**Atmospheric
Measurement
Techniques
Discussions**

Interactive comment on "Multispectral information for gas and aerosol retrieval from TANSO-FTS instrument" *by* H. Herbin et al.

Anonymous Referee #1

Received and published: 18 February 2013

The authors perform an information content analysis of GOSAT spectral channels with respect to retrieval of CO_2, CH_4 and H_2O. The question they are addressing is whether the combined use of spectral channels will improve retrieval of trace gases compared to the use of single channels.

I consider the style of the manuscript and scientific rigorousness insufficient for publication in a scientific journal. The description of the method is confusing and contains a number of obvious errors. Previous literature on GOSAT trace gas retrieval (algorithms) is hardly discussed. The results in their present form seem to have limited scientific value.

These are some of the major concerns:

(i) I consider the style of the manuscript (phrasing, spelling, grammar) insufficient for publication in a scientific journal. I had to read many sentences a number of times before I understood them (or at least thought I did so). I will give a number of typical examples to illustrate my point, but the authors should carefully check the entire manuscript. I am not a native English speaker myself, so I know that it is sometimes a struggle to find the right words in a foreign language. However, there are many online editing services which can help to improve the text.

p.8437, l.27-28: "The forward model which allows treating simultaneously high resolution infrared and visible measurements and performed to study gas composition in presence of scattering particles is presented in the Sect. 3." -> "... and is used to study..." ? p. 8448, l.4-7: "Obviously, the gas concentrations retrieval in the presence of an aerosol layer without taking it into account in the forward model, either the calculation cannot converge, either the results will be obtain with a very large error or bias." -> ? "Obviously, if gas concentrations are retrieved when aerosols are present but not accounted for, retrieved concentrations will be biased or retrieval will not converge." Many more of these type of sentences can be found in the manuscript. p.8442, l.4: "As emphasis by Rodgers (2000)" -> emphasised p.8440, l.13: "(Engelen and stephens, 2004)" -> "(Engelen and Stephens, 2004)" p.8439, l.19: "...the fours bands of TANSO-FTS..." -> four p.8442, l.14-15: "The details of the theoretical elements of this procedure and examples are provided in the reference (L'Ecuyer et al., 2006)." -> Wordy; rephrase for example as "This procedure is described in L'Ecuyer et al. (2006)." p.8437, l.3: "(spectral range, spectral resolution, multiple viewing angles, polarization...)" -> Don't put an ellipsis (...) there. Make explicit what you mean: if more items should be added to this list in your opinion, add them; if not, stop after "polarization". Ellipses appear in a number of places throughout the manuscript. p.8441: AMT gives clear instructions on typesetting mathematical symbols: matrices should be printed in bold face and vectors in bold face italics. The authors do not follow this convention, and the authors are even inconsistent within their own style. For example, l.23, Eq. 5: Sa, K and Se are all matrices, but only Sa is typeset correctly; l.24-25:

matrix K is printed differently within the same sentence. Many more examples in this section.

(ii) Description of the method (p.8440-8441, sect.4) is sloppy and contains a number of obvious errors. This does not make a very trustworthy impression.

In p.8440, l.9-10 the authors state: "The theoretical elements relevant for the present information content analysis are similar to those described by Rodgers (2000). They are only briefly summarized hereafter." Still, in the next paragraphs a detailed derivation of the information content of the measurement H follows, which is for the most part irrelevant for this paper. The authors even try to give a definition of thermodynamic entropy (l.18-19). Information content of the measurement H within Rodgers' optimal estimation framework has a specific definition (cf. equation 2.72 in Rodgers, 2002). Eq. 2 is supposed to give this definition, but it is wrong (natural logarithm, NOT logarithm with base 2). Eq. 3 giving a definition of the Degrees of Freedom for the Signal, which is the basis for their analysis, is also wrong. I hope this is only a typing error, but that they did their calculations correctly. The correct formula can be copy-pasted from Rodgers (2002), for example equation 2.80. In addition, notation is sloppy: the a posteriori covariance matrix is denoted as S2 and Sˆ interchangeably. This confuses the reader. Matrices Sm and Sb are not well explained and the authors' notation is inconsistent with Rodgers (2002). Sm and Sb have different meanings in Rodgers (2002) and in this manuscript (cf. equations 3.16, 3.18 and 3.19 in Rodgers, 2002), which is very confusing. This section should have started with an explanation of the forward model, but it ends with it.

However, I wonder why, after having tried to explain information content H, no results concerning H whatsoever are presented?

(iii) I also have concerns about the authors' assumptions on a priori errors.

The authors state that they "assume in this theoretical study a very small prior knowledge. This choice is justified by the fact that this study focuses on the information

content to perform retrievals in a general case and thus highlight information coming from the measurement. Therefore, we will always assume Sa as a diagonal matrix with an error (Perror) of 100% on the prior state vector xa" (p.8442, l23-25, p.8443, l.1-2). No further discussion and references for this assumption are given. I find it quite arbitrary to simply assume a prior errors of 100% without any further discussion. In principle, I understand the motivation to choose large a prori errors, but will you get the same results when you assume errors of 200%? The point is that I find it difficult to interpret the Degrees of Freedom for the Signal (DOFS) in an absolute sense in the case of unrealistically large a priori errors, as the authors do (see remark vi below). If they want to investigate the information content of the measurement itself, why not assume infinitely large a priori errors? (But then the DOFS would be equal to the number of state vector elements). When performing an information content analysis as they present it (Degrees of Freedom for The Signal, and information content of the measurement H), I would first expect a more extensive discussion of available a priori information. DOFS and H can only be interpreted with respect to the assumed a priori errors.

As to the assumed errors in model parameters, please provide references for the values that you assume (e.g. error of 5% in the surface emissivity).

In the second set of simulations, the authors take aerosol into account to investigate a "situation where one wants to use aerosol information from ancillary data (e.g. Retrieval from other instruments mainly dedicated to aerosol study)." (p.8448, l.11-12). However, no discussion of uncertainties associated with these ancillary data is provided. Instead, the authors again assume that "The aerosol parameters [...] are supposed to be known with an uncertainty of 100 %." (p.8448, l.26-28) Thus, I would say the simulations actually do NOT investigate a situation of having available ancillary information.

(iv) The authors claim that Fig. 2, 4, and 7 show the Degrees of Freedom for The Signal. However, it seems that they are showing the diagonal elements of the averaging kernel matrix. This is not properly explained.

These figures provide indications about the height sensitivity of retrieval for each GOSAT channels. But why not tell the full story and show complete averaging kernels? The authors should also investigate in a systematic way the dependence of the averaging kernels on properties of the surface-atmosphere. For example, they only present DOFS for retrieval over sea. But how about sensitivities near the surface if we are over bright land?

(v) As a general remark, however, the scientific contribution of their comparison of the sensitivities of GOSAT's spectral channels is not clear to me. I am not that familiar with GOSAT literature, but I expect that many previous (pre-launch) studies have already addressed this question. Many people have probably already done similar analyses. However, no literature addressing a similar research question is cited. I find this odd. Also, except for Yoshida et al. (2011) no literature on existing (GOSAT) trace gas retrieval algorithms is discussed. For example, the ACOS team has published many papers on their CO2 algorithm (e.g. O'Dell et al., AMTD, 2012 and references therein).

(vi) The authors conclude at the end of their paper: "To summarize, the first part demonstrates that in case of clear sky condition and given the instrumental characteristics of TANSO-FTS instrument, we can retrieve between 1 and 2 columns for CH4, 2 columns for CO2 and at least 6 columns for H2O from ground to 20 km, with a good accuracy, with a reduced selection of channels (1000) mainly from Band 4." (p.8451, l.20-24) I don't know what the reason for this conclusion is. It is not explained in the text. To determine how many subcolumns can be retrieved, I would need to know the DOFS, which is not reported (see remark iv). Hence, I need to know the area below the curves of Fig. 2 and Fig. 4.

(vii) There are many statements in the paper which are awkward or simply incorrect. p.8439, l.3: "The absorption lines computation includes Lorentz, Doppler and Voigt line-shape..." -> Voigt profile is a convolution of Lorentz and Gauss (Doppler) profiles. p.8446, l.12-15: "Indeed, it is important to note that the total errors are regularly governed by the a priori error Sa. This is due to the fact that we deliberately chose to set

a large uncertainty on the prior state vector xa (100% uncertainty) which deteriorates dramatically the total error when the measurement sensitivity decreases (cf. Eq. 5)." -> This statement is incorrect and the authors even contradict themselves. According to Eq.5, to which they are referring: if Sa is large, Sa will hardly have an effect on the total error (thus, Sa will NOT govern total error). p.8436, l.15-18 (abstract): "This work was conducted in order to develop a powerful tool that allows retrieving simultaneously not only the gas concentrations but also the aerosol characteristics by selecting the so called "best channels", i.e. the channels that bring most of the information concerning gas and aerosol." -> No tool is developed in this paper. The methodology is described in Rodgers (2000). Selection of "best channels" (please avoid quotation marks to indicate special use of words; simply explain what you mean with best channels) is done according to L'Ecuyer et al. (2006). p.8443, l.14: As far as I understand FTS instruments, isn't the noise dependent on the average intensity of the spectrum?

(viii) As a general remark, I think it is better to have separate sections presenting and discussing results. This will allow the authors to provide a more detailed comparison of their work and the work by Yoshida et al. (2011) as well as by other authors, which I think is appropriate here. In the present manuscript, results and interpretation are confounded. For example:

p.8446, l.28 – p.8447, l.2: "So, Bands 2 and 3, although less sensitive, are commonly used for total columns retrievals, but, the lack of sensitivity requires perfectly constrain the inversion by an adapted Sa matrix (Yoshida et al., 2011)." -> I found this sentence hard to understand. It seems that the authors draw a conclusion based on their results and discuss the work by Yoshida et al. (2011) in the same sentence.

(ix) Concerning the figures:

p.8461, fig.1: This figure is not well explained: Please state explicitly in the text (or caption) what you mean with normalized radiance and with spectra showing contributions of individual absorbers. Why show the contributions from all the different isotopologues

of water? What does this information contribute to the story you are trying to tell in your manuscript? What is the surface albedo you assume? You show normalized solar irradiance spectra for bands 1-3: what does this information add?

As a general advice, I encourage the authors to reflect upon what story a figure is supposed to be telling? What message are you trying to get across?

p.8462, fig.2: Some lines are overlapping, which you mention in the text. Please also mention this in the caption. However, perhaps you should think of some other representation that is more easy to understand (generally, I find eleven plotting lines in a single figure too much). I initially had trouble reading this figure, because the dotted line in the figure seems to be different from the dotted line in the legend (is this true?).

p.8467, fig.7: This figure is too small to read. I cannot read what is on the x-axis for the DOFS plots. You suggest it is optical depth, which is strange.

(x) However, I must say that I appreciate the authors' effort to include in their analyses model parameter errors instead of only taking measurement noise into account. The question of selecting wavelengths within a spectral channel is interesting and important (reduction of computation time). Unfortunately, they only briefly discuss this issue. They present results of a selection procedure described in L'Ecuyer et al. (2006). I think this could be a more interesting research question to further elaborate.