Reply to Christopher O`Dell

Dear Referee,

Thank you very much for your effort and time to review our manuscript amtd-7-12173-2014 “A linear method for the retrieval of sun-induced chlorophyll fluorescence from GOME-2 and SCIAMACHY data”. Your constructive criticism has led us to rethink the way of presenting our retrieval approach and results. It has helped us to significantly improve our manuscript in its revised version. In the item-by-item response hereinafter, comments are listed in black and responses in blue. Please note that we have additionally listed the most significant revisions in the general final comment.

Sincerely,

Philipp Köhler on behalf of the authors

Overview

This paper presents a method to retrieve solar-induced Chlorophyll fluorescence (SIF) from satellite measurements with only moderate spectral resolution. The paper does an OSSE-type validation using synthetic data, and then applies the method to both GOME-2 and to SCIAMACHY data.

General Comments

I have both positive and negative comments regarding this work. The application to SCIAMACHY has not been done before and in itself merits publication. However, the primary problem with this paper is that, so far as this reviewer can tell, the method simply takes an established method (Joiner et al, 2013) and makes two minor modifications to it. Those modifications, however, have a significant impact on the results, and may in fact be a WORSENING of the algorithm beyond the originally published method. Indeed, in certain areas with active Vegetation, the seasonal cycle amplitude of SIF changes by a factor of 2-3 between the original and modified algorithm. This is a huge difference, and the cause should be better understood. Is this due purely to the underlying algorithm or due to the filtering method chosen (ie which soundings to average together)?

Thank you for pointing out that there are indeed large uncertainties in absolute SIF values. We agree that the presented retrieval method is an evolution of that proposed by Joiner et al. (2013) and highlight this throughout the manuscript. Hence, it is remarkable that there are nevertheless large differences in retrieval results. Although spatio-temporal patterns compare well ($r2=0.9$ for global aggregates, see new Fig. 13), it is immediately noticeable that the amount of SIF obtained from the presented retrieval is about two times higher. In this context, it should be considered that a change in the algorithm of Joiner et al. has lead to a significant decrease in absolute SIF values from version 14 to version 25. That decrease can be primarily attributed to the change in fitting window, which was decreased to 734—758 nm. This is different to the fitting window used here. We are not able to judge which absolute values are closer to reality, but our revised manuscript clearly shows that the presented retrieval approach is not less valid than that from Joiner et al.. In particular, we compare long-term averages (overlapping period) of GOME-2 SIF (our approach and V25 from Joiner et al., 2014) and the recently processed SCIAMACHY data with GOSAT results (new Fig. 14). It appears that there is a higher correlation to the GOSAT SIF data set for the presented retrieval algorithm.
We believe that the manuscript was not clear enough to justify our approach. Therefore, the relevant sections have been revised accordingly.

Please note that, building upon previous works from Guanter et al. (2013) and Joiner et al. (2013), our approach solves the important issue of the arbitrary selection of the number of free model parameters. For this reason and in view of the results, we claim that this is not a minor modification.

Considering these large differences, I cannot recommend publication until more foundational work is done to show if there is indeed a problem with the original Joiner algorithm (or the latest Version of the Joiner algorithm available). The simulation-based validation may be sufficient for this purpose. In fact, because it appears that the identical simulations were used for validation of both the original Joiner et al. method as well as the presented method, it should be possible to do a more extensive validation with that synthetic data set. It would be best to provide a table like that of Table 1 Joiner et al. (2013), comparing your results directly to the Joiner results for the synthetic data set. Indeed, the method presented appears to have some significant problems, which I describe in the specific comments section below. These comments should be fully addressed, and if necessary, modifications to the algorithm should be made to address them.

We agree that our manuscript suffered from a lack of information on this point. We therefore extended the simulation-based validation to enhance confidence and to prove that there are indeed no significant problems. In particular, we provide a correlation error matrix for a sample fit (new Fig. 8) in order to show that the reference fluorescence emission spectrum (hf) has no significant correlation with any of the provided PCs. Furthermore, we show that systematic effects from different illumination angles, water vapour contents and aerosol optical thicknesses can be excluded. Moreover, we evaluate the number of selected PCs as well as the total number of selected coefficients. We also provide a table of basic results from different retrieval windows (Table 1). A key finding is that the described retrieval approach suggests to use significantly less PCs than Joiner et al (2013). In this context, it should be noted that the recent v25 also uses a smaller number of PCs (12 instead of 25), which is due to a confined retrieval window.
We agree that a further direct comparison to the Joiner et al (2013) algorithm would be beneficial, but -in view of the considerable extent of our revised sensitivity analysis- beyond the scope of this paper. Furthermore, we believe that the revised manuscript contains valuable results.

Specific Comments
To motivate why a departure from Joiner et al is necessary, this work should state what the deficiencies of the Joiner method are. The fact that the Joiner forward model is nonlinear is not one of them. If it takes three of four steps to converge to the cost function minimum, so what? That is not a valid reason, unless three or four steps (instead of one) makes the algorithm too slow to apply to SCIAMACHY or GOME-2, which explicitly appears not to be the case. So it must be a reason concerning the accuracy of that algorithm.

We fully agree with the reviewer that the non-linearity is not a deficiency of the Joiner et al. forward model. It was our intention to emphasize differences from the beginning on. However, we now realize that the manuscript was misleading in this point. We removed all relevant statements accordingly. Please note that a comparable forward model to our approach was proposed by Guanter et al. (2013) and recently by Guanter et al. (2015). From that point of view, our approach should be not less valid, and we feel that it is unnecessary to motivate a departure from the Joiner forward model. However, the primary reasons to use a linear model are an easier implementation and an increased computationally efficiency of the backward elimination algorithm. The main and unique feature which arises through that step is an automated selection of model parameters. It should be taken into account, particularly because of the different results, that Joiner`s results have also evolved since the 2013 publication.

Please note that equation (1) is identical to Joiner et al. (2013) equation (1).

The concerned equation has been replaced according to a comment from reviewer 1. In the revised manuscript, we derive the forward model starting from the most simplistic formulation of the TOA signal.

In all your equations, please State that bolded items indicate vectors (ie quantities with a spectral dependence), or write out the wavelength-dependence explicitly.

Thank you for pointing out that such a statement was missing. We added the explanation to the first equation.

Equation (3) - please explain why the po term can be (and has been) neglected. Certainly Rayleigh scattering could play a role and it seems strange to ignore it, when it is pretty easy to include at least to first order. I realize Joiner et al. (2013) also neglect this term.

Thanks for this comment. In general, it should be considered that the TOA signal is reconstructed with a data-driven approach that makes it impossible to distinguish between different contributions from the atmosphere. We have to acknowledge that the previous argumentation was inconsistent with respect to the equations. Please note that we have revised the section in order to remain consistent.
In particular, we argue that the planetary reflectance (without SIF, including path radiance, surface reflectance and atmospheric transmittance) can be modeled by a combination of spectrally low and high frequency components. Our data-driven/statistically interpretation refers more to the reconstruction of the signal than to the single components of the equation. The main point thereby is to reproduce the measurement with a sufficient accuracy to be able to evaluate the changes in the fractional depth of solar lines by fluorescence. We therefore revised the manuscript accordingly including the new Eq. (1), (2) and (5).

The linearization method described in equation (4) takes a reasonably valid forward model and makes it less valid. This is because the coefficients which determine $T_{\uparrow\downarrow}$ also determine $T_{\uparrow}$ (via your equation 4). But then you determine $T_{\uparrow\downarrow}$ separately. This doesn't make sense. One needs to solve for all these quantities simultaneously (and yes this makes the problem nonlinear!). Perhaps this is only a very small error, but this may explain why the performance of the "linear" method is relatively poor. It certainly seems like a wholly unnecessary step.

Thank you for this comment that showed us that the explanation of the concerned step was not clear enough. We modified this issue in the revised manuscript. From our point of view, solving for $T_{\uparrow\downarrow}$ and $T_{\uparrow}$ simultaneously has no significant advantage with respect to decoupling them. Please note that all simulations presented by Guanter et al. (2015) are also based on that decoupling.
Those results could serve as evidence to support our statement. We discuss further implications in the revised manuscript in order to justify this approach. In particular, we show by means of simulated TOA radiances that the error due to an in-filling of atmospheric absorption lines is negligible.

Therefore, we have tested two scenarios without instrumental noise, a medium and a large difference in SIF emission (2.1 and 4.3mW/m^2/sr/nm), and calculated $T_{\uparrow}$ as described, whereas resulting changes are only marginal (new Fig. 4 in the revised manuscript).

Further, the step in equation (5) seems bizarre. The authors state that Joiner et al. solve for (i*j + 1) variables. In fact, they solve for (i+j+1). So if i=4 (# of reflectance coefficients) and j=10 (# of PCs to determine atmospheric transmittance), they would solve for 15 parameters. i*j+l = 4*10+1 = 41 which is many more (and is simply incorrect). So the motivation to even perform this step seems to disappear.

Apologies if the impression has been given that Joiner et al. solve for i*j+1 model parameters. We explicitly mention in the revised manuscript that a consequence of our approach is an increased number of state vector elements compared to Joiner et al. In the extended sensitivity analysis we are further able to show that our algorithm selects about 7 PCs and 15 coefficients respectively. In this context, it is to be remarked that Joiner et al. proposed to use 25 PCs for an equivalent retrieval window. Despite the significantly lower number of used PCs, the main advantage of using our approach is that the number of selected coefficients remains stable regardless of how many PCs are provided.

It may be that reducing the number of solved-for parameters would be useful, as this paper proposes, but it would be good to see that method applied directly to the model of Joiner et al (which uses the correct forward model, and keeps the transmittance parameters separated from the surface albedo parameters, which in itself greatly helps to reduce the number of parameters needed).

We agree that it would be interesting to apply a model selection also to the forward model of Joiner et al., but we do not expect an improvement for following reasons:
1) Our algorithm selects the required model parameters automatically with respect to the goodness of fit balanced by model complexity. Thus it should be advantageous to provide an increased number of potentially solved-for parameters in order to obtain the 'best' possible fit.
2) We doubt that we use an incorrect forward model. Again, in view of other studies (Guanter et al., 2013 and Guanter et al., 2015), our sensitivity analysis and presented results, it can be claimed that our forward model is not less valid than that of Joiner et al. (2013).
3) Joiner et al. (2014) v25 results used a reduced number of PCs (12) in a different fitting window (734–758nm).

Nevertheless, it might be worthwhile to do so in future investigations.

Equation (6) is not generally applicable. This assumes Gaussian white noise. Many instruments have a noise that increases with increasing signal – in particular grating Systems. Therefore, I question whether this is really the noise model that is suggested by the GOME-2 instrument team. Can the authors verify that this is the case? One should always use the noise model suggested by the instrument team, unless you have good reason to believe you have a more representative noise model for that particular instrument.

The reason for using that noise model is the estimation of a realistic noise for the simulated radiance. It may have been unclear that the signal level is actually taken into account. The reference SNR is determined for each measurement (simulated and real) from the averaged radiance between 757.7–758 nm using calculations of the SNR vs. level 1B calibrated radiance performed by EUMETSAT. We clarified this issue in the revised manuscript. An effect of the increasing noise with the signal level can be observed in the new Fig. 6, where lower illumination angles cause a slightly higher variance in the bias (retrieved minus simulated SIF). Indeed, it would be best to use the noise estimate attached to the L1 files for the real satellite data, but we just adopted the noise estimation of the sensitivity study for the sake of simplicity. Please realize that a similar procedure was used by Joiner et al. (2013).

Section 4.1 - please state if this is the identical synthetic data set used by Joiner et al. (2013). Both had 230,400 samples so it seems that that is the case.

Yes, in principle we are using the same synthetic data set. The only difference is that we add the instrumental noise with respect to the calculated SNR from GOME-2 (in relation to the radiance level at a reference wavelength) provided by EUMETSAT via Eq. 7, while Joiner et al. (2013) used a fixed SNR of 1000 and 2000 respectively. The missing reference has been added in the revised manuscript.

Please write the equation for "BIC". Is it the same as the reduced chi-squared of the fit to the measurements? If so, please used reduced chi-squared instead, as it is much more common than "BIC". If not, please State why BIC is preferable, with a reference if possible.

Thanks for this comment. We recognize that there was a lack of information on this point. We therefore extended the backward elimination section including the equation. Regarding your question why the BIC is preferable, it has to be mentioned that there are also other methods to compare and select models. Here, we decided to use the BIC because it penalizes the number of model parameters the most.

Joiner et al. (2013) provide a statement that they have determined the theoretical error on the retrieval and it closely matches the true error, for the synthetic data test. Have you done so as well? What is the result?

Please note that we show error estimates based on monthly averages in the revised Sect. 5.3 “Spatio-temporal composites”. On the contrary, we examine uncertainties due to different atmospheric states in our sensitivity analysis. Hence, we do not provide a comparison between theoretical and true error. Nevertheless, we presume that the statement also holds for our approach.

Page 12195, line 25. Does the SD achieved with the presented method agree with that calculated from theory (ie via simple error propagation of the assumed instrument noise model via your equation 7)?

Please note that this figure has been replaced by a long term average in the revised manuscript. Nevertheless, it must be taken into account that the standard deviation covers instrumental noise as well as natural variability. Hence, the value of such a comparison would be limited.

Page 12196, line 15. Why would one expect high reflectance surfaces to be associated with higher errors? I assume it is because scattering is enhanced (because of scattered light reflected from the atmosphere to the ground that can make it to the satellite). It would be good to state a hypothesis here.

Please note that this paragraph has been removed due to the replacement of the associated figure.

Section 5.4. This is interesting. Is this also seen in the Joiner et al. data set? If so, it would provide further proof of an instrument artifact rather than something particular to your retrieval algorithm.

Yes, it was confirmed that there is also a slight offset in the Joiner et al. data set. Please note that this section has also been extended in order investigate more in detail if this bias is really related to an instrumental issue. This is most likely the case, although the magnitude can be decreased by an appropriate sampling of the training set.

Technical/Grammar Comments

- Pg 12177 line 17. "Comprehends". Suggest replace with "encompasses"
- Pg 12182 line 21. "in a sufficient precision" -> "with sufficient precision"
- Pg 12183 line 1. "that a removal" -> "that the removal"
- Pg 12183 line 12. "SIF free spectrum in an" -> "SIF-free spectrum with an"
- Pg 12184 line 1. "by applying a previous step." -> "by applying an additional step prior to minimization."
- Page 12187 line 17. "A further filtering beside" "Further filtering besides"
- Pg 12188 line 3 "simulations comprehend" -> "simulations include" or "simulations contain"
- Pg 12189 line 5 "TOCs." ^ "TOC SIF spectra."
- Pg 12189 line 10. "which impact the retrieval of SIF inevitably." -> "which inevitably impact the retrieval of SIF."
- Pg 12192 line 6. "should amount at least" -> "should be at least"
- Pg 12195 line 17. "discrepance" "discrepancy". Also pg 12196 line 4.
- Pg 12196 line 3. "congruent" -> "similar" (at least this is what I think you are trying to say)
- Pg 12199 line 2. "most homogeneous" -> "mostly homogenous"
- Pg 12199 line 8. Suggest you cite Frankenberg et al. (2012) here, which shows how the SIF signal is impacted from cloud effects directly.
- Pg 12200 line 2. "against cloud contamination" -> "against moderate cloud contamination". To make clear it is not impervious to all cloud contamination.

Thank you very much for these correction suggestions which we fully adopted (where still applicable after revision).