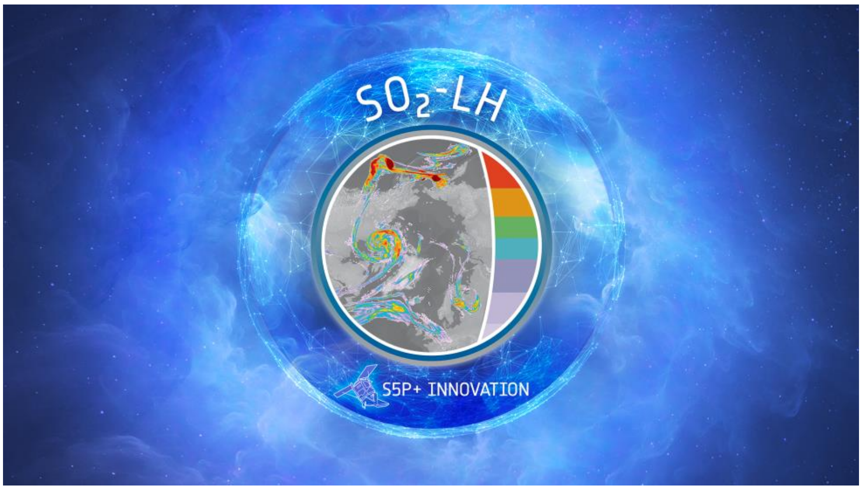

S5P+I: SO2LH project

Scientific Roadmap [D10]

Doc. ID	S5P+I_SO2LH_SR_D10
Issue	1.0
Date	2022-02-17
Prepared by	P. Hedelt (DLR)
Status	Draft

German Aerospace Centre - Remote Sensing Technology Institute (DLR-IMF)
Aristotle University of Thessaloniki – Laboratory of Astrophysics (AUTH-LAP)
University of Oxford (Univ. Oxford) – Earth Observation Data Group (EODG)
European Centre for Medium-Range Weather Forecasts (ECMWF)

DOCUMENT APPROVAL RECORD

	Digital signature:
Prepared by:	
Checked by:	

CHANGE RECORD

Issue	Date	Change
0.9	2021-12-20	First initial version
1.0	2022-02-17	Updated version after feedback from ESA

TABLE OF CONTENTS

1. Introduction ... 4
1.1 Purpose ... 4
1.2 References ... 4
1.2.1 Applicable Documents ... 4
1.2.2 Reference Documents ... 4
1.3 Terms and Abbreviations ... 5
2. Overview of the S5p+I SO2LH project ... 6
3. Analysis and strategic actions for fostering a transition from research to operational activities 7
3.1 Implementation into an operational processing environment ... 7
3.2 Scientific roadmap ... 7
3.3 Application to other UV satellite sensors ... 9
3.4 Conclusions ... 9

LIST OF TABLES

Table 1 Applicable Documents ... 4
Table 2 Reference Documents ... 4

1. Introduction

1.1 Purpose

This document is the Scientific Roadmap (Deliverable D10) as part of the Sentinel-5 precursor Innovation: SO2 LH project. This document presents a critical analysis of all feedback from scientists and institutions that have accessed the prototype product; strategies for integrating the developed methods and models into existing large scientific initiatives and operational frameworks as well as scientific development strategies improving the developed methods and product.

1.2 References

1.2.1 Applicable Documents

The following project documents contain provisions which, through reference in this text, become applicable to the extent specified in this document.

Table 1 Applicable Documents

Document Title		Document ID	Issue
[AD01]	Statement of Work, ESA Express Procurement Plus - [EXPRO+] Sentinel-5p Innovation (S5p+I)	EOP-SD-SOW-2018-049	2.0 (20.08.2018)

1.2.2 Reference Documents

The following standards or documents are referenced in this document. They have been used (in the sense of tailoring) to prepare the document on hand.

Table 2 Reference Documents

Title		Document ID	Issue
[RD01]	Sentinel-5P TROPOMI SO_2 LH ATBD	S5P+I-SO2LH-D4-ATBD-v4	4.0
[RD02]	Sentinel-5P TROPOMI SO_2 LH Validation Report	S5P+I-SO2LH-D5-VR-v2	2.0
[RD03]	Sentinel-5P TROPOMI SO_2 LH Requirements Baseline	S5P+I_SO2LH_RB_D1	1.0
[RD04]	Sentinel-5 Level-2 Prototype Processor Development Requirements Specification	S5-RS-ESA-GR-0131	1.7

1.3 Terms and Abbreviations

Abbreviations specific to the report are found in the following table:

Abbreviation	Meaning
DOAS	Differential Optical Absorption Spectroscopy
FP_ILM	Full-Physics Inverse Learning Machine
LH	Layer height
LUT	Look-up table
NN	Neural Network
PCA	Principal Component Analysis
RTM	Radiative Transfer Model
S5P	Sentinel-5 precursor
SCD	Slant column density
SO_2	Sulfur dioxide
TROPOMI	Tropospheric Ozone Measurement Instrument
VCD	Vertical column density

2. Overview of the S5p+I SO2LH project

The **ESA Sentinel-5p+ Innovation project (S5p+I)** has been initiated to develop novel scientific and operational applications, products and retrieval methods that exploit the potential of the Sentinel-5P mission's capabilities beyond its primary objectives.

Accurate determination of the location, height and loading of SO_2 plumes emitted by volcanic eruptions is essential for aviation safety. SO_2 in the atmosphere has important impacts on chemistry and climate at both local and global levels. The SO_2 layer height is furthermore one of the most critical parameters that determine the impact on the climate. The height of volcanic ash columns are often estimated by local observers with mostly unknown accuracy. The plume height can also be determined using aircraft, ground-based radar or LIDAR but such observations are not often available and many volcanic eruptions in remote areas remain not observed. In addition, volcanic plumes containing SO_2 but not ash cannot be directly seen.

Although retrievals of the SO_2 plume height have been carried out using satellite UV backscatter measurements, such as from OMI (Ozone Monitoring Instrument on Aura) or GOME-2 (Global Ozone Monitoring Experiment-2 on MetOp), until now such algorithms are very time-consuming, since the spectral information content and its characterization require computationally demanding radiative transfer modelling. Due to the high spatial resolution of TROPOMI (Tropospheric Ozone Measurement Instrument) aboard S5p (Sentinel-5p) and consequent large amount of data, an SO_2 layer height algorithm has to be very fast.

The **SO_2 Layer Height (SO2LH)** theme is dedicated to the generation of an SO_2 layer height product for Sentinel-5p considering data production timeliness requirements. Within this project machine learning techniques have been applied to retrieve the SO_2 LH information from Sentinel-5p/TROPOMI data to overcome the issue of current state-of-the-art direct fitting approaches to directly retrieve the SO_2 LH, which use computationally expensive radiative transfer calculations.

The retrieval of the SO_2 LH based on Sentinel-5P/TROPOMI measurements is performed using the 'Full-Physics Inverse Learning Machine' algorithm (FP_ILM, Hedelt et al. 2019) and combines a Principal Component analysis (PCA) and a Neural Network (NN) approach to retrieve the SO_2 LH based on Sentinel-5P/TROPOMI backscattered UV Earthshine measurements in the wavelength range between 311 and 335 nm.

In general, the FP_ILM algorithm creates a mapping between the spectral radiance and the atmospheric parameter (here the SO_2 LH) using machine learning methods. The main advantage of the FP_ILM algorithm over classical direct fitting approaches is that the time-consuming training phase involving complex radiative transfer (RT) modelling and NN training is performed offline; the final trained inversion operator itself is robust and computationally simple and therefore extremely fast and can be applied in near-real-time (NRT) processing environments. The algorithm is described in detail in the SO_2 LH Algorithm Theoretical Baseline Document (ATBD, [RD01])

The algorithm has been extensively validated against other datasets, which is described in detail in the SO_2 LH Validation Report (VR, [RD02]) as well as in Koukouli et al. (2021)

3. Analysis and strategic actions for fostering a transition from research to operational activities

The S5P SO_2 LH L2 product that has been generated in the framework of this Sentinel-5p Innovation: SO2LH project has been extensively used and analyzed by several end users, including ECMWF/CAMS, which already actively assimilating the product for forecasting volcanic SO_2 plumes. Building upon their feedback, an implementation into existing operational processing environments as well as an improvement of the algorithm is foreseen. In the following, the roadmap for possible ways of transition of the algorithm into an operational processing environment as well as the scientific way forward is presented.

3.1 Implementation into an operational processing environment

One of the main requirements of near-real time operational processing facilities such as the S5P processing ground segment is the processing speed of an L2 algorithm, i.e. that an L2 NRT algorithm must be fast enough such that the L2 product is generated within 3 h after sensing, see [RD03] and [RD04]. According to [RD01], the S5P SO_2 LH algorithm fulfills this requirement, since the total time to read the L1 data and retrieve the SO_2 LH is estimated to be around 330s for the very pessimistic scenario that about 50,000 pixels (3% of an TROPOMI orbit) need to be retrieved. Note that this estimate is based on the processing speed of a usual personal computer. In the framework of an operational environment with multi-threading, parallelization options etc, the retrieval is of course significantly faster.
The algorithm can in principle be easily transferred to any operational processing facility. The only possible challenge in integrating the developed algorithm into existing large scientific initiatives and operational frameworks is the NeuralNetwork part that has been implemented in Python using the ScikitLearn library. Operational frameworks are usually based on C++ or FORTRAN such that they are fast and easily scalable.

The S5P/TROPOMI SO_2 LH product is currently already generated in a semi-operational quasi-NRT processing environment in the framework of the DLR INPULS project (Innovative Products for Analyses of Atmospheric Composition), i.e. on an hourly basis a script reads all available S5P NRT data and applies the SO_2 LH algorithm to it. The results are immediately pushed to ECMWF/CAMS via FTP.
Note that it is currently foreseen to implement the S5P SO_2 LH algorithm into the Copernicus Sentinel-5P Product Algorithm Laboratory (PAL) environment, hence enabling the testing and application of the algorithm in a quasi-operational processing environment.

3.2 Scientific roadmap

According to several end users, including ECMWF/CAMS, the S5P SO_2 LH L2 product works good for strong & explosive volcanic eruptions, which has been shown for a number of cases, see [RD06], Inness et al. (2022) and Koukouli et al. (2021).
Nevertheless, the following aspects have been identified that will be addressed in future algorithm improvements:

- Application to sources with low SO_2 amounts (i.e. SO_2 <15DU)
- Application to high-altitude SO_2 plumes
- Underestimation of SO_2 LH in the presence of ash/aerosols
- Dependence of SO_2 LH in case of cloud screening
- Improved row dependence

The product already fulfills the scientific requirement of the upcoming Sentinel-5 satellite mission of having an uncertainty of the SO_2 LH to be smaller than 1 km (breakthrough) to 2 km (threshold)

for SO_2 VCD > 25 DU (see [RD03]). Nevertheless, the S5P SO_2 LH algorithm cannot be applied to low/medium SO_2 sources (i.e. SO_2 <15DU) and therefore misses weak eruptions and weak parts of the volcanic plume, which are very important for an end-user. Although the algorithm is in principle applicable for medium SO_2 VCD (i.e. 15-20DU), the retrieved SO_2 LH is biased towards higher LHs.
In the framework of the S5P+I: SO2LH project not only a fine-tuning of the PCA operator and NN retrieval has been performed, but also a sensitivity study in order to improve the retrieval and decrease the lower SO_2 VCD threshold. Unfortunately, no further improvements could be found in this respect so far. Nevertheless, Artificial Neural Networks allow for a multitude of different retrieval approaches, hence in the future, the retrieval approach itself might be revisited by using a different NN architecture or by specifically filtering out only SO_2 related information from the reflectance spectrum using a revisited PCA approach. Furthermore, other dimensionality reduction techniques such as Autoencoders might be considered, which allow for e.g. taking into account non-linear dependencies of the input spectra.
An additional approach to improve the NN training is the usage of information from ground-based in-situ measurements of the true volcanic SO_2 LH as a constraint to train the NN using real data in addition to simulated training data.
Another completely different approach to retrieve the SO_2 LH based on TROPOMI data could build upon a covariance-based retrieval algorithm, which has recently been successful applied to retrieve SO_2 slant columns based on TROPOMI measurements, see They et al. (2021). With this approach, the retrieval of the SO_2 LH for lower SO_2 amounts could in principle be possible.

One important aspect in the algorithm development is the applicability of the algorithm to a broad range of (volcanic) emission sources. Although the algorithm performs well for most volcanic eruptions (even huge and extended SO_2 plumes) it can completely fail for other eruptions, since the atmospheric conditions can vary significantly and the algorithm might not be trained well-enough for special conditions. This was especially the case for the explosive Hunga Tonga Hu'apaai eruption in January 2022 which produced a strong SO_2 plume as high as 30km in the atmosphere – an altitude for which the SO_2 LH retrieval algorithm have not been trained for. These special conditions which can occur sometimes will be addressed in any future algorithm improvements.

In general, SO_2 LH retrieval algorithms developed so far have issues in retrieving the correct SO_2 LH in the presence of volcanic ash and aerosols due to the broad spectral absorption signal of ash/aerosol particles. The retrieved SO_2 LH can be significantly underestimated in this case, which is especially an issue in fresh volcanic plumes, in which ash and SO_2 are usually collocated. In order to compensate for this issue, a dedicated NN operator could be trained with a simulated or measured spectral dataset including ash/aerosol absorption and could be optimized to handle such scenarios. In addition, the information based on the dedicated operational S5P AI and ALH L2 products for the aerosol index and layer height, respectively, could be taken into account in the SO_2 LH retrieval.

Another aspect which could be improved in future updates of the algorithm is the treatment of the instrument pixel row, which directly relates to the viewing angle dependency of the retrieval. Since the TROPOMI instrument features a row-dependent instrument slit function, this dependency must properly be taken into account. Currently, the S5P SO_2 LH algorithm has been trained for a set of instrument rows and in the retrieval the SO_2 LH is interpolated to the measured instrument row. While this approach yielded very good results and no scan-angle dependency of the final results were observed, this approach can be certainly improved, e.g. by taking into account the row dependency of the spectra in the NN training directly.

As a final aspect, which will be investigated in the future is the dependency of the SO_2 LH in the presence of clouds, especially if the SO_2 is located below the cloud layer or mixed with it. Currently no filtering of clouds or a dependency of the retrieval on cloud parameters is included in the retrieval.

3.3 Application to other UV satellite sensors

The FP_ILM SO_2 LH algorithm which has been improved and optimized in the framework of this S5P+I: SO2LH project is very versatile and can be easily applied to other UV satellite sensors. In principle the high-resolution spectral training data only needs to be convolved with the instrument slit function of any other UV instrument and can then be applied after some NN and PCA hyperparameter optimization.
Although this appears like a straight-forward approach, the algorithm development and optimization process in this project has shown that there are many aspects that need to be treated carefully and are not easy understood and need to be investigated for several volcanic eruptive cases.

The S5P SO_2 LH algorithm as well as its related validation methodology can be easily applied to Sentinel-4 and Sentinel-5 observations in the future. Lessons learned on the training of the NN and PCA parametrizations will be used as starting point for the application of the FP_ILM SO_2 LH algorithm to the S4 and S5 instrument characteristics, e.g. wavelength resolution, slit-function, row/viewing zenith angle dependencies, etc. Note that the algorithm has already been applied to both GOME2/Metop (Efremenko et al. 2017) and OMI/Aura (Fedkin et al. 2021) observations with success. As far as the validation is concerned, the scarcity of quality-assured ground-based observations in the locations of most significant volcanoes, renders the inter-satellite validation the most appropriate indicator for establishing the accuracy of any new products. Furthermore, the well-established channels of communication to the ECMWF/CAMS assimilation team will greatly facilitate the quick set-up of forecasting experiments for the new sensors.

3.4 Conclusions

The S5P SO_2 LH algorithm and L2 product is already in a very mature stage such it can be easily implemented in an operational processing environment. The SO_2 LH retrieval is already implemented and performed in a quasi-operational NRT environment in the framework of the DLR INPULS project and actively used by ECMWF/CAMS.
The developed algorithm is very versatile such that it can be easily adopted to any current and future UV satellite instrument, including GEMS, Sentinel-4 and Sentinel-5. It is already used for the (non-operational) SO_2 LH retrieval based on Metop/GOME-2 (Efremenko et al. 2017) and AURA/OMI (Fedkin et al. 2021). Further improvements of the resulting SO_2 LH retrieval might be expected with increased instrumental spectral resolution.

REFERENCES

D.S. Efremenko, D.G.R. Loyola, P. Hedelt, R.J.D. Spurr: **Volcanic SO_2 plume height retrieval from UV sensors using inverse learning machines.** International Journal of Remote Sensing, 38, sup1, 1-27, https://doi.org/10.1080/01431161.2017.1348644, 2017.

N. M. Fedkin, C. Li, N.A. Krotkov, P. Hedelt, D.G. Loyola, R.R. Dickerson, R. Spurr: **Volcanic SO_2 Effective Layer Height Retrieval for OMI Using a Machine Learning Approach**, Atmos. Meas. Tech., 14, 3673–3691, https://doi.org/10.5194/amt-14-3673-2021, 2021.

P. Hedelt, D.S. Efremenko, D.G. Loyola, R. Spurr, L. Clarisse: **SO_2 Layer Height retrieval from Sentinel-5 Precursor/TROPOMI using FP_ILM,**
Atmos. Meas. Tech., 12, 5503–5517, https://doi.org/10.5194/amt-12-5503-2019, 2019

A. Inness, M. Ades, D. Balis, D. Efremenko, J. Flemming, P. Hedelt, M.E. Koukouli, D. Loyola, and R. Ribas: **Evaluating the assimilation of S5P/TROPOMI near real-time SO_2 columns and layer height data into the CAMS integrated forecasting system (CY47R1), based on a case study of the 2019 Raikoke eruption**, Geosci. Model Dev., 15, 971–994, https://doi.org/10.5194/gmd-15-971-2022, 2022.

M.E. Koukouli, K. Michailidis, P. Hedelt, I. A. Taylor, A. Inness, L. Clarisse, D. Balis, D. Efremenko, D. Loyola, R. G. Grainger, C. Retscher: **Volcanic SO2 Layer Height by TROPOMI/S5P; validation against IASI/MetOp and CALIOP/CALIPSO observations**, submitted to ACP, 2021, *under review.*

N. Theys, V. Fioletov, C. Li, I. De Smedt, C. Lerot, C. McLinden, N. Krotkov, D. Griffin, L. Clarisse, P. Hedelt, D. Loyola, T. Wagner, V. Kumar, A. Innes, R. Ribas, F. Hendrick, J. Vlietinck, H. Brenot, M. Van Roozendael: **A sulfur dioxide Covariance-Based Retrieval Algorithm (COBRA): application to TROPOMI reveals new emission sources**, Atmos. Chem. Phys., 21, 16727–16744, https://doi.org/10.5194/acp-21-16727-2021, 2021.