your results for ACOS vs. TCCON (though better vs. worse, I'm not sure).
We have added the line as suggested in the manuscript. We only use data with a quality flag of zero at all step of the study. This is now made clear in the manuscript.

Technical/Grammatical:
L92: "Although, the NN technique"à"Although the NN technique"
L124: "For each OCO-2 observation"
L129: "cosmic flux anomaly"à"cosmic ray flux anomaly"
L151: "lowest pressure" "lowest pressures"
L181: please replace "classical" with "standard" or "traditional"
L250: "that are described in this paper."
Done (thanks !)

XCO2 estimates from the OCO-2 measurements using a neural network approach

Leslie David, François-Marie Bréon, Frédéric Chevallier

Laboratoire des Sciences du Climat et de l'Environnement/IPSL,
CEA-CNRS-UVSQ, Université Paris-Saclay, F-91198 Gif-sur-Yvette, France

Correspondence to: Francois-Marie Breon (fmbreon@cea.fr)

Abstract. The OCO-2 instrument measures high-resolution spectra of the sun radiance reflected at the Earth surface or scattered in the atmosphere. These spectra are used to estimate the column-averaged dry air mole fraction of CO2 (XCO2) and the surface pressure. The official retrieval algorithm (NASA's Atmospheric CO2 Observations from Space retrievals - ACOS) is a *full physics algorithm* and has been extensively evaluated. Here we propose an alternative approach based on an artificial neural network (NN) technique. For the training and evaluation, we use as reference estimate (i) the surface pressures from a numerical weather model and (ii) the XCO2 derived from an atmospheric transport simulation constrained by surface air-sample measurements of CO2. The NN is trained here using real measurements acquired in nadir mode on cloud-free scenes during even months and is then evaluated against similar observations of odd months. The evaluation indicates that the NN retrieves the surface pressure with a root-mean-square error better than 3 hPa and XCO2 with a 1-sigma precision of 0.8 ppm. The statistics indicate that the NN, that has been trained with a representative set of data, allows excellent accuracy, slightly better than that of the full physics algorithm. An evaluation against reference spectrophotometer XCO2 retrievals indicates similar accuracy for the NN and ACOS estimates, with a skill that varies among the various stations. The NN-model differences show spatio-temporal structures that indicate a potential for improving our knowledge of CO2 fluxes. We finally discuss the pros and cons of using this NN approach for the processing of the data from OCO-2 or other space missions.

Deleted: integrated CO2
Deleted: uses
Deleted: so-called
Formatted: Font: Italic
Deleted: a
Deleted: 7
Deleted: sunphotometer
Deleted:
Deleted: OCO-2

1. Introduction

During the past decades, natural fluxes have absorbed about half of the anthropogenic emissions of CO2 (Knorr, 2009), but there is large uncertainty on the spatial distribution of this sink over time and therefore on the processes that control it. A growing network of high-precision atmospheric CO2 measurements has been used together with meteorological information to constrain the sources and sinks of CO2 using a technique known as *atmospheric inversion* (e.g., Peylin et al., 2013), but the lack of data in large regions of the globe like the tropics does not allow monitoring these fluxes with enough space-time resolution. Early attempts to complement this network with satellite retrievals from sensors that were not specifically designed for this purpose were not successful (Chevallier et al., 2005), but a series of dedicated instruments have been put in orbit since the Greenhouse Gases Observing Satellite (GOSAT, Yokota et al., 2009) and the second Orbiting Carbon Observatory (OCO-2 Eldering et al., 2017a), launched in 2009 and 2014, respectively, and still operated at the time of writing. This new and evolving constellation is directly supported by Japanese, US, Chinese and European space agencies (CEOS Atmospheric Composition Virtual Constellation Greenhouse Gas Team, 2018). All missions have adopted the same CO2 observation principle that consists in measuring the solar irradiance that has been reflected at the Earth's surface in selected spectral bands. Along the double atmospheric path (down-going and up-going), the sunlight is absorbed by atmospheric molecules at specific wavelengths. The resulting absorption lines on the measured spectra makes it possible to estimate the amount of gas between the surface and the top of the atmosphere. CO2 shows many such absorption lines around 1.61 and 2.06 µm that are used to estimate the CO2 column. Similarly, the oxygen lines around 0.76 µm are used to estimate the surface pressure and can also be used to infer the sunlight atmospheric path, leading to the column-averaged dry air mole fraction of CO2, referred to as XCO2 (O'Brien and Rayner, 2002, Crisp et al., 2004).

One main difficulty in the retrieval of XCO2 from the measured spectra results from the presence of atmospheric particles that scatter light and change its atmospheric path. Accounting for aerosols, in particular, is challenging because aerosols are very variable in amount and in vertical distribution. Another major difficulty results from modelling errors. The radiative transfer models that are used for the retrieval leave significant residuals between the measured and modelled spectra, even after the XCO2 and aerosol amount have been inverted for a best fit (Crisp et al., 2012; O'Dell et al., 2018). As a consequence of the various uncertainties in the retrieval process, raw XCO2 retrievals show significant biases against reference ground-based retrievals (Wunch et al., 2011b, 2017). These biases, together with the comparison against modelling results, led to the development of empirical corrections to the retrieved XCO2. In the case of the OCO-2 v8r retrievals generated by NASA's Atmospheric CO2 Observations from Space (ACOS), these corrections amount to roughly half that of the "signal", i.e. of the difference between the prior and the retrieved XCO2 (O'Dell et al. 2018).

The limitations in the full-physics retrieval method, despite considerable efforts and progresses (e.g., O'Dell et al. 2018, Reuter et al. 2017, Wu et al., 2018 in the case of OCO-2), encourage developing alternative approaches. Here, we want to re-evaluate the potential of an artificial neural network technique (NN) to estimate XCO2 from the measured spectra. A NN-based technique was already used by Chédin et al. (2003) for a fast retrieval of mid-tropospheric-mean CO_2 concentrations from some meteorological satellite radiometers. These authors trained their NNs on a large ensemble of radiance simulations made by a reference radiation model and assuming diverse atmospheric and surface conditions. NN-based approaches are also commonly used for the retrieval of other species from various high-spectral-resolution satellite radiance measurements because of their computational efficiency (e.g., Hadji-Lazaro et al. 1999).

A NN approach requires a large and representative training dataset. A standard method for problems similar to that discussed here is to use a radiative transfer model and to generate a large ensemble of pseudo observations based on assumed atmospheric and surface parameters. However, as mentioned above, the radiative transfer models have deficiencies that are rather small, but nevertheless significant with respect to the high precision objective of the CO2 measurements. In addition, there may be some wrong assumptions and unknown instrumental defects that are not accounted for in the forward modeling. We thus prefer to avoid using such radiative transfer models and rather base the training on a fully empirical approach (see, e.g., Aires et al., 2005). We use real OCO-2 observations together with collocated estimates of the surface pressure and XCO2. The retrievals from the NN approach are evaluated against model estimates of surface pressure and XCO2, as well as observations from the Total Carbon Column Observing Network (TCCON, Wunch et al., 2011). In the following, section 2 presents the approach while section 3 describes the results. Section 4 discusses the results and the way forward.

2. Data and method

Our NN estimates XCO2 and the surface pressure from nadir spectra measured by the OCO-2 satellite over land. OCO-2 has eight cross-track footprints (e.g., Eldering et al., 2017), but we only use footprint #4 in the following for simplicity.

Deleted: During

Deleted:

Deleted: normalize

Formatted: English (UK)

Deleted: average CO2

Formatted: English (UK)

Deleted: molecules

Deleted: optimized

Deleted: encourages

Deleted: neural network

Deleted: comprehensive

Deleted: classical

Deleted: The evaluation results show that the NN approach leads to estimates that are of similar quality as those from the official OCO-2 product.

Deleted: presents

If successful, the same approach can be applied to all footprints. The focus on nadir measurements here is motivated by the complication introduced by the Doppler effect in glint mode, which is the other pointing mode for OCO-2 routine science operations: the absorption lines affect pixel elements that vary among the spectra. These variations of the position of the absorption line may cause additional difficulty to the NN training. The solar lines in the nadir spectra are also affected by Doppler shifts due to the motion of the Earth and satellite relative to the sun, but this concerns a limited set of spectral elements that are affected by the solar (Fraunhofer) lines. The development of a glint-mode NN is therefore left for a future study.

We use spectral samples in the three bands of the instrument (around 0.76, 1.61 and 2.06 µm). They have footprints of ~ 3 km^2 on the ground. In principle, each band is described by 1016 pixel elements but some are marked as bad either because some of the corresponding detectors died at some stage or because of known temporary or permanent issues. We systematically remove 15 pixel elements that are flagged in about 80% of the spectra and 478 pixels in the band edges. Conversely, we do not remove the spectra that are affected by the deep solar lines, and we let the NN handle these specific features. Because the information in the spectrum is mostly in the relative depth of the absorption lines, and not in their overall amplitude, we normalize each spectrum by a radiance that is representative of the offline values (i.e. the mean of the 90-95% range for each spectrum). This essentially removes the impact of the variations in the surface albedo and in the sun irradiance linked to the solar zenith angle. Other choices for the input may be attempted in the future.

As input to the NN, we add the observation geometry (sun zenith angle and relative azimuth). The sun zenith angle drives the atmospheric pathlength and is then required for the interpretation of the absorption line depth in terms of atmospheric optical depth. The azimuth was not included in our first attempts but when included, it led to a significant improvement in the results. Although the NN technique does not allow for a clear physical interpretation, we assume that the information brought by the relative azimuth is linked to the polarization of the molecular scattering contribution to the measurements that varies with the azimuth.

The NN exploits these 2557 input variables to compute 2 variables only: XCO2 and the surface pressure. It is structured as a Multilayer Perceptron (Rumelhart et al. 1986) with one hidden layer of 500 neurons that use a sigmoid activation function. The number of hidden layers is somewhat arbitrary and based on a limited sample of trials. Lower quality estimates were obtained with 50 neurons whereas the training time increased markedly for 1000 neurons and more. The weights of the input variables to the hidden neurons and the weights of the hidden variables to the output variables are adjusted iteratively with the standard Keras library (Chollet, 2015). Figure A1 in the appendix illustrates the convergence process. The NN cost function (aka loss) becomes fairly constant for a test dataset after about 100 iterations, while it continues to decrease for the training dataset, indicating an over-fitting of the data. The iteration is stopped when there is no decrease of the test loss for 50 iterations. There is a factor of 3 to 4 between the loss of the training dataset and that of the test, which confirms the over-fitting of the former.

Note that the NN estimate does not use any a priori information on surface pressure or the CO2 profile after the training is done. Also, no explicit information is provided on the altitude, location or time period of the observation. The NN estimates are therefore only driven by the OCO-2 spectrum measurements, together with the observation geometry (sun zenith and relative azimuth). The observation geometry varies with the latitude and the season so that the NN may infer some location information from this input. Conversely, it is the same from one year to the next and, at a given date, for all longitudes. Thus, there is no information on the longitude or the year of observation in the geometry parameters that are provided to the network.

Deleted: .

Deleted: can be

Deleted: its maximum radiance value.

Deleted: viewing

Deleted:

Deleted: it results

Deleted: in

Deleted: .

Deleted: ,

Deleted: our evaluation set

Deleted: We thus stop

Deleted: optimization after 100

Deleted: -

Deleted: spectra measurements, together with the observation geometry.

Deleted: for

The NN training is based on OCO-2 radiance measurements (v8r) acquired during even months between January 2015 and August 2018. The 4-year period allows varying the global background CO2 dry air mole fraction by ~ 2%, as much as typical XCO2 seasonal variations in the northern extra-tropics (see, e.g., Fig. 1 of Agustí-Panareda et al., 2019). Our evaluation dataset is based on observations during the odd months of the same period. In both cases, we make use of XCO2 estimates and the quality control filters of the ACOS L2Lite v9r products: only observations with *xco2_quality_flag*=0 are used. We also consider the *warn level*, *outcome flag* and *cloud_flag_idp* that are provided in the v8r L2lite and L2Std files. For the training of the NN, we only use the best quality observations, i.e. those with a *warn level* lower or equal to 2, a *cloud_flag* of 3 (very clear) and an *outcome flag* of 1. This choice is based on an evaluation of the surface pressure estimates that is described below (with the description of Figure 3). This distinction leads to about 131 000 observations for the training. For the evaluation of the NN estimates, we use less restrictive criteria and accept observations with outcome_flag of either 1 and 2, and cloud_flag of 2 or 3. These choices are justified below. The spatial distribution of the observations that are used for the training is shown in Figure A2 of the appendix. The training dataset covers most regions of the globe with the exception of South America. The underrepresentation of this sub-continent stems for both the high cloudiness and impact of cosmic rays that leads to missing pixel elements (see below).

For the reference surface pressure (training and evaluation), an obvious choice is the use of numerical weather analyses corrected for the sounding altitude. Indeed, the typical accuracy for surface pressure data is on the order of 1 hPa (Salstein et al. 2008). For convenience, we use the surface pressure that is provided together with the OCO-2 data and that is derived from the Goddard Earth Observing System, Version 5, Forward Processing for Instrument Teams (GEOS5-FP-IT) created at Goddard Space Flight Center Global Modeling and Assimilation Office (Suarez et al. 2008 and Lucchesi et al. 2013). There is no such obvious choice for XCO2 as there is no global-scale highly-accurate dataset of XCO2 and we thus rely here on best estimates from a modelling approach. We use the CO2 atmospheric inversion of the Copernicus Atmosphere Monitoring Service (CAMS, atmosphere.copernicus.eu, last access: 28 January 2020, Chevallier et al., 2010); version 18r2). This product was released in July 2019 and contributed, e.g., to the Global Carbon Budget 2019 of Friedlingstein et al. (2019). It results from the assimilation of CO2 surface air-sample measurements in a global atmospheric transport model run at spatial resolution 1.90° in latitude and 3.75° in longitude over the period 1979-2018 and using the adjoint of this transport model. Neither satellite retrievals nor TCCON observations were used for this modelling. For each OCO-2 observation, XCO2 is computed from the collocated concentration vertical profile, through a simple integration weighted by the pressure width of the model layers. Note that the model layers use “dry” pressure coordinates so that there is no need for a water vapor correction in the vertical integration. The GEOS5-FP-IT surface pressure and the XCO2 from CAMS are used both for the training and the evaluation, although using independent datasets (odd and even months).

Many measured spectra lack one or several spectral pixels. This is particularly the case over South America, as a consequence of the South Atlantic cosmic ray flux anomaly that impacts the OCO-2 detector in this region. We therefore devised a method to interpolate the spectra and to fill the missing pixels. Our method first sorts all spectral pixels as a function of the measured radiance in a large number of complete measured spectra. The pixel ranks are averaged to generate a rank representative of the full dataset. Then, when a pixel element is missing in a spectrum, we look for its typical rank and we average the radiances of the two pixel elements that have the ranks just above and below. The procedure is applied even when several pixel elements are missing in a spectrum, except when these are successive in the typical ranking. The procedure described here fills the missing elements, and the NN can then be applied to the corrected spectrum to estimate the surface pressure and XCO2.

Deleted: v8 algorithm, i.e. we
Deleted: use OCO-2
Deleted:
Formatted: Font: Italic
Formatted: Font: Italic
Deleted: ,
Deleted: 155 000
Deleted: both
Deleted: and
Deleted: evaluation.
Deleted: pixel
Deleted: precision
Deleted: ppm
Deleted: surface-based
Deleted: observations
Deleted:

3. Results

Figure 1 shows a density histogram of the GEOS5 FP-IT surface pressure analysis and of the NN estimate for the evaluation dataset (odd months). Clearly, there is an excellent agreement between the two over a very wide range of surface pressures. There is no significant bias and the standard deviation is 2.9 hPa. The equivalent ACOS v8r retrieval shows a bias of 1.5 hPa and a standard deviation of 3.4 hPa, slightly larger than that of the NN approach. Note that the ACOS statistics are those of the ACOS retrieval-minus-prior statistics (see Section 2). Interpreting them in terms of error is counter-intuitive because the Bayesian retrieval is supposed to be better than the prior, but in practice radiation modelling errors lead to a different interpretation (see, e.g., the discussion in Section 4.3.4 of O'Dell et al. 2018).

Deleted: 7
Deleted: v9
Deleted: 4
Deleted: 0

Both NN and ACOS correlations with GEOS5 FP-IT are very high (0.997 and 0.996) although the best fit shows a very small deviation from the 1:1 line. Interestingly, the best fit deviations from the 1:1 line are of opposite sign (slopes 0.99 and 1.01). The results of the NN are surprisingly good given the simplicity of the approach and given that the NN estimate does not use any a priori information or ancillary information such as the surface altitude or temperature profile, contrarily to the ACOS estimate. The quality of the NN results for the estimate of the surface pressure is a first demonstration of the potential of the approach. Note that the retrieval accuracy holds over a very large range of surface pressures (the relative variations of XCO2 are much smaller), although there is some indication of biases for the lowest pressures that are under-represented in the training dataset. These biases of ≈5 hPa affect the observations over high elevation surfaces such as the Tibetan Plateau or the US Rocky Mountains.

Deleted: 998
Deleted: 997
Deleted: 98
Deleted: -
Deleted: pressure
Deleted: pressure

Figure 2 is similar to Figure 1 but for XCO2. There is no significant bias between the NN estimate and the CAMS model, while the standard deviation is 0.84 ppm. The bias-corrected ACOS retrievals show a slight bias against the CAMS model and the standard deviation (1.14 ppm) is larger than that of the NN approach. Note that the statistics given here are affected by CAMS modelling errors that may eventually be corrected with the help of the satellite information. The best fit slope deviations from the 1:1 line are larger than for the surface pressure: the slopes are 0.93 for the NN and 0.87 for ACOS.

Deleted: 74
Deleted: 0.93
Deleted: 94
Deleted: 95

Figure 1 and 2, together with the quantitative assessment of the precision are given for the observations that are clear according to ACOS (Cloud Flag=2 or 3), that have a warn level of 2 or less, that may include missing pixel elements, and that have an outcome flag of 1 or 2. This choice is based on a prior performance analysis. We have analyzed how the performance of the NN approach varies with the quality indicators. For this objective, we have compared the retrieved surface pressure against the value derived from the numerical weather data, as in Figure 1, and we have evaluated the statistic of their difference as a function of the quality flags. First (figure not shown), there is no significant difference between the cases when the measured spectra are complete and those when one or several missing pixel elements have been interpolated with the method described above. Conversely, the statistics vary with the cloud flag and the warn level, as shown in Figure 3. We only use the spectra for which an ACOS retrieval is available. Among those, and according to the flag *cloud_flag_idp*, about 53% are labeled as "very clear" while 43% are "probably clear". The statistics are slightly better for the former than they are for the latter. Conversely, the rather rare "definitely cloudy" and "probably cloudy" show deviations that are significantly larger. This result was well expected since our NN did not learn how to handle clouds in the spectra, so that all "definitely cloudy" and "probably cloudy" soundings are outside the domain covered by the training dataset. Note also that the observations used here have all been classified as "clear" by the ACOS pre-processing. Thus, most OCO-2 observations are not used here and Figure 3 should not be interpreted as the ability to retrieve the surface pressure in cloudy conditions. Most (78%) of the observations have a warn level of 0. The deviation statistics increase with the warn level, both in terms of bias and standard deviation. In comparison, the difference in the

Deleted: term

statistics for an outcome flag of 1 and 2 are small. Besides, more than half of the ACOS retrievals have an outcome flag of 2 which encourages us not to reject those for further use. Based on this analysis, we retain all spectra that are very clear (cloud flag of 2 or 3) and that have a warn level of 2 or less.

We have made a similar figure as Figure 3 but based on the XCO2 estimates (not shown). Although the results are similar in terms of sign (i.e. increase of the deviations with the warn levels), the signal is not as obvious (there is less relative difference between a warn level and another, or for the various cloud flags). Our interpretation is that the relative accuracy of the surface pressure that is used as a reference estimate is much better than that of the NN retrieval, whereas the accuracy of the XCO2 from CAMS is not much better than that of the NN. As a consequence, variations in the accuracy of the NN do not show up as clearly for XCO2 than they do for the surface pressure.

A standard method to evaluate an algorithm that estimates XCO2 from spaceborne observation is the comparison of its products against estimates from TCCON retrievals. These estimates use ground-based solar absorption spectra recorded by Fourier transform infrared spectroscopy and have been tuned with airborne in-situ profiles (Wunch et al. 2010). To take advantage of the full potential of the TCCON retrievals for the bias-correction and validation of the XCO2 estimates, the OCO-2 platform can be oriented so that the instrument field of view is close to the surface station. The ACOS full-physics algorithm can handle these spectra that are acquired in neither nadir nor glint geometries, but the NN was trained solely on nadir spectra and cannot be applied yet to the observations acquired in target mode. We thus have to rely on nadir measurements acquired in the vicinity of TCCON sites. In the following, we use nadir measurements that are within 5 degrees in longitude and 1.5 degrees in latitude to the TCCON site. The XCO2 estimates (either from ACOS, the NN, or the model sampled at the OCO-2 measurement location) are averaged for a given overpass. Similarly, we average the TCCON estimates of XCO2 within 30 minutes of the satellite overpass. No attempt was made to correct for the different weighting functions of the surface and spaceborne remote sensing estimates. The comparisons are shown in Figure 4 for each TCCON station of Table 1.

Overall, the biases and standard deviations of the differences to TCCON observations are -0.34±1.40 ppm for the NN, -0.47±1.49 ppm for ACOS and 0.04±1.09 ppm for CAMS. Statistics per stations are provided in Table 1. Two stations, Paris and Pasadena, show a large negative bias for both estimates, which may be interpreted as the impact of the city on the atmosphere sampled by the TCCON measurement, while the atmosphere sampled by the distant satellite may be less affected. Conversely, there is no such negative bias for other stations that are located close to large cities, such as Tsukuba that is in the suburb of the Tokyo Metropolitan area. Zugspitze is rather specific due to its high altitude. The comparison against TCCON indicates that the NN approach has a similar performance as ACOS, if not better. The dispersion is larger for one versus the other for some stations, while the opposite is true for others. Note also that the CAMS model performs better than both satellite retrievals for most stations. This observation provides further justification to the use of this model for training the NN.

The evaluation of the algorithm performance is limited by the distance between the satellite estimate and its surface validation. This is inherent to the use of nadir-only observations that are seldom located close to the TCCON sites. A reduction of the distance results in less coincidences, which leads to a validation dataset of poor representativeness. Note that the CAMS model was sampled at the location of the satellite observations, so that the higher performance of the model versus the satellite products cannot be caused by a higher proximity to the TCCON station.

We now investigate whether the model-minus-NN differences are purely random or contain some spatial or temporal structures. This question is important as, if the differences show a random structure, there is little hope to use these data to improve the surface fluxes used in the CAMS product. Conversely, if the XCO2 differences do show some structures,

Deleted: classical

Deleted: the

Deleted: networkmeasurements

Deleted:

Deleted: Transform Infrared

Deleted: tuned

Deleted: againstcalibrated

Deleted: networkmeasurements

Deleted: tuning

Deleted: Only cases (overpasses) with at least 10 ACOS sample that meet the spatial criteria are kept.

Deleted: suburbs

Deleted: Note that there is no coincidence for the Lauder station due to the fact that most OCO-2 observations in the vicinity are acquired in Glint or Target modes.

Deleted: observation

Deleted: XCO2 differences do show some structures, they may be attributed to surface flux errors in the CAMS product that may then be corrected through inverse atmospheric modelling. Conversely, if the they can be attributed to surface flux errors in the CAMS product that may then be corrected through inverse atmospheric modelling. There is no certainty, however, as a spatial structure in the NN-minus-CAMS difference can also be interpreted as a bias in the satellite estimate.

We first show (Figure 5) the difference between the NN estimate of the surface pressure and that from the numerical weather analyses. These are monthly maps of the NN-minus-CAMS difference for the 3 years of the period at a 5°×5° resolution. We only present the odd months as the others months have been used for the training, and therefore do not show any significant differences. There are very clear spatial patterns of a few hPa which are not expected and should be interpreted as a bias in the NN approach. The biases over the high mountains and plateaus have already been mentioned. In addition, positive biases tend to occur in the high latitudes, and negative biases toward the tropics. The structures show additional spatial and temporal patterns and are therefore more complex than just a latitude function. The same figure but based on the ACOS retrievals (Figure A3) displays large-scale structures with different spatial patterns: the surface pressure bias is mostly negative over the Northern latitudes and positive over the low latitudes. A histogram (Figure 6) of the monthly differences such as those shown on Figure 5 confirms that the amplitude of the surface pressure biases is larger with ACOS than it is with the NN. The NN (resp. ACOS) surface pressure bias is -0.33 hPa (resp 1.39 hPa) and the standard deviation is 2.12 (resp. 2.79 hPa).

Deleted: the surface pressure

Deleted: analysis

Deleted: In particular

Deleted: altitude

Deleted: A1) display

Deleted: but

Deleted: positive

Deleted: negative

Deleted: 13

Deleted: 41

Deleted: 37

Deleted: 76

Figure 7 is similar as Figure 5 but for XCO2 difference between the NN estimate and the CAMS model. As for the surface pressure, there are clear spatial patterns, with amplitudes of 1 to 2 ppm. The question is whether these are mostly linked to monthly biases in the CAMS model or to the NN. The first hypothesis is of course more favorable as it would indicate that the satellite data can bring new information to constrain the surface fluxes. However, the analysis of the surface pressure that shows biases of several hPa suggests that the NN XCO2 estimate may also show biases with spatially coherent patterns. Interestingly, the patterns vary in time and are not correlated with those of the surface pressure. Further analysis, in particular atmospheric flux inversion, is necessary for a proper interpretation of the NN-CAMS differences.

The differences of ACOS estimates to the CAMS model also show patterns of similar amplitude as those in Figure 7 (Figure A4). However, there is no clear correspondence between these patterns and those obtained using the NN product. The differences between the satellite products and the CAMS model are small, but these contain the information that may be used to improve our knowledge on the surface fluxes. The absence of a clear correlation between the spatio-temporal pattern from the NN and ACOS approaches indicate that their use would lead to very different corrections on the surface fluxes, if used as input of an atmospheric inversion approach. Figure 6, top, shows the histogram of these monthly mean differences. The histograms are very similar for the two satellite products, although the standard deviation of the difference to the CAMS model is slightly larger for ACOS than it is for the NN approach (0.89 vs 0.83 ppm).

Deleted: A2

Deleted:

Deleted: 82

4. Discussion and Conclusion

The use of the same product for the NN training and its evaluation may be seen as a weakness of our analysis. One may argue that the NN has learned from the model and generates an estimate (either the surface pressure or XCO2) that is not based on the spectra but rather on some prior information. Let us recall that the NN input does not contain any information on the location or date of the observation. This is a strong indication that the information is derived from the spectra as the NN does not “know” the CAMS value that corresponds to the observation location. Yet, the NN input also includes the observation geometry (sun angle and azimuth) that is somewhat correlated with the latitude and day-in-the-year. One may then argue that the NN learns from this indirect information on the observation location and then generates an estimate that is based on the corresponding CAMS value. However, since the observation geometry is exactly the same from one year to the next, there is no information, direct or indirect, on the observation year in the NN input. Thus, the XCO2 growth rate, that is accurately retrieved by the NN method (see Fig. 7), is necessarily derived from the spectra. A similar argument can be made on the spatial variation across the longitudes.

To further demonstrate that the NN retrieves XCO2 from the spectra rather than from the prior, we made an additional experiment. The training is based only on even months. As a consequence, the prior does not include any direct information on the odd months. For the odd months, the best prior estimate here is a linear interpolation between the two adjacent even months. We can then analyze how the NN estimate compares with the CAMS product, that accounts for the true synoptic variability, and a degraded version of CAMS that is based on a linear interpolation between the two adjacent months. This comparison is shown in Figure 8. The center figure compares the true CAMS value and that derived from the temporal interpolation. As expected, both are highly correlated (the seasonal cycle and the growth rate are kept in the interpolated values) but show nevertheless a difference standard deviation of 0.89 ppm. This value can be interpreted as the synoptic variability of XCO2 that is present in CAMS but is not captured in the interpolated product. The comparison of the NN estimate against CAMS (right) and the interpolated CAMS (left) shows significantly better agreement to the former. Thus, the NN product does reproduce some XCO2 variability that is not contained in the training prior. It provides further demonstration that the NN estimates relies on the spectra rather than on the time/space variations of the training dataset.

The results shown above indicate that the NN approach allows an estimate of surface pressure and XCO2 with a precision that is similar or better than that of the operational ACOS algorithm. The lack of independent "truth" data does not allow a full quantification of the product precision and accuracy. However, there are indications that the accuracy on the surface pressure is better than 3 hPa RMS, while the precision (standard deviation) of XCO2 is better than 0.9 ppm. The data used for the XCO2 product evaluation has its own error that is difficult to disentangle from that of the estimate based on the satellite observation. It may also contain a bias that is propagated to the NN through its training.

Deleted: evaluation
Deleted: .
Deleted: precision
Deleted: ppm
Deleted: that
Deleted: 8
Deleted: Indeed, the

One obvious advantage of the NN approach is the speed of the computation that is several orders of magnitude higher than that of the full physics algorithm. This is significant given the current re-processing time of the OCO-2 dataset despite the considerable computing power that is made available for the mission. It also bears interesting prospects for future XCO2 imaging missions that will bring even higher data volume (e.g., Pinty et al., 2017).

Another advantage is that the NN approach described in this paper does not require the extensive de-bias procedure which is necessary for the ACOS product (O'Dell et al 2018, Kiel et al. 2019). Per construction, there is no bias between the NN estimates and the dataset that is used for the NN training. The NN approach requires therefore less effort and manpower.

Deleted: as shown

There are however a number of drawbacks for the NN approach that is described in this paper.

One obvious drawback is the use of a CO2 model simulation in the training while the main purpose of the satellite observation is to improve our current knowledge on atmospheric CO2 and its surface fluxes. Our argument is that, although the CAMS simulation used here has high skill (as demonstrated in Figure 4), it may have positive or negative XCO2 biases for some months and some areas. These biases are independent from the measured spectra so that the NN training will aim at average values. As a consequence, the NN product could in principle be of higher quality than the CAMS product, even though the same model has been used as the reference estimate for the training (see, e.g., Aires et al., 2005).

Another drawback of the NN approach is that it does not directly provide its averaging kernel. The averaging kernel vector reports the sensitivity of the retrieved total column to changes in the concentration profile (Connor et al., 1994). It is a combination of physical information (about radiative transfer) and of statistical information (about the prior information). It is needed for a proper comparison with 3D atmospheric models (e.g., Chevallier 2015). When comparing with model simulations, for instance for atmospheric inversion, we may wish to neglect the NN implicit prior information: this hypothesis leads to a homogeneous pressure weighting over the vertical, as this is the product that the NN was trained to simulate. Alternatively, we could decide to neglect the difference in prior information between the NN and the full physics algorithm and use typical averaging kernels of the latter. A third, more involving, option would be to perform a detailed sensitivity study of the NN, based on radiative transfer simulations.

Deleted: an

Similarly, the current version of our neural network does not provide a posterior uncertainty. A Monte Carlo approach using various training datasets could be use in the future for such an estimate.

Also, the NN that was developed cannot be safely used to process observations that are acquired later than a few weeks after the last data of the training dataset, in order to keep the application within the variability range of the training data and despite the CO_2 growth rate. Therefore, the use of the neural network approach for near real time applications would require frequent updates of the training phase.

We acknowledge the fact that the NN product that is evaluated here is not fully independent from the ACOS product. Indeed, we use the cloud flag and the quality diagnostic from ACOS to select the spectra that are of sufficient quality. If we aim at some kind of operational product, there is a need to design a procedure to identify these good quality spectra. One option would be to compare the surface pressure retrieved by the NN to the numerical weather analysis estimate, and to reject cases with significant deviations (e.g. differences larger than 3 hPa).

Deleted: data

Despite these drawbacks, the results presented here do show that a neural network has a large potential for the estimate of XCO_2 from satellite observations such as those of OCO-2, of the forthcoming MicroCarb (Pascal et al. 2017) or the CO2M constellation (Sierk et al. 2018) that aims at measuring anthropogenic emissions. It is rather amazing that a first attempt leads to trueness and precision numbers that are similar or better than those of the full physics algorithm. There are several ways for improvement: one is to provide the NN with some ancillary information such as the surface altitude or a proxy of the atmospheric temperature. Another one is to train the NN with model estimates (such as those of CAMS used here) but that have been better sampled for their assumed precision, for instance through a multi-model evaluation. Also, one could train the NN with observations acquired during a few days of each month, rather than the even months as done here, so that the evaluation dataset would provide a better evaluation of the seasonal cycle.

Deleted: shown in this paper

Our next objective is to attempt a similar NN approach but for the measurements that have been acquired in the glint mode. As explained above, the glint observations may be more difficult to reproduce by the NN than those acquired in the nadir mode. However, we have been very much surprised by the ability of the NN with the nadir data, and cannot exclude to be surprised again. Last, we shall analyze the spatial structure of the NN retrievals in regions that are expected to be homogeneous and in regions where structures of anthropogenic origin are expected (e.g., Nassar et al., 2017; Reuter et al., 2019).

Acknowledgments

This work was in part funded by CNES, the French space agency, in the context of the preparation for the MicroCarb mission. OCO-2 L1 and L2 data were produced by the OCO-2 project at the Jet Propulsion Laboratory, California Institute of Technology, and obtained from the ACOS/OCO-2 data archive maintained at the NASA Goddard Earth Science Data and Information Services Center. TCCON data were obtained from the TCCON Data Archive, hosted by the Carbon Dioxide Information Analysis Center (CDIAC) - tccon.onrl.gov. We warmly thank those who made these data available.

Code/Data availability

The codes used in this paper and the CAMS model simulations are available, upon request, from the author. The OCO-2 and TCCON data can be downloaded from the respective websites.

Author contributions

FMB designed the study. LD developed the codes and performed the computations. All authors shared the result analysis.

Competing interests

The author declare no competing interests.

References

Agustí-Panareda, A., Diamantakis, M., Massart, S., Chevallier, F., Muñoz-Sabater, J., Barré, J., Curcoll, R., Engelen, R., Langerock, B., Law, R. M., Loh, Z., Morguí, J. A., Parrington, M., Peuch, V.-H., Ramonet, M., Roehl, C., Vermeulen, A. T., Warneke, T., and Wunch, D.: Modelling CO_2 weather – why horizontal resolution matters, *Atmos. Chem. Phys*., 19, 7347–7376, https://doi.org/10.5194/acp-19-7347-2019, 2019.

Aires, F., C. Prigent, and W. B. Rossow (2005), Sensitivity of satellite microwave and infrared observations to soil moisture at a global scale: 2. Global statistical relationships, *J. Geophys. Res*., 110, D11103, doi:10.1029/2004JD005094.

Blumenstock, T., F. Hase, M. Schneider, O.E. Garcia, E. Sepulveda. 2017. TCCON data from Izana, Tenerife, Spain, Release GGG2014R1. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, Pasadena, CA, U.S.A. https://doi.org/10.14291/tccon.ggg2014.izana01.R1

CEOS Atmospheric Composition Virtual Constellation Greenhouse Gas Team: A constellation architecture for monitoring carbon dioxide and methane from space, Report from the Committee on Earth Observation Satellites (CEOS) Atmospheric Composition Virtual Constellation (AC-VC), available at: http://ceos.org/document_management/Virtual_Constellations/ACC/Documents/CEOS_AC-VC_GHG_White_Paper_Version_1_20181009.pdf, 2018.

Chédin, A., S. Serrar, N. A. Scott, C. Crévoisier, and R. Armante (2003b), First global measurement of midtropospheric CO2 from NOAA polar satellites: Tropical zone, *J. Geophys. Res*., 108(D2), 4581, doi:10.1029/2003JD003439.

Chevallier, F., Fisher, M., Peylin, P., Serrar, S., Bousquet, P., Bréon, F.-M., Chédin, A., and Ciais, P.: Inferring CO2 sources and sinks from satellite observations: Method and application to TOVS data, J. Geophys. Res., 110, D24309, https://doi.org/10.1029/2005JD006390, 2005.

Chevallier, F., Ciais, P., Conway, T. J., Aalto, T., Anderson, B. E., Bousquet, P., Brunke, E. G., Ciattaglia, L., Esaki, Y., Frohlich, M., Gomez, A., Gomez-Pelaez, A. J., Haszpra, L., Krummel, P. B., Langenfelds, R. L., Leuenberger, M., Machida, T., Maignan, F., Matsueda, H., Morgu, J. A., Mukai, H., Nakazawa, T., Peylin, P., Ramonet, M., Rivier, L., Sawa, Y., Schmidt, M., Steele, L. P., Vay, S. A., Vermeulen, A. T., Wofsy, S., and Worthy, D.: CO2 surface fluxes at grid point scale estimated from a global 21 year reanalysis of atmo- spheric measurements, J. Geophys. Res.-Atmos., 115, D21307, https://doi.org/10.1029/2010JD013887, 2010.

Chevallier, F.: On the statistical optimality of CO_2 atmospheric inversions assimilating CO_2 column retrievals, Atmos. Chem. Phys., 15, 11133–11145, https://doi.org/10.5194/acp-15-11133-2015, 2015.

Chollet F et al. (2015) Keras developer guides, https://keras.io/guides/. Last accessed June 1st, 2020

Connor, B. J., Siskind, D. E., Tsou, J. J., Parrish, A., and Remsberg,E. E.: Ground-based microwave observations of ozone in the upper stratosphere and mesosphere, J. Geophys. Res., 99, 16757–16770, 1994

Crisp, D., Atlas, R. M., Breon, F.-M., Brown, L. R., Burrows, J. P., Ciais, P., Connor, B. J., Doney, S. C., Fung, I. Y., Jacob, D. J., Miller, C. E., O'Brien, D., Pawson, S., Randerson, J. T., Rayner, P., Salawitch, R. J., Sander, S. P., Sen, B., Stephens, G. L., Tans, P. P., Toon, G. C., Wennberg, P. O., Wofsy, S. C., Yung, Y. L., Kuang, Z., Chudasama, B., Sprague, G., Weiss, B., Pollock, R., Kenyon, D., and Schroll, S.: The Orbiting Carbon Observatory (OCO) mission, Adv. Space Res., 34, 700–709, doi:10.1016/j.asr.2003.08.062, 2004.

Crisp, D., Fisher, B. M., O'Dell, C., Frankenberg, C., Basilio, R., Bösch, H., Brown, L. R., Castano, R., Connor, B., Deutscher, N. M., Eldering, A., Griffith, D., Gunson, M., Kuze, A., Mandrake, L., McDuffie, J., Messerschmidt, J., Miller, C. E., Morino, I., Natraj, V., Notholt, J., O'Brien, D. M., Oyafuso, F., Polonsky, I., Robinson, J., Salawitch, R., Sherlock, V., Smyth, M., Suto, H., Taylor, T. E., Thompson, D. R., Wennberg, P. O., Wunch, D., and Yung, Y. L.: The ACOS CO 2 retrieval algorithm – Part II: Global X CO 2 data characterization, Atmos. Meas. Tech., 5, 687–707, https://doi.org/10.5194/amt-5-687-2012, 2012

De Maziere, M., M. K. Sha, F. Desmet, C. Hermans, F. Scolas, N. Kumps, J.-M. Metzger, V. Duflot, J.-P. Cammas. 2017. TCCON data from Reunion Island (La Reunion), France, Release GGG2014R0. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, Pasadena, CA, U.S.A. https://doi.org/10.14291/tccon.ggg2014.reunion01.R1

Deutscher, N., N. M., Diao, M., Elkins, J. W., Gerbig, C., Gottlieb, E., Griffith, D. W. T., Hurst, D. F., Jiménez, R., Keppel-Aleks, G., Kort, E. A., Macatangay, R., Machida, T., Matsueda, H., Moore, F., Morino, I., Park, S., Robinson, J., Roehl, C. M., Sawa, Y., Sherlock, V., Sweeney, C., Tanaka, T., and Zondlo, M. A.: Calibration of the Total Carbon Column Observing Network using aircraft profile data, *Atmos. Meas. Tech.*, 3, 1351–1362, https://doi.org/10.5194/amt-3-1351-2010, 2010.

Deutscher, N., J. Notholt, J. Messerschmidt, C. Weinzierl, T. Warneke, C. Petri, P. Grupe, K. Katrynski. 2017. TCCON data from Bialystok, Poland, Release GGG2014R2. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, Pasadena, CA, U.S.A. http://doi.org/10.14291/tccon.ggg2014.bialystok01.R2

Dubey, M., B. Henderson, D. Green, Z. Butterfield, G. Keppel-Aleks, N. Allen, J.-F. Blavier, C. Roehl, D. Wunch, R. Lindenmaier. 2017. TCCON data from Manaus, Brazil, Release GGG2014R0. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, Pasadena, CA, U.S.A. https://doi.org/10.14291/tccon.ggg2014.manaus01.R0/1149274

Eldering, A., R. Pollock, R. Lee, R. Rosenberg, F. Oyafuso, D. Crisp, R. Granat (2015), Orbiting Carbon Observatory (OCO)-2, Level 1B theoretical basis document (2015),Version 1.2, 40 pp. , http://disc.sci.gsfc.nasa.gov/OCO-2/documentation/oco-2-v6/OCO2_L1B%20ATBD.V6.pdf

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

Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Hauck, J., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Quéré, C. L., Bakker, D. C. E., Canadell, J. G., Ciais, P., Jackson, R. B., Anthoni, P., Barbero, L., Bastos, A., Bastrikov, V., Becker, M., Bopp, L., Buitenhuis, E., Chandra, N., Chevallier, F., Chini, L. P., Currie, K. I., Feely, R. A., Gehlen, M., Gilfillan, D., Gkritzalis, T., Goll, D. S., Gruber, N., Gutekunst, S., Harris, I., Haverd, V., Houghton, R. A., Hurtt, G., Ilyina, T., Jain, A. K., Joetzjer, E., Kaplan, J. O., Kato, E., Klein Goldewijk, K., Korsbakken, J. I., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lenton, A., Lienert, S., Lombardozzi, D., Marland, G., McGuire, P. C., Melton, J. R., Metzl, N., Munro, D. R. Nabel, J. E. M. S., Nakaoka, S.-I., Neill, C., Omar, A. M., Ono, T., Peregon, A., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rödenbeck, C., Séférian, R., Schwinger, J., Smith, N., Tans, P. P., Tian, H., Tilbrook, B., Tubiello, F. N., Werf, G. R. van der, Wiltshire, A. J. and Zaehle, S.: Global Carbon Budget 2019, *Earth System Science Data*, 11(4), 1783–1838, doi:https://doi.org/10.5194/essd-11-1783-2019, 2019.

Griffith, D. W. T., V. A. Velazco, N. Deutscher, C. Murphy, N. Jones, S. Wilson, R. Macatangay, G. Kettlewell, R. R. Buchholz, M. Riggenbach. 2017. TCCON data from Wollongong, Australia, Release GGG2014R0. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, Pasadena, CA, U.S.A. https://doi.org/10.14291/tccon.ggg2014.wollongong01.R0/1149291

Griffith, D. W. T., N. Deutscher, V. A. Velazco, P. O. Wennberg, Y. Yavin, G. Keppel Aleks, R. Washenfelder, G. C. Toon, J.-F. Blavier, C. Murphy, N. Jones, G. Kettlewell, B. Connor, R. Macatangay, C. Roehl, M. Ryczek, J. Glowacki, T. Culgan, G. Bryant. 2017. TCCON data from Darwin, Australia, Release GGG2014R0. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, Pasadena, CA, U.S.A. http://doi.org/10.14291/tccon.ggg2014.darwin01.R0/1149290

Hadji-Lazaro, J., C. Clerbaux, and S. Thiria, an inversion algorithm using neural networks to retrieve atmospheric CO total columns from high-resolution nadir radiances, *J. Geophys. Res*., 104, 23,841-23,854, 199

Hase, F., T. Blumenstock, S. Dohe, J. Gross, M. Kiel. 2017. TCCON data from Karlsruhe, Germany, Release GGG2014R1. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, Pasadena, CA, U.S.A. https://doi.org/10.14291/tccon.ggg2014.karlsruhe01.R1/1182416

Iraci, L., J. Podolske, P. Hillyard, C. Roehl, P. O. Wennberg, J.-F. Blavier, J. Landeros, N. Allen, D. Wunch, J. Zavaleta, E. Quigley, G. Osterman, R. Albertson, K. Dunwoody, H. Boyden. 2017. TCCON data from Armstrong Flight Research Center, Edwards, CA, USA, Release GGG2014R1. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, Pasadena, CA, U.S.A. https://doi.org/10.14291/tccon.ggg2014.edwards01.R1/1255068

Kiel, M., O'Dell, C. W., Fisher, B., Eldering, A., Nassar, R., MacDonald, C. G., and Wennberg, P. O.: How bias correction goes wrong: measurement of XCO2 affected by erroneous surface pressure estimates, *Atmos. Meas. Tech*., 12, 2241–2259, https://doi.org/10.5194/amt-12-2241-2019, 2019.

Kivi, R., P. Heikkinen, E. Kyro. 2017. TCCON data from Sodankyla, Finland, Release GGG2014R0. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, Pasadena, CA, U.S.A. https://doi.org/10.14291/tccon.ggg2014.sodankyla01.R0/1149280

Knorr, W. (2009), Is the airborne fraction of anthropogenic CO2 emissions increasing? *Geophys. Res. Lett*., 36, L21710, doi:10.1029/2009GL040613.

Liu C., Wang, W., Sun, Y. 2018. TCCON data from Hefei, China, Release GGG2014R0. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, Pasadena, CA, U.S.A. http://dx.doi.org/10.14291/tccon.ggg2014.hefei01.R0

Lucchesi, R.: File Specification for GEOS-5 FP-IT (Forward Processing for Instrument Teams), Tech. rep., NASA Goddard Spaceflight Center, Greenbelt, MD, USA, available at: https: //ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20150001438.pdf (last access: 4 December 2018), 2013.

Morino, I., T. Matsuzaki, A. Shishime. 2017. TCCON data from Tsukuba, Ibaraki, Japan, 125HR, Release GGG2014R2. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, Pasadena, CA, U.S.A. http://doi.org/10.14291/tccon.ggg2014.tsukuba02.R2

Morino, I., N. Yokozeki, T. Matzuzaki, A. Shishime. 2017. TCCON data from Rikubetsu, Hokkaido, Japan, Release GGG2014R2. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, Pasadena, CA, U.S.A. https://doi.org/10.14291/tccon.ggg2014.rikubetsu01.R2

Morino, I., Natraj, V., Notholt, J., O'Brien, D. M., Oyafuso, F., Polonsky, I., Robinson, J., Salawitch, R., Sherlock, V., Smyth, M., Suto, H., Taylor, T. E., Thompson, D. R., Wennberg, P. O., Wunch, D., and Yung, Y. L.: The ACOS CO2 retrieval algorithm – Part II: Global XCO2 data characterization, Atmos. Meas. Tech., 5, 687– 707, https://doi.org/10.5194/amt-5-687-2012, 2012.

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

Notholt, J., C. Petri, T. Warneke, N. Deutscher, M. Buschmann, C. Weinzierl, R. Macatangay, P. Grupe. 2017. TCCON data from Bremen, Germany, Release GGG2014R0. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, Pasadena, CA, U.S.A. https://doi.org/10.14291/tccon.ggg2014.bremen01.R0/1149275

O'Brien, D. M. and Rayner, P. J.: Global observations of the carbon budget 2. CO2 column from differential absorption of reflected sunlight in the 1.61 µm band of CO2, *J. Geophys. Res*., 107, ACH6-1, doi:10.1029/2001JD000617, 2002.

O'Dell, C. W., Eldering, A., Wennberg, P. O., Crisp, D., Gunson, M. R., Fisher, B., Frankenberg, C., Kiel, M., Lindqvist, H., Mandrake, L., Merrelli, A., Natraj, 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 Velazco, V. A.: Improved retrievals of carbon dioxide from Orbiting Carbon Observatory-2 with the version 8 ACOS algorithm, *Atmos. Meas. Tech*., 11, 6539–6576, https://doi.org/10.5194/amt-11-6539-2018, 2018.

Pascal, V., Buil, C., Loesel, J., Tauziede, L., Jouglet, D., and Buisson, F.: An improved microcarb dispersive instrumental concept for the measurement of greenhouse gases concentration in the atmosphere, *Proc. SPIE* 10563, 10563K-1–10563K-9, https://doi.org/10.1117/12.2304219, 2017.

Peylin, P., Law, R. M., Gurney, K. R., Chevallier, F., Jacobson, A. R., Maki, T., Niwa, Y., Patra, P. K., Peters, W., Rayner, P. J., Rödenbeck, C., van der Laan-Luijkx, I. T., and Zhang, X.: Global atmospheric carbon budget: results from an ensemble of atmospheric CO_2 inversions, *Biogeosciences*, 10, 6699–6720, https://doi.org/10.5194/bg-10-6699-2013, 2013.

Pinty, B., G. Janssens-Maenhout, M. Dowell, H. Zunker, T. Brunhes, P. Ciais, D. Dee, H. Denier van der Gon, H. Dolman, M. Drinkwater, R. Engelen, M. Heimann, K. Holmlund, R. Husband, A. Kentarchos, Y. Meijer, P. Palmer and M. Scholze (2017) An operational anthropogenic CO_2 emissions monitoring & verification support capacity - Baseline requirements, Model components and functional architecture, doi:10.2760/39384, European Commission Joint Research Centre, EUR 28736 EN

Reuter, M.; Buchwitz, M.; Schneising, O.; Noël, S.; Rozanov, V.; Bovensmann, H.; Burrows, J.P. A Fast Atmospheric Trace Gas Retrieval for Hyperspectral Instruments Approximating Multiple Scattering—Part 1: Radiative Transfer and a Potential OCO-2 XCO2 Retrieval Setup. Remote Sens. 2017, 9, 1159.

Reuter, M., Buchwitz, M., Schneising, O., Krautwurst, S., O'Dell, C. W., Richter, A., Bovensmann, H., and Burrows, J. P.: Towards monitoring localized CO_2 emissions from space: co-located regional CO_2 and NO_2 enhancements observed by the OCO-2 and S5P satellites, Atmos. Chem. Phys., 19, 9371–9383, https://doi.org/10.5194/acp-19-9371-2019, 2019.

Rumelhart, D. E., G. E. Hinton, and R. J. Williams (1986), Learning internal representations by error propagation. Parallel Distributed Processing: Explorations in the Microstructure of Cognition 1, D. E. Rumelhart and J. L. McClelland, Eds., The MIT Press, 318–362.

Salstein, D. A., Ponte, R. M., and Cady-Pereira, K.: Uncertainties in atmospheric surface pressure fields from global analyses, J. Geophys. Res., 113, D14107, https://doi.org/10.1029/2007JD009531, 2008.

Sherlock, V., B. Connor, J. Robinson, H. Shiona, D. Smale, D. Pollard. 2017. TCCON data from Lauder, New Zealand, 125HR, Release GGG2014R0. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, Pasadena, CA, U.S.A. https://doi.org/10.14291/tccon.ggg2014.lauder02.R0/1149298

Shiomi, K., Kawakami, S., H. Ohyama, K. Arai, H. Okumura, C. Taura, T. Fukamachi, M. Sakashita. 2017. TCCON data from Saga, Japan, Release GGG2014R0. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, Pasadena, CA, U.S.A. https://doi.org/10.14291/tccon.ggg2014.saga01.R0/1149283

Sierk, B., Bézy, J.-L., Löscher, A., and Meijer, Y.: TheEuropean CO2 Monitoring Mission: Observing anthropogenic greenhouse gas emissions from space, International Conference on Space Optics, SPIE, 237–250, https://doi.org/10.1117/12.2535941, 2018.

Strong, K., S. Roche, J.E. Franklin, J. Mendonca, E. Lutsch, D. Weaver, P.F. Fogal, J.R. Drummond, R. Batchelor, R. Lindenmaier. 2017. TCCON data from Eureka, Canada, Release GGG2014R3. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, Pasadena, CA, U.S.A. http://doi.org/10.14291/tccon.ggg2014.eureka01.R3

Suarez, M. J., Rienecker, M. M., Todling, R., Bacmeister, J., Takacs, L., Liu, H. C., Gu, W., Sienkiewicz, M., Koster, R. D., and Gelaro, R.: The GEOS-5 Data Assimilation System-Documentation of Versions 5.0.1, 5.1.0, and 5.2.0, Tech. rep., NASA Goddard Spaceflight Center, Greenbelt, MD, USA, available at: https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20120011955.pdf (last access: 4 December 2018), 2008.

Sussmann, R., M. Rettinger. 2017. TCCON data from Garmisch, Germany, Release GGG2014R2. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, Pasadena, CA, U.S.A. https://doi.org/10.14291/tccon.ggg2014.garmisch01.R2

Te, Y., P. Jeseck, C. Janssen. 2017. TCCON data from Paris, France, Release GGG2014R0. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, Pasadena, CA, U.S.A. https://doi.org/10.14291/tccon.ggg2014.paris01.R0/1149279

Warneke, T., J. Messerschmidt, J. Notholt, C. Weinzierl, N. Deutscher, C. Petri, P. Grupe, C. Vuillemin, F. Truong, M. Schmidt, M. Ramonet, E. Parmentier. 2017. TCCON data from Orleans, France, Release GGG2014R1. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, Pasadena, CA, U.S.A. https://doi.org/10.14291/tccon.ggg2014.orleans01.R1

Wennberg, P. O., D. Wunch, C. Roehl , J.-F. Blavier, G. C. Toon, N. Allen. 2017. TCCON data from California Institute of Technology, Pasadena, California, USA, Release GGG2014R1. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, Pasadena, CA, U.S.A. https://doi.org/10.14291/tccon.ggg2014.pasadena01.R1/1182415

Wennberg, P. O., D. Wunch, C. Roehl, J.-F. Blavier, G. C. Toon, N. Allen, P. Dowell, K. Teske, C. Martin, J. Martin. 2017. TCCON data from Lamont, Oklahoma, USA, Release GGG2014R1. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, Pasadena, CA, U.S.A. https://doi.org/10.14291/tccon.ggg2014.lamont01.R1/1255070

Wennberg, P. O., C. Roehl, D. Wunch, G. C. Toon, J.-F. Blavier, R. Washenfelder, G. Keppel-Aleks, N. Allen, J. Ayers. 2017. TCCON data from Park Falls, Wisconsin, USA, Release GGG2014R1. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, Pasadena, CA, U.S.A. http://doi.org/10.14291/tccon.ggg2014.parkfalls01.R1

Wu, L., Hasekamp, O., Hu, H., Landgraf, J., Butz, A., aan de Brugh, J., Aben, I., Pollard, D. F., Griffith, D. W. T., Feist, D. G., Koshelev, D., Hase, F., Toon, G. C., Ohyama, H., Morino, I., Notholt, J., Shiomi, K., Iraci, L., Schneider, M., de Mazière, M., Sussmann, R., Kivi, R., Warneke, T., Goo, T.-Y., and Té, Y.: Carbon dioxide retrieval from OCO-2 satellite observations using the RemoTeC algorithm and validation with TCCON measurements, Atmos. Meas. Tech., 11, 3111–3130, https://doi.org/10.5194/amt-11-3111-2018, 2018.

Wunch, D., G.C. Toon, J.-F.L. Blavier, R.A. Washenfelder, J. Notholt, B.J. Connor, D.W.T. Griffith, V. Sherlock, P.O. Wennberg. The Total Carbon Column Observing Network. *Phil. Trans. R. Soc.* A (2011a) 369, doi:10.1098/rsta.2010.0240

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

Wunch, D., Wennberg, P. O., Osterman, G., Fisher, B., Naylor, B., Roehl, C. M., O'Dell, C., Mandrake, L., Viatte, C., Kiel, M., Griffith, D. W. T., Deutscher, N. M., Velazco, V. A., Notholt, J., Warneke, T., Petri, C., De Maziere, M., Sha, M. K., Sussmann, R., Rettinger, M., Pollard, D., Robinson, J., Morino, I., Uchino, O., Hase, F., Blumenstock, T., Feist, D. G., Arnold, S. G., Strong, K., Mendonca, J., Kivi, R., Heikkinen, P., Iraci, L., Podolske, J., Hillyard, P. W., Kawakami, S., Dubey, M. K., Parker, H. A., Sepulveda, E., García, O. E., Te, Y., Jeseck, P., Gunson, M. R., Crisp, D., and Eldering, A.: Comparisons of the Orbiting Carbon Observatory-2 (OCO-2) XCO2 measurements with TCCON, *Atmos. Meas. Tech.*, 10, 2209–2238, https://doi.org/10.5194/amt10-2209-2017, 2017.

Wunch, D., Toon, G. C., Wennberg, P. O., Wofsy, S. C., Stephens, B. B., Fischer, M. L., Uchino, O., Abshire, J. B., Bernath, P., Biraud, S. C., Blavier, J.-F. L., Boone, C., Bowman, K. P., Browell, E. V., Campos, T., Connor, B. J., Daube, Yokota, T., Yoshida, Y., Eguchi, N., Ota, Y., Tanaka, T., Watanabe, H., and Maksyutov, S.: Global Concentrations of CO2 and CH4 Retrieved from GOSAT: First Preliminary Results, SOLA, 5, 160–163, 2009.

Zhou, M., Dils, B., Wang, P., Detmers, R., Yoshida, Y., O'Dell, C. W., Feist, D. G., Velazco, V. A., Schneider, M., and De Mazière, M. (2016), Validation of TANSO-FTS/GOSAT XCO_2 and XCH_4 glint mode retrievals using TCCON data from near-ocean sites, *Atmos. Meas. Tech.*, 9, 1415-1430, doi:10.5194/amt-9-1415-2016.

Figures and Tables

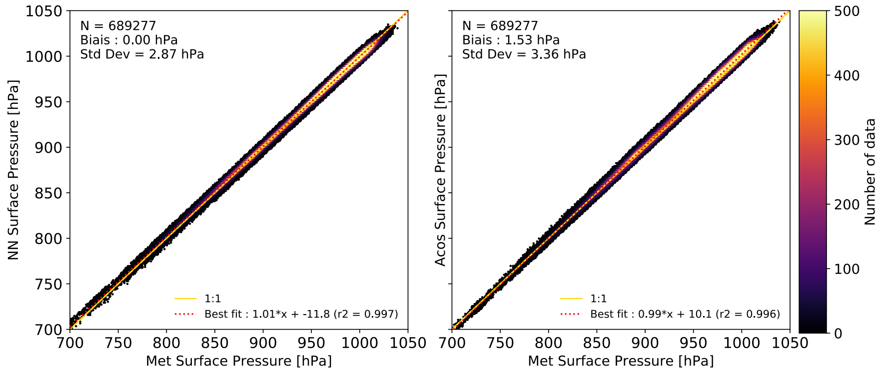

Figure 1: Density histogram of the surface pressure retrieved from the OCO-2 satellite measurements against that derived from GEOS-FP-IT. The left figure is for the NN approach while the right figure is for the ACOS v9r retrieval (using the official bias-correction). The figure insets provide the number of data points, the bias, the standard deviation, the equation of the best linear fit and the correlation. The yellow line is the 1:1 line whereas the red dotted line is the best linear fit.

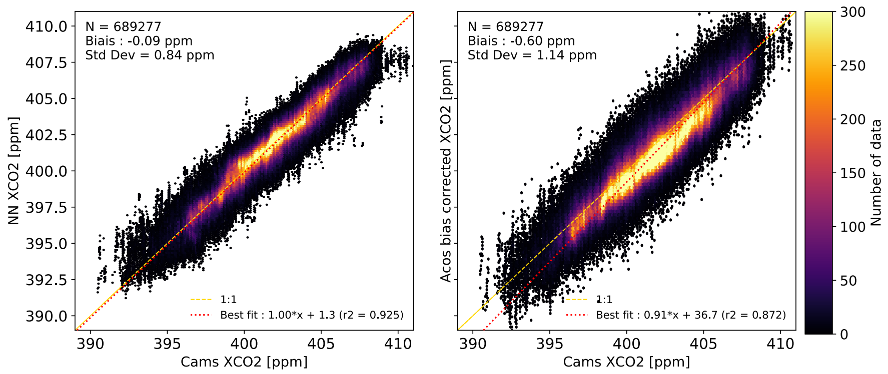

Figure 2: Same as Figure 1 but for XCO2. In this case, the reference data is the CAMS v18r2 simulation.

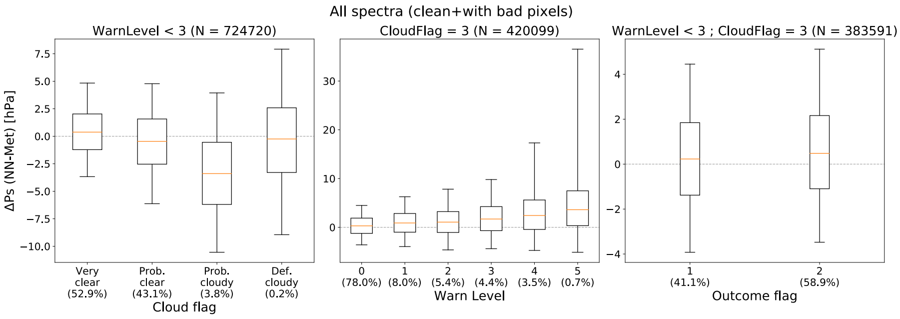

Figure 3: Statistics on the difference between the surface pressure retrieved by the NN approach and that derived from the weather analyses, as a function of various quality parameters. In these figures, the red line is the median, the boxes indicate the 25 and 75% percentiles and the whiskers indicate the 5-95% range. The left figure shows the statistics as a function of the cloud flag, the middle figure is as a function of the warn level, while the right figure is as a function of the outcome flag.

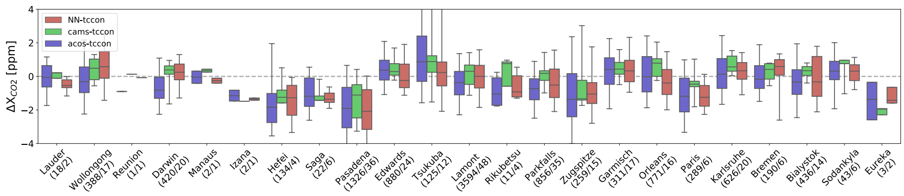

Figure 4: Statistics of the differences between the NN retrieval (red), the CAMS model (green) or the bias-corrected ACOS retrievals (blue) and the TCCON retrievals. The boxes indicate the 25-75% percentiles and the median is shown by the horizontal line within the box. The whiskers indicate the 5-95% percentiles. Stations are ordered by increasing latitudes. The numbers below the station name indicate the number of individual observations and coincidence days used for the statistics. The references of the various TCCON observations are provided in table 1.

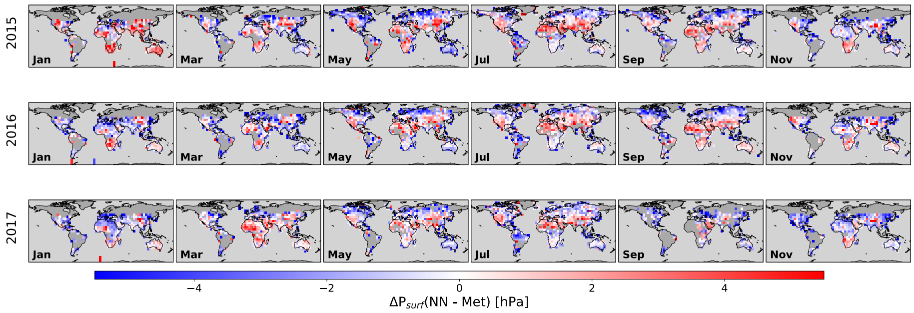

Figure 5: Difference between the NN estimate of the surface pressure and the numerical weather analyses. The differences have been averaged at monthly and spatial 5°×5° resolutions. The results are shown for 3 years and only for the months that were not used for the training.

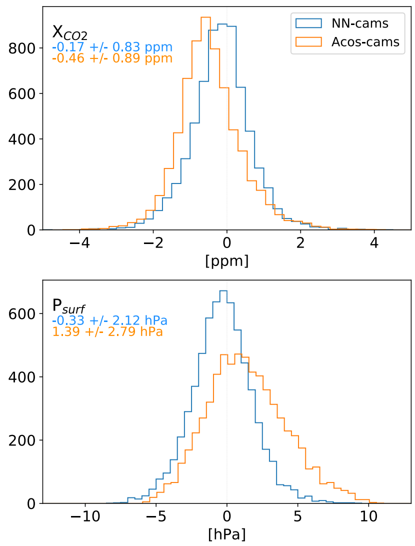

Figure 6: Histogram of the monthly mean differences, at 5° resolution (such as those shown in Figure 5), between the satellite retrievals and the CAMS model. The top figure is for XCO2 while the bottom figure is for the surface pressure. The blue line is for the NN product while the orange line is for ACOS.

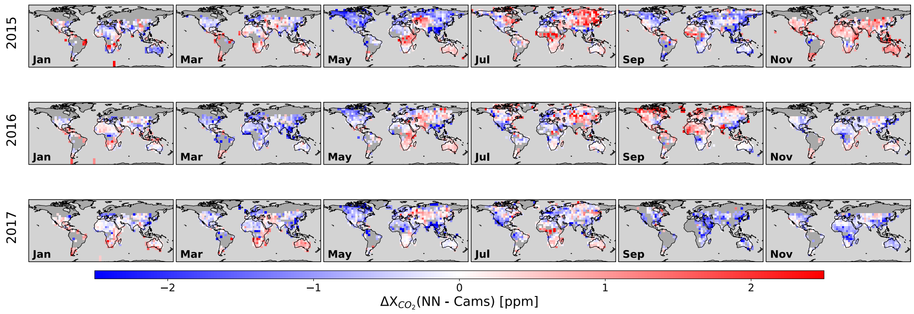

Figure 7: Same as Figure 5 but for the difference between the XCO2 estimated by the NN approach and that derived from the CAMS model.

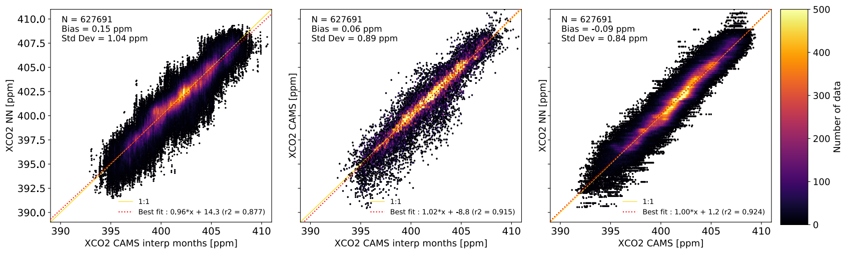

Figure 8: Scatter plots of XCO2 estimated by the NN, the CAMS model, and the CAMS model that has been interpolated in time from adjacent months (see text for details). Note that the number of points is less than in Figure 2 because the edge months could not always be interpolated.

Table 1: TCCON stations used in this paper (Figure 4). The data have been obtained from the tccondata.org web site at during the summer of 2019.

Stations	[lat ; lon]	Altitude [m]	Reference	Biases NN/ACOS/CAM	Std Dev NN/ACOS/CAM
Lauder	[-45.04 ; 169.68]	370	Sherlock et al. 2017	-0.48/-0.25/0.076	0.43/1.51/0.16
Wollongong	[-34.41 ; 150.88]	30	Griffith et al. 2017	0.60/-0.20/0.42	1.21/1.32/0.60
Reunion	[-20.90 ; 55.49]	90	De Maziere et al. 2017	-0.08/-0.90/0.13	-/-/-
Darwin	[-12.43 ; 130.89]	30	Griffith et al. 2017	0.19/-0.69/0.23	0.80/1.09/0.72
Manaus	[-3.21 ; -60.6]	50	Dubey et al. 2017	-0.25/-0.05/0.34	0.43/1.04/0.26
Izana	[28.3 ; -16.48]	2300	Blumenstock et al. 2017	-1.35/-1.14/-1.48	0.18/0.92/0.0
Hefei	[31.90 ; 118.67]	30	Liu et al. 2018	-1.47/-1.58/-1.01	1.11/1.76/0.63
Saga	[33.24 ; 130.29]	10	Shiomi et al. 2017	-1.36/-1.03-1.15	0.57/1.22/0.59
Pasadena	[34.14 ; -118.13]	240	Wennberg et al. 2017	-2.12/-1.87/-1.41	1.57/1.64/1.17
Edwards	[34.96 ; -117.88]	700	Iraci et al. 2017	0.07/0.41/0.50	1.00/1.01/0.64
Tsukuba	[36.05 ; 140.12]	30	Morino et al. 2017	0.42/1.43/1.05	2.13/2.53/1.61
Lamont	[36.6 ; -97.49]	320	Wennberg et al. 2017	-0.03/-0.38/0.16	1.07/1.21/0.94
Rikubetsu	[43.46 ; 1473.77]	390	Morino et al. 2017	-0.57/-0.84/0.47	0.84/1.07/0.98
Parkfalls	[45.94 ; -90.27]	440	Wennberg et al. 2017	-0.41/-0.75/0.11	1.15/1.01/0.72
Zugspite	[47.42 ; 11.06]	2960	Sussmann and Rettinger 2017	-0.85/-1.14/-0.83	1.45/1.85/1.36
Garmisch	[47.48 ; 11.06]	740	Sussmann and Rettinger 2017	0.40/0.28/0.43	0.98/1.29/0.62
Orleans	[47.97 ; 2.11]	130	Warneke et al. 2017	-0.35/0.13/0.66	1.06/1.38/0.67
Paris	[48.85 ; 2.36]	60	Te et al. 2017	-1.29/-1.24/-0.62	1.30/1.66/1.23
Karlsruhe	[49.1 ; 8.44]	110	Hase et al. 2017	0.26/0.21/0.75	0.80/1.29/0.55
Bremen	[53.10 ; 8.85]	7	Notholt et al. 2017	0.30/-0.07/0.36	1.11/1.02/0.45
Bialystok	[53.23 ; 23.02]	180	Deutscher et al. 2017	-0.11/-0.32/0.33	1.31/1.30/0.42
Sodankyla	[67.37 ; 26.63]	190	Kivi et al. 2017	0.26/0.24/0.61	0.79/1.36/0.80
Eureka	[80.05; -86.42]	600	Strong et al. 2017	-1.02/-1.50/-2.16	1.01/2.25/0.41

Appendix

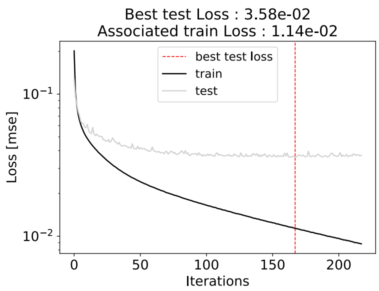

Figure A1: Illustration of the iterative convergence of the NN during its training. The loss is an indicator of the difference between the NN estimate and the dataset. One dataset is used for the best estimate of the NN weights whereas another independent one is used for the evaluation of the NN capability. The NN is stopped when there is no further reduction of the loss for the test dataset for 50 iterations. The weight for the NN are those obtained for the lowest loss of the test dataset (iteration 167 on the figure).

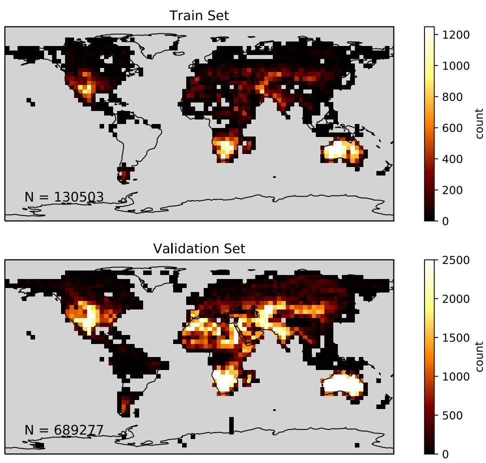

Figure A2: Spatial density of the observation that have been used for the training (top) and validation (bottom) processes.

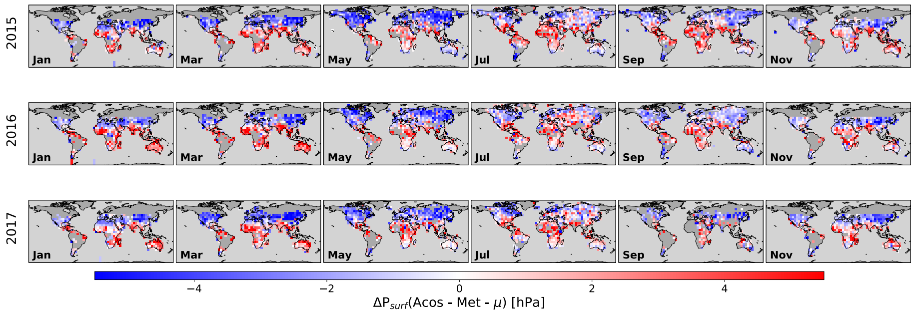

Figure A3: Same as Figure 5 but for the surface pressure retrieved by the ACOS algorithm. The mean bias over the full period (μ) is removed so that the differences are centered on zero.

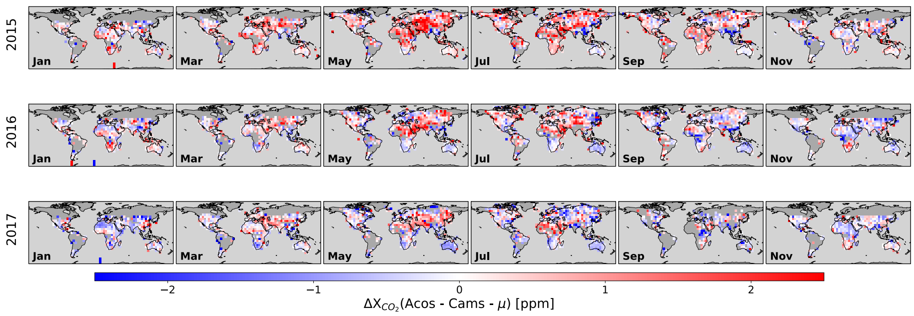

Figure A4: Same as Figure 7 but for the XCO2 retrieved by the ACOS algorithm. The mean bias over the full period (μ) is removed so that the differences are centered on zero.