# Reply to Reviewer #1

We thank the reviewer for their careful reading of the manuscript and for their constructive comments and suggestions. Their comments have helped us improve both the clarity and the scientific quality of the manuscript. Below we provide a point-by-point response to every comment. Reviewer comments are shown in **bold** font while our responses are given in normal font.

**"This paper presents an error-correction framework for the CAMS solar radiation service (CRS) using an XGBoost-based regression model, with SHAP used for feature attribution. While the attempt to improve operational irradiance products is relevant, the work faces significant limitations in terms of scientific novelty, rigorous benchmarking, and the conceptual definition of "uncertainty." The presentation of the methodology in Section 2.3 is unnecessarily opaque, and the scope of the study lacks a broader, more robust framework suitable for a general audience or global applications."**

We thank the reviewer for the careful assessment of our manuscript and for highlighting the concerns. We have substantially revised the manuscript in response to the comments.

In particular, we have clarified the scientific motivation and novelty of the work by emphasizing that the proposed framework is intended to analyse and interpret the residual errors of an operational radiative-transfer model rather than the physical controls of solar irradiance itself. We have also clarified the rationale for the residual-learning approach, simplified Section 2.3 by removing unnecessary parts of the algebraic derivation and revised the terminology throughout the manuscript to distinguish deterministic residual modelling from probabilistic uncertainty quantification.

Furthermore, we have strengthened the discussion of the interpretation and limitations of the SHAP analysis, including the roles of representativeness effects, correlated predictors, and the distinction between statistical associations and physical causality. Finally, we have streamlined the presentation by moving several repetitive diagnostic figures to the Appendix while retaining the key results in the main text.

The specific responses to each of the reviewer's comments are provided below.

**"1. The methodology relies on standard tools (XGBoost + SHAP) applied to existing data, which limits the novelty of the contribution."**

We appreciate this comment. We agree that neither XGBoost nor SHAP are novel algorithms by themselves. The novelty of this work lies in their application to a new scientific problem. Rather than using explainable machine learning to directly predict the surface irradiance, we train the model to predict the errors of an operational radiative transfer model and use SHAP to identify the atmospheric and surface conditions associated with these errors in the model. This allows us to investigate the error characteristics of the actual implementation of the radiative transfer code with its necessary assumptions and parameterizations, rather than the physical controls of solar radiation, which are already well understood. We have revised the Introduction and Discussion to clarify this distinction and better highlight the scientific contribution of the work.

**"2. The findings regarding the drivers of error (e.g., cloud and aerosol influences) are largely intuitive and established in the literature. The paper would benefit from a more critical discussion on why a data-driven approach is necessary here if the drivers are already known, and what specific new insights these models provide beyond confirming known physics."**

We thank the referee for this important comment. We agree that the general importance of clouds and aerosols for solar radiation is well known. However, the objective of the present work is different. Rather than identifying the factors controlling irradiance itself, we identify the factors associated with the residual errors of an operational radiation service.

The value of the data-driven framework lies in its ability to analyse operational model inputs simultaneously and to identify specific variables and conditions that are most strongly associated with the residuals in the actual CAMS model implementation. For example, the analysis highlights the importance of individual cloud properties as seen by the CAMS cloud retrieval and recently introduced aerosol species in the CAMS IFS such as ammonium and nitrate aerosol optical depths. While SHAP analysis alone does not identify the underlying physical cause of the residuals, it provides a practical tool for identifying variables and modelling components that warrant further investigation and may offer opportunities for future improvements in the retrieval and modelling framework. It also helps to quantify future development priorities and offers a reproducable tool to quantify development progress in future.

To clarify this point, we have expanded the Conclusions section to better distinguish between identifying the physical controls of irradiance and identifying the variables associated with residual errors in the operational CRS products.

**"3. The authors should clarify why they chose to model the error specifically, rather than attempting to predict the irradiance or clear-sky index directly, and whether a direct prediction model was considered or tested as a baseline."**

The choice to model the residual error rather than irradiance directly was motivated by the objective of improving an existing operational radiation product rather than replacing it with a purely data-driven model. The CRS radiative-transfer framework already captures the dominant physical controls of surface irradiance, and the machine-learning model therefore only needs to learn the remaining systematic residuals.

This residual-learning strategy has several advantages. First, the learning problem is simplified because the machine-learning model focuses only on the unexplained component of the signal. Second, the physical consistency and interpretability of the underlying radiative-transfer model are retained. Third, the resulting correction can be applied directly as a post-processing step to existing CRS products without modifying the operational radiative-transfer calculations. Furthermore, the results quantify the different impacts and help to quantify priorities for further development of the model.

We have expanded the manuscript to clarify the motivation for modelling residual errors rather than irradiance directly and have added a discussion of the advantages of this approach.

**"4. There is a major concern regarding the use of the term "uncertainty." In the atmospheric sciences and remote sensing, uncertainty implies a probabilistic, quantified measure (e.g., variance, confidence intervals). This work presents a point-regression model, which does not provide a probabilistic view. The authors should either adopt more precise terminology (e.g., "systematic deviation" or "bias") or incorporate formal uncertainty quantification techniques (e.g., quantile regression, ensemble methods, or evaluating metrics like CRPS, PIT, or reliability diagrams)."**

We thank the referee for this important comment and agree that the term "uncertainty" was used too broadly in the original manuscript. The present framework employs a deterministic point-regression model and therefore does not provide a probabilistic estimate of uncertainty in the statistical sense.

To avoid ambiguity, we have revised the manuscript by replacing uncertainty with more precise terminology such as residual, prediction error, or systematic bias depending on the context. We also clarify that the objective of the proposed framework is to analyse and reduce systematic patterns in the residuals of the CRS products rather than to quantify predictive uncertainty. We have also removed the word "uncertainty" from the title of the manuscript and the title now reads as "Machine-learning-based approach for solar radiation model prediction error identification, attribution and correction". We believe this terminology more accurately reflects the scope of the present study.

**"5. The study lacks a comparison with alternative methods. To justify the use of XGBoost, it is essential to benchmark against other statistical or simpler ML models (e.g., linear regression, random forests) to demonstrate the specific value added by this approach."**

The objective of this work was not to identify the optimal regression algorithm, but to investigate the error attribution using an explainable machine-learning framework. XGBoost was selected because of its excellent predictive performance, robustness and its compatibility with the exact TreeSHAP algorithm. A comprehensive benchmark of alternative machine-learning methods is an interesting topic, but falls outside the scope of the study. We have clarified this limitation in the manuscript.

**"6. The current approach of training a single model for the entire disk is likely inappropriate, as errors in solar radiation retrievals are highly dependent on climate, seasonal regimes, and sky conditions. The authors should consider a foundation model in their future work."**

We agree that solar radiation errors depend on geographical location, climate, season, and atmospheric conditions. However, the present framework uses a comprehensive set of predictors describing cloud properties, aerosol composition, surface characteristics, atmospheric state, and solar geometry. Furthermore, it relies on a carefully quality controlled set of ground observation stations in various climates to cover seasonal regimes and ensuring to cover all relevant sky conditions with a large number of cases each. These variables enable the XGBoost model to learn different error characteristics across a wide range of environmental conditions through its tree-based partitioning of the feature space, without explicitly defining separate regional or climate-specific models. For the needs of the operational CAMS service, the existence of a single error assessment model is a critical constraint. Localized models may describe individual locations better, but are not suited to be used in an operational context.

Nevertheless, we agree that alternative modelling strategies, including foundation model approaches, represent an interesting direction for future research. We have added this perspective to the Conclusions.

**"7. Section 2.3 is unnecessarily complex. The algebraic derivation provided in Eq. (2) adds little value and obscures a straightforward procedure. The section could be significantly condensed by simply stating that a regression model was developed on the retrieval error."**

We thank the referee for this suggestion. We agree that the algebraic derivation was unnecessarily detailed and did not substantially contribute to understanding the methodology. We have therefore simplified Section 2.3 by removing the intermediate derivation and instead directly describing the residual-learning approach. The revised section now focuses on the rationale for modelling the residual error and on how the machine-learning model is used to correct the CRS estimates.

**"8. The paper is saturated with repetitive plots that do not offer proportional informational gains. A more concise synthesis of the findings, perhaps moving some supplemental diagnostic plots to an appendix, would strengthen the narrative and focus the reader's attention on the key results."**

We agree that several figures present similar types of analyses for the different irradiance components. To improve the flow of the manuscript and reduce repetition, we have moved the corresponding DHI and BNI diagnostic SHAP figures to the Appendix while retaining the key GHI figures in the main text. The main findings remain unchanged and the appendix provides the complete analyses for readers interested in the detailed results.