We thank the review for the insightful comments and suggestions. The responses to the reviewer are in red and modifications to the paper content are in *red and italic with indentation*. The equations, lines and sections in the comments are referred to the updated version of the manuscript.

General comments:

Atmospheric correction is an essential part of satellite remote sensing of land surface. Yin et al. describe and evaluate the Sensor Invariant Atmospheric Correction (SIAC) algorithm for atmospheric correction. The existing atmospheric correction methods can be improved and therefore further development of the algorithms is welcome.

SIAC is capable of carrying out atmospheric correction both for Sentinel 2 and Landsat 8 satellite data. Furthermore, SIAC is a Bayesian (statistical) algorithm so it can take advantage of prior information and it produces uncertainty estimates for the surface reflectances produced. The algorithm is tested and validated with ground-based AERONET and RadCalNet data. There are no significant steps taken in new method development in this work, but SIAC combines well existing methods. The results shown in the manuscript show that SIAC is capable of carrying out good quality atmospheric correction.

The selections made in the development of SIAC are mostly well justified and based on previously published literature. It is very good that the authors have shared the codes for others to be used.

My main criticism is in the presentation. The quality of presentation in the manuscript varies. Time to time the text is well written and smooth but quite often the text is difficult to read. The manuscript is quite long and structured so that first SIAC and results are explained in general terms followed by the Discussion and Conclusions, and then all the technical details are mostly included in the Appendices. As this is a method development manuscript, I find this a bit difficult for the reader as it is needed to browse back and forth while reading. Furthermore, the manuscript heavily relies on use of acronyms and symbols. It is good that the main symbols are explained in a table but this also makes the manuscript very slow to read, especially for a person who is not that familiar with the field. I would strongly recommend the authors to think the use of acronyms (even single letter acronyms) and symbols, and possibly shorten and re-organise the manuscript for improved readability. Also, there were some typos in the text so proofreading is recommended. Most of the font sizes in figures are too small and very difficult to read.

This paper describes the theoretical basis, practical implementation, and the validation results, which inevitably makes the paper long and relatively complex. We have decided to put implementation details and further testing in appendices to have a more readable main paper by a wider audience. We have added some text clarifying that the appendices are there to provide detail and additional testing towards the end of Section 2.2:

> *In this paper, we present the theoretical underpinnings of the method and major results, relegating details of the implementation and additional results to the appendices.*

Regarding acronyms, the ones used are in common use throughout the literature. Symbols are described in Table 2 and used within the detailed implementation sections of the paper, so we assume the reader is familiar with the paper and overall method at that stage.

We have updated the font size in the figures to improve readability.

Specific comments:

l.15 Abstract mentions efficient emulators. However, in the text this was a bit unclear how these were used. Could it be clarified?

l.129 In the abstract this was probably mentioned as "statistical emulation", neural networks are not really statistical emulation.

We have removed the qualifier “statistical” from the abstract and modified the text to be clearer around L129:

> *Running the model atmospheric model many times is computationally costly, and is often approximated by using e.g, look-up tables. Here, we provide fast surrogate approximations to the full atmospheric model, called emulators (Gómez-Dans et al., 2016). These approximations are based on fully connected artificial neural networks (ANNs) and provide an estimate of the pi terms as a function of the model inputs Xac . Additionally, the Jacobian of the atmospheric model (needed for efficient gradient descent minimisation and for uncertainty propagation) is also approximated by the emulator making use of backpropagation techniques (Hecht-Nielsen, 1992)*

l.53 AERONET is based on remote sensing, not in-situ measurements.

Changed to “AERONET estimates” in L203

l.83 ".., we calculate:". "Calculate atmospheric parameter estimation"?

Thanks for pointing this out; it has been changed to:

SIAC comprises two major steps:

l.84 The list is difficult to follow.

We modified the list to be a bit more readable.

l.133 "This may cause errors". Can you estimate the significance of these errors

We have changed the text to be clearer:

This choice of aerosol type may cause errors when conditions strongly depart from it, such as situations dominated by urban, maritime or biomass burning conditions. See (Tirelli et al., 2015; Shen et al., 2019) for analysis of the impacts of aerosol types.

We have also added a clarification of this in Sect 5.3 ("Future developments"):

In this paper, the atmospheric composition is set by a model (6S in this case), and by a choice of aerosol optical properties (continental aerosol model). The use of emulators of the RT model makes it easy to change the RT model entirely in the code, or to use a different configuration of the model used. We can also extend the scheme to retrieve independent aerosol species concentrations by both modifying the RT model (and thus extending the number of parameters that go in the inference), and by using data on species distribution available from CAMS and extending the prior to cover these. A similar approach has been implemented in the MAJA processor (Rouquié et al., 2017), which uses the CAMS aerosol species data to define the aerosol types for the atmospheric correction and has found improved atmospheric correction results over deserts. This approach may well be valuable in areas of high dust aerosol loading, or in situations where biomass burning results in an important contribution to aerosol concentrations.

l.169 "Matrix D" was this defined?

It is defined as first-order spatial difference constraint, and the matrix is given in the updated version of the paper (Eq. 7)

l.183 "We assume that mean atmospheric parameters...are constant..." Can you estimate the significance of this assumption?

We have added support for this statement:

We calculate the pa,b,c with the mean atmospheric parameters x and the auxiliary data (Ozone and elevation) at MODIS spatial grid Gc. A linear interpolation is then used to re-sample the pa,b,c to the target sensor grid Gm, which is then used to derive the mean surface reflectance r with Eq. 8. The simple linear interpolation method used to resample the pa,b,c to sub-MODIS scale is justified as atmospheric parameters are known to exhibit much larger correlation lengths (100s of km) (Anderson et al., 2003; Chatterjee et al., 2010).

l.221 Can you give an example of "other artifacts"

We have added a list of likely artifacts:

[...](such as saturation of pixel value, cloud shadow, modelling error from the RadCalNet surface reflectance to TOA reflectance, etc.)

l.223 Why the tolerance of 10% was selected? How this filtering affects the evaluation of the results?

We have added a justification in the text:

5% is used as the target uncertainty of the S2/L8 TOA reflectance and the RadCalNet TOA uncertainty is around 2-5% for non-absorption bands (Wenny and Thome, 2022), which lead us to choose 10% as the threshold to filter out bad samples

l.300 "corrected to R", what is R?

R is surface reflectance (changed in the text)

l.333 Why different TCWV gamma values are used for S2 and L8?

We have pointed out that this is discussed in one of the appendices:

We use γ values of of 5 for S2 and L8 AOT, 5 for S2 TCWV and 0.1 for L8 TCWV in SIAC. Cross-validation studies suggest that there is a wide range suitable values for γ (Sect. G for additional details on this choice and its implications).

p.23 Figure 13. What are the colors of the bars?

p.26 Figure 15. What are the colors?

We have added a clarification on the boxplots colors

Boxplot colors are the same as colors in Fig. 2.

l.374 Is IQR defined?

It has been updated as the interquartile range (IQR) with the reference added.

l.375 Validation of surface reflectance uncertainty was a bit unclear. Should be clarified.

We have added a summary statement to this paragraph.

In summary, our tests show that we have a conservative estimate of uncertainty. The overestimation of the variance in surface reflectance is more marked for L8 than S2.

l.390 In many sentences (especially on page 25) you use "that" & "this" and it is unclear to which word these are referring to

We have updated the text to be clearer in that paragraph.