SEVIRI FIR Update Validation Report

Doc.No.	:	EUM/RSP/REP/20/1161182
Issue	:	v1D e-signed
Date	:	9 December 2020
WBS/DBS	:	

EUMETSAT
Eumetsat-Allee 1, D-64295 Darmstadt, Germany
Tel: +49 6151 807-7
Fax: +49 6151 807 555
http://www.eumetsat.int

© EUMETSAT
The copyright of this document is the property of EUMETSAT.

Change Record

<List all approved versions and summarise the changes below. Ensure Version/Date in Change Record & Document Headers are the same as on the Profile in the DM Tool.>

Version	**Date**	**DCR* No.** *if applicable*	**Description of Changes**
1	01/04/2020		First issue
1A	23/04/2020		Update of document after first reviews
1B	13/10/2020		Update of document after final reviews
1C	02/11/2020		Final version for internal approval
1D	09/12/2020		Same as version 1C for publication

DCR = Document Change Request

Table of Contents

Table of Figures

Table of Tables

1 INTRODUCTION

1.1 Purpose and Scope

This document presents the validation activities for the updated MSG SEVIRI active fire monitoring (FIR) product. The update introduces improvements to the detection algorithm, the addition of a fire probability indicator, and an improvement of the land-sea mask.

The algorithm update was introduced after Greece and Bulgaria raised an anomaly on FIR in October 2018, reporting that significant forest fires close to coastlines on Greek islands were not properly detected in the 2013-2018 period. On the 13.06.2019, the production of the updated FIR product started in the MPEF VAL processing chain, in parallel to the original algorithm being nominally produced on the MPEF OPE chain. On the 08.08.2019, the updated FIR product was then rolled-out on MPEF OPE in the release 2.8.

The validation is based on the comparison of the updated and original FIR products against comparable fire detection products from the MODIS and VIIRS LEO missions. Additionally, also the LSA-SAF Fire Detection and Monitoring (FD&M) product based on SEVIRI data was compared. The validation was performed over a total of three continuous weeks of MSG data in the 2019 summer.

1.2 Applicable Documents

	Document Title	Reference
AD-1	Active Fire Monitoring with MSG - Algorithm Theoretical Basis Document	EUM/MET/REP/07/0170 v2 (Original Algorithm)
AD-2	MSG ATBD for Active Fire Monitoring	EUM/RSP/DOC/20/1161774 (Updated Algorithm)

1.3 Reference Documents

<List any reference material for gaining a better understanding of this source document.>

	Document Title	Reference
RD-1	MODIS Active Fire Product User Guide	http://modis-fire.umd.edu/files/MODIS_C6_Fire_User_Guide_B.pdf
RD-2	MODIS Active Fire Product Algorithm Theoretical Basis Document	https://eospso.nasa.gov/sites/default/files/atbd/atbd_mod14.pdf
RD-3	VIIRS Active Fire Product User Guide	https://lpdaac.usgs.gov/documents/427/VNP14_User_Guide_V1.pdf
RD-4	VIIRS Active Fire Product Algorithm Theoretical Basis Document	https://lpdaac.usgs.gov/documents/133/VNP14_ATBD.pdf
RD-5	LSA-SAF Fire Detection and Monitoring Product User Guide	https://landsaf.ipma.pt/GetDocument.do?id=672
RD-6	LSA-SAF Fire Detection and Monitoring Algorithm Theoretical Basis Document	https://landsaf.ipma.pt/GetDocument.do?id=671

1.4 Terminology

Acronyms and Abbreviations

Acronym/Abbr.	Explanation
ATBD	Algorithm Theoretical Basis Document
FD&M	LSA-SAF Fire Detection and Monitoring Product
F1	F1 Score/Measure
FIR	SEVIRI active fire monitoring product
GEO	Geostationary Orbit
IR	Infrared
LEO	Low Earth Orbit
LSA-SAF	Land Surface Analysis – Satellite Applications Facility
MODIS	Moderate Resolution Imaging Spectroradiometer
MPEF	Meteorological Product Extraction Facility
MSG	Meteosat Second Generation
OPE	Operational
POD	Probability of Detection
PRE	Precision
ROI	Region Of Interest
SEVIRI	Spinning Enhanced Visible and Infrared Imager
VAL	Validation
VIIRS	Visible Infrared Imaging Radiometer Suite

Definitions

Definition/Term	Explanation
FIRProb	Fire Probability Indicator in the updated (New) FIR product
New FIR	Updated SEVIRI FIR algorithm (operational since 08.08.2019, MPEF release 2.8).
Old FIR	Original SEVIRI FIR algorithm

1.5 Document Structure

Section 2 describes the principle of the FIR detection algorithm, together with a description of the algorithm updates that lead to this validation.
Section 3 describes the validation setup, including the used data, the adopted methodology and monitored metrics.
Section 4 reports and discusses the results obtained from the validation. The results are split in two parts: quantitative results analysing the full dataset and day/night splits, and qualitative results analysing a test case.
Section 5 presents the known limitations of the methodology and reports conclusions and indications for users.

2 FIR ALGORITHM DESCRIPTION

2.1 Original Algorithm (Old FIR)

This section provides a brief overview of the original fire detection algorithm as implemented before the update. This algorithm and the resulting data is denoted as "Old FIR" throughout this document. The physics background and the full detailed algorithm description are provided in [AD-1].

The algorithm is based on a series of threshold tests using the 3.9 µm and 10.8 µm channels. The algorithm is applied for every repeat cycle and over all land surface pixels, excluding desert/bare soil surface and coastal pixels. The bare soil pixel classification is extracted from climatological background information for the MSG field of view, in addition to a IR10.8-IR8.7 difference test. For the processed pixels, the FIR algorithm uses the following five criteria to check for potential fire pixels, while reducing the false alarms:

1. Brightness temperature of channel IR3.9 has to exceed a certain threshold
2. Brightness temperature difference of channels IR3.9 and IR10.8 has to exceed a certain threshold
3. Difference of the standard deviations of channel IR3.9 and IR10.8 has to exceed a certain threshold
4. Standard deviation of channel IR10.8 has to be lower than a certain threshold
5. Standard deviation of channel IR3.9 has to exceed a certain threshold

Each criteria is checked against computed thresholds, following the rationale that the IR3.9 channel is highly sensitive to fire hotspots, while fires affect considerably less the IR10.8 channel. The standard deviation is calculated on a 3x3 pixel array around each MSG pixel. The thresholds are computed dynamically using predicted backgrounds extracted from ECMWF forecasts. Each test is performed iteratively in a binary passed/non-passed way. A pixel is classified as "possible" fire pixel if all tests are passed. A further set of conditions with stricter thresholds further classifies "probable" fire pixels. Separated sets of threshold computation coefficients are applied for day and night operation.
The full tests explanation and the threshold computation formulas, including the numerical coefficients, are available in [AD-1].

2.2 Updated Algorithm (New FIR)

The following sections describe the different updates on the FIR algorithm. The updated algorithm is denoted as "New FIR" throughout this document. The physics background and the full detailed algorithm description are provided in [AD-2].

2.2.1 Revision of Performed Tests

The standard deviation test on the IR3.9 (criterion nr. 5) was found not to provide any additional information, and was therefore removed from the algorithm. Moreover, the standard deviation test on the IR10.8 channel (criterion nr. 4) eliminated too many real fires in regions with inhomogeneous surfaces, and was thus removed as well.

2.2.2 Introduction of Bayesian-Type Filtering and Fire Probability Indicator

In replacement of the binary tests of the original algorithm, a new set of tests with a Bayesian approach was implemented. Following criteria are being tested:

1. Brightness temperature of channel IR3.9 has to exceed a certain threshold
2. Brightness temperature difference of channels IR3.9 and IR10.8 has to exceed a certain threshold
3. Difference of the standard deviations of channel IR3.9 and IR10.8 has to exceed a certain threshold

Each individual test has a minimum and maximum threshold. Values below the minimum threshold indicate no fire (confidence=0), while values above the maximum threshold indicate fire (confidence=1). Values between the thresholds are linearly interpolated between 0 and 1. The combination of all tests, through consecutive multiplications, yields a fire probability indicator (FIRProb) between 0 and 100%. This fire probability indicator is now available in the product. "Possible" fires (mid confidence) are defined as pixels with a probability higher or equal than 40% and lower than 80%, while "probable" fires (high confidence) have a probability higher or equal than 80%.

2.2.3 Improvement of the Sea-Land Mask

The handling of areas close to water bodies and coasts was optimized for this update. In the original algorithm, a minimum distance to water bodies was defined, below which a pixel would not be processed. In this update, pixels marked as water and mixed land/water are not processed, while land pixels are always processed disregard of their distance to water bodies. This enables the detection of fires closer to coastlines without drastically increasing the false alarms.

3 VALIDATION SETUP

3.1 Validation Period

During the validation period, data from both Old FIR (OPE chain) and New FIR (VAL chain) algorithms was stored for a total of three weeks (18.06.2019 – 24.06.2019 and 15.07.2019 – 31.07.2019), containing 2254 SEVIRI repeat cycles (full Earth disk scans with 15 min periods). The validation described in this document was performed over this three-week period.

3.2 Validation Data

3.2.1 MODIS and VIIRS Comparison Data

For the three validation weeks, comparison data from the LEO instruments MODIS and VIIRS was fetched from the LAADS DAAC portal. For the swaths query, an ROI over central Africa with the bounding box coordinates [W -17.2, N 19.5, E 52.6, S -29.6] was selected to concentrate the analysis over an area with frequently detected fires. Figure 1 depicts the ROI area in the data portal. Swaths from all available pass times (day/night/terminator), intersecting the defined ROI, were included.

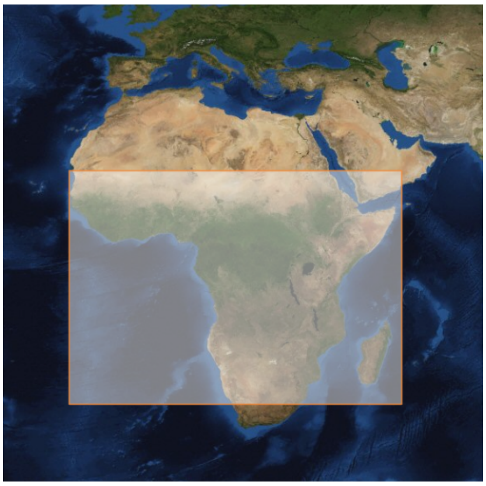

Figure 1: Selected ROI for LEO fire products query.

Additional information about the LEO fire products is provided in the following:

- MODIS active fires product [RD-1, RD-2]
    - L2 Product name: MOD14 (Terra satellite), MYD14 (Aqua satellite)
    - Spatial resolution: 1 km
    - Data provided as segments of orbital swaths (approx. 5 min of recording, 2340 × 2030 km along-scan and along-track size)
    - Swath pixels geolocation separately contained respectively in the MOD03 and MYD03 geolocation products
    - Data from reprocessing collection 6
- VIIRS active fires product [RD-3, RD-4]
    - L2 Product name: VNP14IMG (Suomi/NPP satellite)
    - Spatial resolution: 375 m

- Data provided as segments of orbital swaths (approx. 6 min of recording, 3060 km along-scan size)
- Swath pixels geolocation separately contained in the VNP03IMGLL geolocation product.
- Data from reprocessing collection 1

For both MODIS and VIIRS, the product contains a fire mask with flags for detected fires (7 = low confidence fire, 8 = nominal confidence fire, 9 = high confidence fire).

3.2.2 LSA-SAF Comparison Data

For an additional validation comparison, data from the LSA-SAF Fire Detection and Monitoring (FD&M) product was downloaded for the same three-week period. While LSA-SAF also produces a fire radiative power product (FRP-PIXEL) with a more optimised algorithm, we selected the FD&M product for this comparison as it is an equivalent product to FIR (active fire detection only). As it originates from SEVIRI data, this product has the same temporal and geometric properties as the FIR products. The fire mask in the product is in the native geostationary projection with full-disk view, with a single flag for detected fires (no further confidence level is provided). Further information about the product and the algorithm can be found in [RD-5, RD-6].

3.3 Validation Methodology

3.3.1 Principle and Workflow

The validation is based on the comparison of detected fires by the "GEO" products (Old FIR, New FIR, LSA-SAF) and the "LEO" products (MODIS, VIIRS). Given the higher resolution of LEO products, the assumption is made that the fire detection capability is better; in general, fires detected by GEO products should be present in LEO products as well. The inverse, however, is not necessarily true. To mitigate this limitation, further studies on aggregation of LEO fire pixels were performed (see section 3.3.2). In the computation of the statistical scores, the LEO products are regarded as the "ground-truth".

The validation workflow is described in Figure 2. The process is repeated for every GEO repeat cycle (Old FIR/New FIR/LSA-SAF) and respectively for every LEO product (MODIS/VIIRS).

Figure 2: Validation workflow for each GEO fires full disk.

Fire pixels from the LEO products are mapped to the GEO pixel grid. The applied resampling method was "bucket" resampling, where each LEO fire pixel is assigned to the GEO pixel that contains its pixel centre. Therefore, each LEO fire contributes only to one SEVIRI GEO grid pixel; it is however possible, that multiple fires from the LEO product are mapped to the same GEO pixel. Note that for the last comparison step the geolocation information from the full swath is required to compare only the GEO fires located within the LEO swath footprint. For this, the full swath is resampled to the GEO grid using a nearest-neighbour approach, unprocessed pixels from the MODIS/VIIRS fire product are disregarded.

The algorithm is based on the Earth observation satellite data processing framework *Pytroll*[1]. Specifically, the code leverages the *Satpy*[2] package for data reading and *Pyresample*[3] package for data resampling.

3.3.2 Resolution Difference Compensation

In a simplified model disregarding geometric pixel deformations, one grid point in the SEVIRI product at nadir (3 km resolution) can fit 9 MODIS pixels (1 km resolution) or 64 VIIRS pixels (0.375 km resolution). A further analysis was run to assess the effects of this resolution difference on the validation results.
In this analysis, after the resampling of the higher resolution MODIS/VIIRS product to the SEVIRI grid, a grid point was considered to contain a fire only if a specified minimum number of MODIS/VIIRS fire pixels were detected within that same grid point. This compensates, at least partially, for the resolution difference, as only larger, wider-spread MODIS/VIIRS fires are considered in the comparison. Different minimum number combinations, considering low/medium confidence and high confidence fires separately, were analysed.

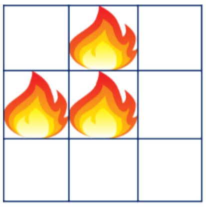

Figure 3: A schematic illustration of three MODIS fires detected within one SEVIRI grid point. The example here shows the case for a minimum threshold of 3.

3.4 Validation Metrics

Once the resampling of a LEO product to a GEO grid was performed and the applicable minimum threshold was set, each GEO grid point within the LEO swath was classified as follows:

- False positive (FP) – false alarm: fire detected in GEO but not in LEO
- False negative (FN) – miss: fire not detected in GEO but detected in LEO
- True positive (TP) – hit: fire detected in both GEO and LEO

[1] http://pytroll.github.io/
[2] https://github.com/pytroll/satpy
[3] https://github.com/pytroll/pyresample

True negative pixels (no-fire pixels in both products) bear no value for this analysis and are therefore not considered.

For the full validation, FP/FN/TP counts were accumulated over all repeat cycles and swaths and the final comparison statistics were computed.

In the following, the list of the used metrics with the respective properties and interpretations are given:

- Probability of Detection (POD)
  - Definition: TP / (TP + FN)
  - Recall, sensitivity of the detector
    - What is the probability of a real fire to be detected by GEO?
    - Amount of correct GEO detections (TP) among all fires detected by LEO (TP+FN)
- Precision (PRE)
  - Definition: TP / (TP + FP)
  - Complementary to the False Alarm Ratio (FAR = 1- PRE)
  - Quality of detections
    - What is the probability of a fire detected by GEO to be a real fire?
    - Amount of correct GEO detections (TP) among all fires detected by GEO (TP+FP)
- F1 Measure/Score (F1)
  - $2 \times \frac{\mathrm{POD} \times \mathrm{PRE}}{\mathrm{POD} + \mathrm{PRE}}$
  - Harmonic mean of POD and PRE
  - Overall performance of detector / algorithm.

The design and tuning of such a binary classification algorithm is a trade-off between probability of detection and precision. The F score, a combination of these two values, was monitored in this study to give a single measure of performance of the detection algorithms. The F1 score was specifically selected to give equal weight to both POD and PRE parameters. Depending on the specific application of the data, however, more weight might be desired in one or the other direction, if more cost is put on missed detections or false alarms. Figure 4 compares the harmonic mean (F1 score) with the geometric mean $\frac{\mathrm{POD} + \mathrm{PRE}}{2}$. Note how a high F1 score can only be achieved by having both high PRE and POD values simultaneously.

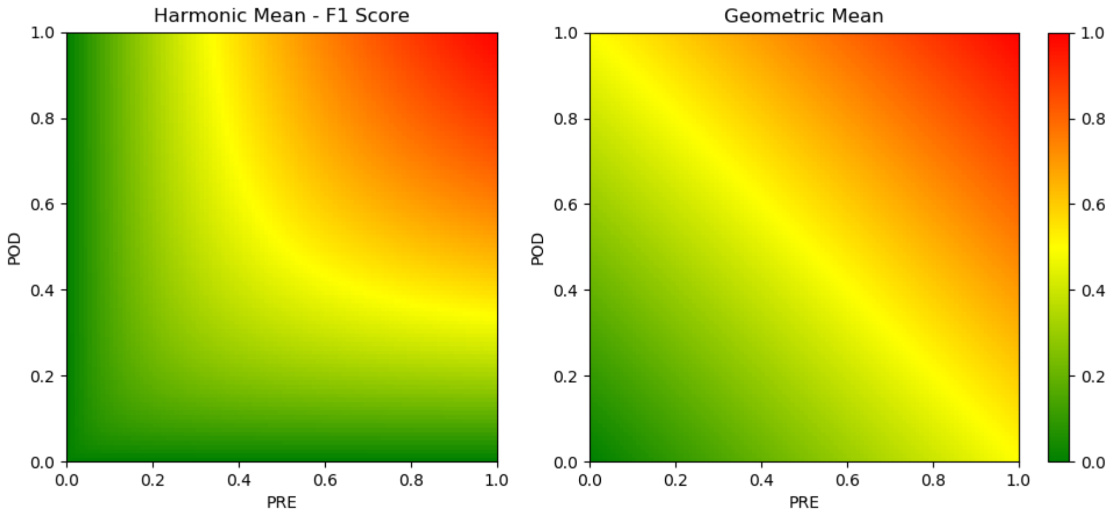

Figure 4: Comparison of harmonic mean (F1 score) with geometric mean of POD and PRE.

4 VALIDATION RESULTS

This chapter reports and discusses the results obtained from the validation methodology described above.

4.1 Quantitative Results

The following sections present the analysis of the quantitative results obtained from the validation comparison.

In each section, tables report the numerical results in terms of POD, PRE and F1 from the comparison of different SEVIRI products with the MODIS and VIIRS active fire detection products.
In each row, results for different SEVIRI products are listed. For the New FIR algorithm, different fire probability indicator (FIRProb) thresholds are analysed. A FIRProb threshold of 40% is currently used to define the lower bound for "possible" (mid confidence) fires (see section 2.2.2). The entry "New FIR 40%" is therefore equivalent to the full fire mask as provided in the product (possible and probable fires combined). The entry "Old FIR" is also to be understood as the full fire mask as in the product. The entry LSA-SAF indicates the results obtained by the comparison with the LSA-SAF FD&M active fire detection product.
In each column, different minimum fire numbers for MODIS/VIIRS fire pixels are indicated, considering all fires (fire mask values 7,8,9) as well as considering only high confidence fires (fire mask values 9).
The table colour coding is selected to aid the reading of the table, with red colours indicating lower, worse scores and green colours indicating higher, better scores.

4.1.1 Full Dataset Analysis

The full dataset comprises of 2254 SEVIRI full disks, out of which 947 contain at least one MODIS or VIIRS swath, and were therefore processed further. In total, 700 disks contained at least one MODIS swath, and 421 at least one VIIRS swath.
Table 1 reports the absolute number of detections included in the analysis for different products.

Product	*Total number of detections*
SEVIRI Products	
New FIR 5%	237246
New FIR 10%	148155
New FIR 20%	89094
New FIR 30%	66277
New FIR 40%	53238
Old FIR	66019
New FIR Probable Only	25630
Old FIR Probable Only	18948
LSA-SAF	18731
LEO Products	
MODIS all confidences	172222
MODIS high confidence	46343
VIIRS all confidences	502833
VIIRS high confidence	54600

Table 1: Total number of detections in different products for the analysed full dataset.

4.1.1.1 MODIS Comparison

Table 2 presents the quantitative results from the MODIS comparison. The table structure and the different items are explained in section 4.1.

		MODIS Fires									
		All							High Confidence Only		
		min 1	min 2	min 3	min 4	min 5	min 6	min 7	min 1	min 2	min 3
POD	New FIR 5%	0.247	0.405	0.608	0.764	0.891	0.927	0.966	0.499	0.704	0.835
	New FIR 10%	0.210	0.355	0.555	0.717	0.864	0.912	0.966	0.447	0.658	0.810
	New FIR 20%	0.161	0.286	0.474	0.641	0.807	0.877	0.925	0.368	0.582	0.756
	New FIR 30%	0.129	0.235	0.409	0.570	0.757	0.848	0.908	0.310	0.518	0.700
	New FIR 40%	0.107	0.199	0.354	0.508	0.695	0.783	0.851	0.266	0.459	0.645
	Old FIR	0.126	0.232	0.408	0.574	0.753	0.845	0.902	0.306	0.517	0.707
	New FIR Probable Only	0.055	0.109	0.213	0.324	0.491	0.582	0.690	0.150	0.287	0.447
	Old FIR Probable Only	0.048	0.101	0.206	0.322	0.497	0.597	0.701	0.137	0.276	0.447
	LSA-SAF	0.054	0.120	0.253	0.406	0.606	0.722	0.839	0.157	0.328	0.536
PRE	New FIR 5%	0.180	0.088	0.033	0.014	0.005	0.002	0.001	0.097	0.038	0.012
	New FIR 10%	0.244	0.124	0.049	0.021	0.008	0.003	0.001	0.140	0.057	0.019
	New FIR 20%	0.311	0.165	0.069	0.031	0.012	0.005	0.002	0.191	0.083	0.030
	New FIR 30%	0.336	0.183	0.080	0.037	0.015	0.007	0.002	0.217	0.100	0.037
	New FIR 40%	0.346	0.193	0.086	0.042	0.018	0.008	0.003	0.231	0.110	0.042
	Old FIR	0.329	0.182	0.080	0.038	0.015	0.007	0.002	0.215	0.100	0.037
	New FIR Probable Only	0.369	0.220	0.108	0.055	0.026	0.012	0.005	0.271	0.143	0.061
	Old FIR Probable Only	0.435	0.275	0.141	0.074	0.035	0.016	0.006	0.335	0.186	0.083
	LSA-SAF	0.496	0.332	0.175	0.094	0.043	0.020	0.008	0.387	0.223	0.100
F1	New FIR 5%	0.208	0.145	0.063	0.028	0.010	0.004	0.001	0.163	0.072	0.024
	New FIR 10%	0.226	0.183	0.089	0.041	0.016	0.006	0.002	0.213	0.104	0.037
	New FIR 20%	0.212	0.209	0.121	0.060	0.024	0.010	0.004	0.252	0.146	0.057
	New FIR 30%	0.187	0.206	0.134	0.070	0.030	0.013	0.005	0.255	0.167	0.070
	New FIR 40%	0.163	0.196	0.139	0.077	0.034	0.015	0.006	0.247	0.178	0.080
	Old FIR	0.182	0.204	0.134	0.071	0.030	0.013	0.005	0.252	0.168	0.071
	New FIR Probable Only	0.096	0.146	0.143	0.094	0.049	0.023	0.009	0.193	0.191	0.107
	Old FIR Probable Only	0.086	0.148	0.167	0.120	0.066	0.032	0.013	0.194	0.222	0.139
	LSA-SAF	0.097	0.177	0.207	0.153	0.081	0.039	0.015	0.223	0.266	0.169

Table 2: Quantitative results from comparisons with MODIS fire product – full dataset.

Several observations can be made by evaluating the table.

- Higher numbers of minimum LEO fires per pixel increase the POD, since only larger fires remain present for comparison. Large fires are generally easier to detect for GEO products. Concurrently, the PRE scores decrease as some of the smaller LEO fires correctly detected by GEO (TP) are removed and become false alarms (FP). These two effects are combined in the F1 score, which shows maxima in the FIR products for minimum fire numbers of 1 and 2. The LSA-SAF product has the maximum for all MODIS fires at min 3, due to a relatively high PRE that overcomes a low POD.
- Comparing only the MODIS high confidence fires has a similar effect as increasing the number of minimum fire pixels, namely lower PRE and higher POD. However, in terms of F1 score, the performance is generally higher, marking the absolute maximum values for FIR products at New FIR 30%. LSA-SAF achieves its maximum at min 2 fires per pixel.
- Decreasing the FIRProb threshold for New FIR includes more detections in the product. Part of these low-FIRProb detections were also detected by MODIS and transform from FN to TP, increasing the POD; part of them become false alarms and decrease the PRE. The two effects are therefore competing for the F1 score; the gain in POD overcomes the

loss in PRE until 10% FIRProb, where the false alarms become too significant and the performance drops.

- The New FIR product at 30% threshold achieves similar results as the Old FIR.
- The probable-only fire masks for New and Old FIR reach similar F1 results, with the new product having slightly worse precision but better probability of detection.

4.1.1.2 VIIRS Comparison

Table 3 reports the numerical results in terms of POD, PRE and F1 from the comparison of different SEVIRI products with the VIIRS active fire detection product. See section 4.1 for the table structure description.

		VIIRS Fires									
		All							High Confidence Only		
		min 1	min 2	min 3	min 4	min 5	min 6	min 7	min 1	min 2	min 3
POD	New FIR 5%	0.132	0.210	0.298	0.357	0.399	0.414	0.409	0.365	0.414	0.377
	New FIR 10%	0.108	0.177	0.257	0.315	0.358	0.379	0.379	0.323	0.380	0.358
	New FIR 20%	0.080	0.134	0.200	0.252	0.294	0.318	0.325	0.263	0.324	0.319
	New FIR 30%	0.063	0.107	0.163	0.208	0.248	0.272	0.281	0.219	0.280	0.286
	New FIR 40%	0.051	0.088	0.136	0.175	0.213	0.236	0.248	0.188	0.248	0.261
	Old FIR	0.060	0.104	0.159	0.205	0.246	0.271	0.282	0.216	0.279	0.285
	New FIR Probable Only	0.025	0.044	0.070	0.093	0.118	0.136	0.147	0.105	0.152	0.174
	Old FIR Probable Only	0.022	0.039	0.064	0.087	0.112	0.130	0.143	0.098	0.146	0.168
	LSA-SAF	0.026	0.050	0.083	0.114	0.147	0.172	0.189	0.125	0.184	0.211
PRE	New FIR 5%	0.265	0.197	0.145	0.108	0.080	0.061	0.046	0.080	0.037	0.018
	New FIR 10%	0.323	0.246	0.185	0.141	0.107	0.083	0.064	0.105	0.050	0.025
	New FIR 20%	0.377	0.294	0.227	0.178	0.139	0.110	0.086	0.135	0.067	0.035
	New FIR 30%	0.395	0.313	0.247	0.197	0.156	0.126	0.100	0.150	0.078	0.042
	New FIR 40%	0.400	0.319	0.255	0.206	0.166	0.135	0.109	0.159	0.085	0.047
	Old FIR	0.393	0.314	0.249	0.200	0.160	0.129	0.103	0.153	0.080	0.043
	New FIR Probable Only	0.411	0.335	0.276	0.229	0.193	0.163	0.135	0.186	0.109	0.066
	Old FIR Probable Only	0.492	0.415	0.351	0.298	0.255	0.218	0.183	0.243	0.147	0.089
	LSA-SAF	0.555	0.487	0.421	0.361	0.309	0.266	0.224	0.285	0.170	0.103
F1	New FIR 5%	0.176	0.204	0.195	0.166	0.134	0.107	0.083	0.131	0.067	0.034
	New FIR 10%	0.162	0.206	0.215	0.195	0.165	0.136	0.109	0.158	0.088	0.046
	New FIR 20%	0.132	0.184	0.213	0.209	0.189	0.164	0.136	0.178	0.111	0.063
	New FIR 30%	0.108	0.159	0.196	0.202	0.192	0.172	0.147	0.178	0.122	0.073
	New FIR 40%	0.091	0.138	0.177	0.189	0.187	0.172	0.151	0.172	0.127	0.080
	Old FIR	0.105	0.156	0.194	0.202	0.194	0.175	0.151	0.179	0.124	0.075
	New FIR Probable Only	0.047	0.078	0.112	0.133	0.146	0.148	0.141	0.134	0.127	0.096
	Old FIR Probable Only	0.041	0.072	0.108	0.134	0.155	0.163	0.160	0.140	0.146	0.117
	LSA-SAF	0.050	0.090	0.139	0.174	0.199	0.209	0.205	0.174	0.177	0.138

Table 3: Quantitative results from comparisons with VIIRS fire product – full dataset.

The general trends observed in the MODIS comparison are present in the VIIRS comparison as well. Further observations are:

- The POD values are lower than in the MODIS comparison, which is expected due to the capability of VIIRS to detect even smaller fires at a product resolution of 375 m.

- The PRE values are higher than for the MODIS comparison. This signifies that some of the fires classified as false alarms (FP) in the MODIS case, are converted to hits (TP) when comparing with VIIRS, indicating a better affinity between the detection algorithms.
- The low POD overcomes the higher PRE and decreases the F1 scores with respect to the MODIS comparison.
- It can be observed how F1 maxima are generally shifted to higher minimum fires numbers than for MODIS, which is expected due to the higher resolution. The absolute maximum is reached by New FIR at 10% with min 3 fires per pixel. The LSA-SAF product reaches the F1 performance maximum at six fires per pixel.

4.1.2 Day/Night Analysis

The following sections present the comparison results achieved by splitting the dataset in day and night cases. Day is defined for SEVIRI repeat cycles starting between 6.00 and 18.00 UTC (479 full-disks), and night the remaining hours (468 full-disks).

Product	***Total number of detections***	
	Day-time	***Night-time***
SEVIRI Products		
New FIR 5%	231402	5844
New FIR 10%	143289	4866
New FIR 20%	85380	3714
New FIR 30%	63264	3013
New FIR 40%	50722	2516
Old FIR	61195	4824
New FIR Probable Only	24311	1319
Old FIR Probable Only	17559	1389
LSA-SAF	18477	254
LEO Products		
MODIS all confidences	162974	9248
MODIS high confidence	42430	3913
VIIRS all confidences	427291	75542
VIIRS high confidence	47579	7021

Table 4: Total number of detections in different products for the analysed day/night dataset.

4.1.2.1 MODIS Comparison

Table 5 reports the results from the day/night analysis with MODIS data. The columns marked as “All” are equivalent to the “All - min 1” columns in the previous tables; the columns “H.c.” are equivalent to the previous “High Confidence Only – min 1”.

		MODIS Fires					
		No Filter		Night Only		Day Only	
		All	H.c.	All	H.c.	All	H.c.
POD	New FIR 5%	0.247	0.499	0.206	0.319	0.250	0.515
	New FIR 10%	0.210	0.447	0.182	0.290	0.211	0.462
	New FIR 20%	0.161	0.368	0.149	0.248	0.162	0.379
	New FIR 30%	0.129	0.310	0.126	0.215	0.130	0.319
	New FIR 40%	0.107	0.266	0.111	0.193	0.107	0.272
	Old FIR	0.126	0.306	0.172	0.285	0.124	0.308
	New FIR Probable Only	0.055	0.150	0.063	0.119	0.054	0.153
	Old FIR Probable Only	0.048	0.137	0.072	0.135	0.047	0.137
	LSA-SAF	0.054	0.157	0.021	0.044	0.056	0.167
PRE	New FIR 5%	0.180	0.097	0.327	0.213	0.176	0.094
	New FIR 10%	0.244	0.140	0.347	0.233	0.240	0.137
	New FIR 20%	0.311	0.191	0.371	0.261	0.309	0.188
	New FIR 30%	0.336	0.217	0.387	0.279	0.334	0.214
	New FIR 40%	0.346	0.231	0.410	0.301	0.343	0.228
	Old FIR	0.329	0.215	0.330	0.231	0.329	0.213
	New FIR Probable Only	0.369	0.271	0.441	0.352	0.365	0.266
	Old FIR Probable Only	0.435	0.335	0.478	0.382	0.432	0.331
	LSA-SAF	0.496	0.387	0.768	0.677	0.492	0.383
F1	New FIR 5%	0.208	0.163	0.253	0.256	0.206	0.160
	New FIR 10%	0.226	0.213	0.239	0.258	0.225	0.211
	New FIR 20%	0.212	0.252	0.213	0.254	0.212	0.252
	New FIR 30%	0.187	0.255	0.190	0.243	0.187	0.256
	New FIR 40%	0.163	0.247	0.175	0.235	0.163	0.248
	Old FIR	0.182	0.252	0.226	0.255	0.180	0.252
	New FIR Probable Only	0.096	0.193	0.110	0.177	0.095	0.194
	Old FIR Probable Only	0.086	0.194	0.125	0.200	0.084	0.194
	LSA-SAF	0.097	0.223	0.041	0.083	0.100	0.233

Table 5: Quantitative results from comparisons with MODIS fire product applying night-only and day-only filters.

Generally, the trends for the FIR products are similar for the day and night cases. The overall results (“No Filter”) are closer to the day-time results due to the much higher number of detections during day that dominate the statistics (see Table 4 for the absolute numbers of detections).
Some more observations can be made:

- Generally, lower POD and higher PRE are registered during night-time.
- The F1 scores are similar between day and night for the high confidence only (H.c.) cases, the drop in performance after New FIR 10% is however not present for the night case due to a better balanced PRE and POD.
- All FIR results perform slightly better during night-time when compared to all MODIS fires. It can be however observed how Old FIR has a better performance during night-time than all New FIR with a FIRProb threshold higher than 10% (while for the day-only case

the performance is matched and surpassed above 30%). This is mostly due to a better balance between POD and PRE in the night-time Old FIR results.

- The LSA-SAF product shows particularly low POD and high PRE during night-time, which result in a low F1 score.

4.1.2.2 VIIRS Comparison

Table 6 reports the comparison results for the day/night analysis with VIIRS data.

		VIIRS Fires					
		No Filter		Night Only		Day Only	
		All	H.c.	All	H.c.	All	H.c.
POD	New FIR 5%	0.132	0.365	0.022	0.048	0.151	0.411
	New FIR 10%	0.108	0.323	0.019	0.044	0.124	0.365
	New FIR 20%	0.080	0.263	0.015	0.039	0.091	0.296
	New FIR 30%	0.063	0.219	0.012	0.034	0.072	0.246
	New FIR 40%	0.051	0.188	0.010	0.031	0.059	0.211
	Old FIR	0.060	0.216	0.018	0.046	0.068	0.242
	New FIR Probable Only	0.025	0.105	0.005	0.020	0.029	0.117
	Old FIR Probable Only	0.022	0.098	0.006	0.024	0.024	0.109
	LSA-SAF	0.026	0.125	0.002	0.009	0.031	0.142
PRE	New FIR 5%	0.265	0.080	0.464	0.094	0.262	0.080
	New FIR 10%	0.323	0.105	0.479	0.104	0.320	0.105
	New FIR 20%	0.377	0.135	0.490	0.121	0.374	0.135
	New FIR 30%	0.395	0.150	0.499	0.128	0.392	0.151
	New FIR 40%	0.400	0.159	0.508	0.145	0.397	0.159
	Old FIR	0.393	0.153	0.467	0.112	0.390	0.154
	New FIR Probable Only	0.411	0.186	0.538	0.186	0.408	0.186
	Old FIR Probable Only	0.492	0.243	0.577	0.206	0.489	0.244
	LSA-SAF	0.555	0.285	0.843	0.431	0.553	0.284
F1	New FIR 5%	0.176	0.131	0.042	0.064	0.192	0.133
	New FIR 10%	0.162	0.158	0.037	0.062	0.178	0.163
	New FIR 20%	0.132	0.178	0.029	0.059	0.147	0.185
	New FIR 30%	0.108	0.178	0.024	0.054	0.121	0.187
	New FIR 40%	0.091	0.172	0.020	0.052	0.102	0.182
	Old FIR	0.105	0.179	0.034	0.065	0.116	0.188
	New FIR Probable Only	0.047	0.134	0.011	0.036	0.054	0.144
	Old FIR Probable Only	0.041	0.140	0.012	0.043	0.046	0.151
	LSA-SAF	0.050	0.174	0.003	0.018	0.058	0.189

Table 6: Quantitative results from comparisons with VIIRS fire product applying night-only and day-only filters.

The observations made for the MODIS case are also valid for this comparison. However, the lower POD and higher PRE differences during night-time are more distinct, leading to a general drop of F1 scores during night for all SEVIRI products.

4.2 Qualitative Results - Test Case Analysis

In this section, the detections on a selected test case are analysed. The SEVIRI repeat cycle with epoch 17.07.2019, 12:30 UTC was selected due to the high number of detected fires. Figure 5 shows the selected repeat cycle full-disk including the LEO swath footprint, the GEO and LEO fires, and the selected test area.

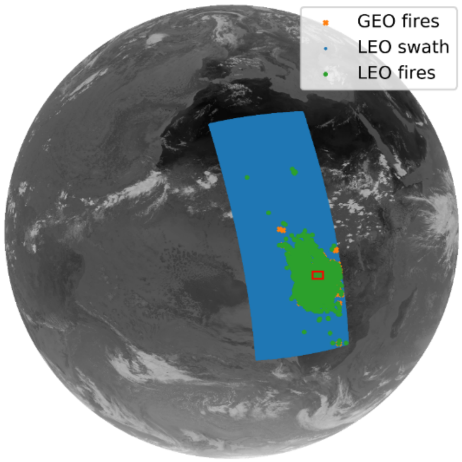

Figure 5: Full-disk view of selected repeat cycle for test case analysis. The analysed area is marked in red.

Following images show the comparison of the detections in the selected test area by different GEO and LEO products. The background image (Figure 6) is the inverted 3.9 µm SEVIRI channel, where hotspots appear in particularly dark colours.

It is recalled that the probability of detection (POD) describes the relationship between true positives (pixels marked in green) and false negatives (pixels marked in blue), while the precision (PRE) describes the relationship between true positives and false alarms (pixel marked in orange). Figure 7 presents the comparison of different SEVIRI products with all MODIS fires (minimum one fire per pixel from all fire detection confidences).

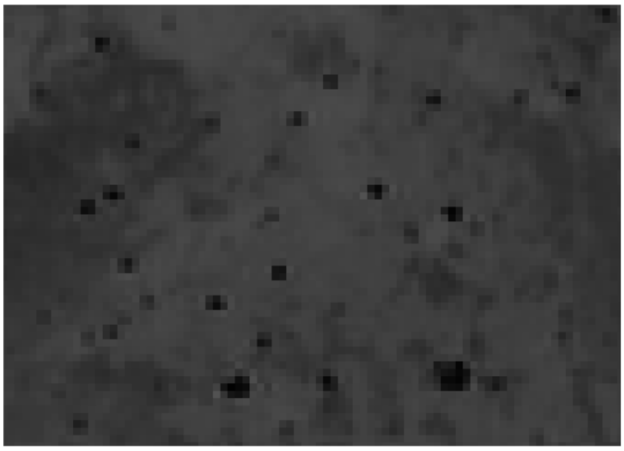

Figure 6: Background image of test area showing the inverted IR3.9 SEVIRI channel.

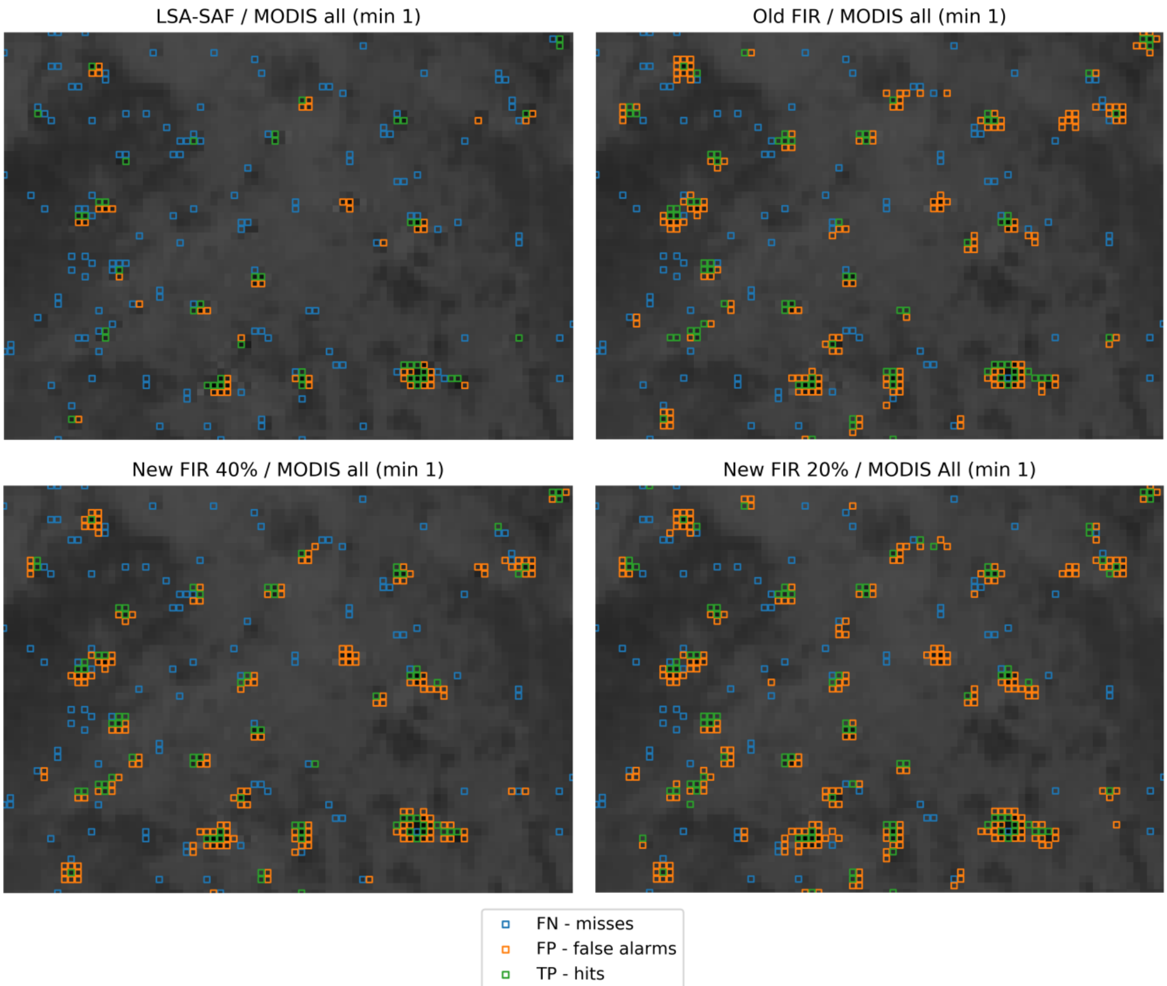

Figure 7: Test case: comparison of LSA-SAF, Old FIR, New FIR 40%, and New FIR 20% with all MODIS fires

It can be observed in the first image of Figure 7 how the LSA-SAF FD&M product generally contains a lower number of detections than the FIR products. This is also reflected in the absolute detection numbers reported in Table 1. As observed in Table 2 (p. 16), this leads to a low POD (high amount of misses compared to hits) but a high PRE (good amount of true positives among detections). The Old FIR comparison shows a higher number of false alarms (lower PRE), but also a higher number of detected fires (higher POD). The New FIR products with a FIRProb threshold of 40% shows a similar behaviour – in the absolute numbers, the PRE is however slightly higher and the POD slightly lower than the Old FIR product. As discussed when analysing Table 2 (p. 16), decreasing the FIRProb threshold includes more detections into the New FIR product; this can be observed in the fourth image with the threshold decreased to 20%, showing a higher number of false alarms but also of hits. For this MODIS-all comparison, the maximum in F1 performance was identified at a FIRProb threshold of 10%

In general, it can be observed how many of the detections in the FIR products are clustered in localised groups. The clusters have cores of true positive pixels, surrounded by false alarm detections. Many of the undetected MODIS fires are smaller, isolated fires appearing in locations where no structure can be identified in the underlying SEVIRI IR3.9 channel background.
Figure 8 shows the FIRProb values for each detection in the New FIR product inside the test area.

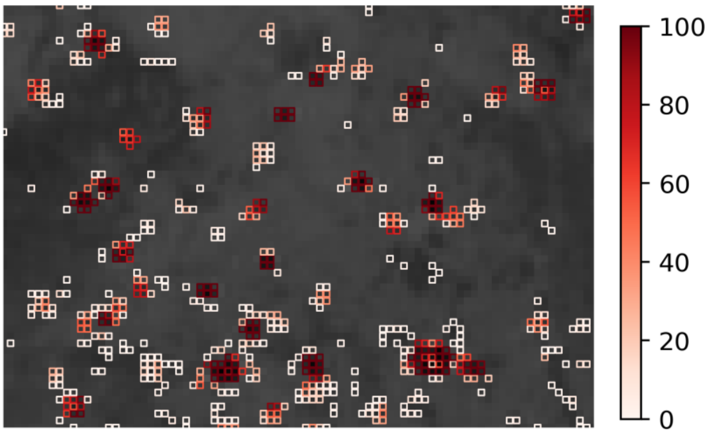

Figure 8: FIRProb values for detections in the test area from the New FIR product. Colour-scale in percentage unit.

The cluster structure can be observed again, with high-probability detections being often surrounded by low-probability detections.

Figure 9 presents the comparison of FIR products with all VIIRS fires (minimum one fire per pixel from all fire detection confidences) on the same test area and same repeat cycle.

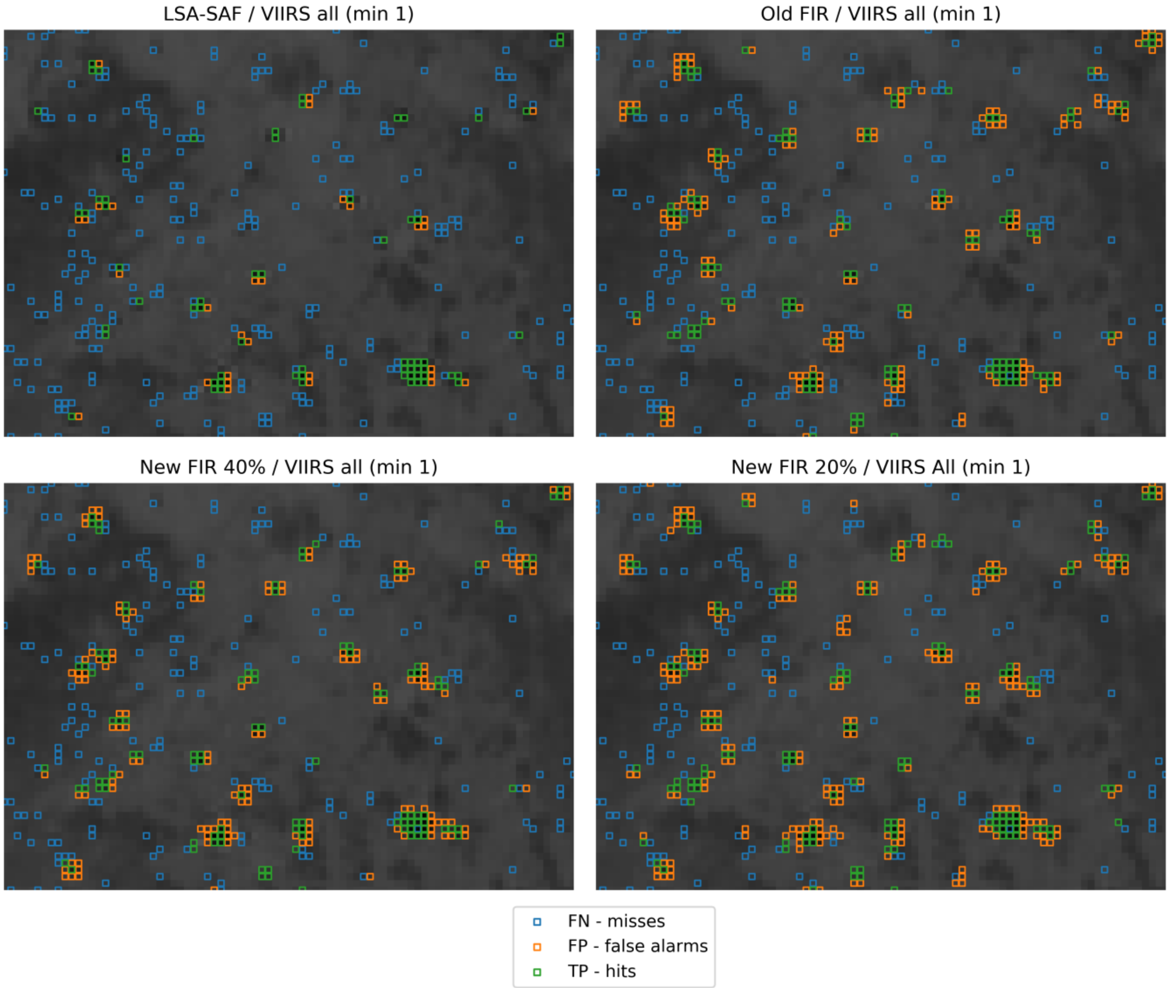

Figure 9: Test case: comparison of LSA-SAF, Old FIR, New FIR 40%, and New FIR 20% with all VIIRS fires

The observations regarding Figure 7 for the MODIS test case are valid also for this VIIRS comparison. The most evident difference is the much higher number of misses, composed by generally small fires that VIIRS is able to detect at its native resolution of 375 m but are invisible to the algorithms on SEVIRI data. A further difference that can be observed is a consistent number of pixels that were categorised as false alarms in the MODIS comparison, but appear as true positives in this VIIRS comparison. This was already observed in the analysis of the quantitative results for the VIIRS comparison in Table 3 (p. 17), where a higher PRE was reported for all SEVIRI products compared to the MODIS comparison.

5 CONCLUSIONS

5.1 Known Limitations of the Adopted Methodology

The intrinsic properties and quality of the available data, together with constraints given by computational resources and implementation complexity, cause limitations to the adopted methodology. In the following, the identified limitations are described:

- In the absence of ground-based active fire observation data, this works uses fire products from LEO instruments (MODIS and VIIRS) as the "ground-truth". This leverages the assumption that the higher ground resolution and the swath geometry (especially at low viewing zenith angles) allows a better detection performance for LEO rather than for GEO instruments. The quality of the validation comparison is therefore directly linked to the quality of the detection algorithms for the LEO data. Especially the quantitative results have to be understood as a relative comparison between different SEVIRI products against LEO products, rather than an analysis of the absolute values of measured metrics. The described validation methodology has therefore to be understood as an inter-comparison of fire products from satellite-based sensors.
- The timing between compared products is a further limitation. The current approach compares LEO swaths whose start time is inside the acquisition time period of a SEVIRI repeat cycle (15 min). This causes unavoidable time gaps between the LEO and GEO acquisition times of each ground pixel, introducing errors caused by the progression of fires and changing cloud cover. Since all compared SEVIRI products are derived from the same data, this does not affect the relative comparison between products.
- Due to the specific acquisition geometry of the MODIS and VIIRS instruments, a degradation of detection quality for higher satellite view angles is expected. As this affects all comparisons the same way, not influencing the relative comparison focus of this study, it was not investigated further.
- The MODIS and VIIRS algorithms are activated also for water pixels, while the SEVIRI algorithms filter water pixels using land-sea masks. This was not corrected in this study, as the fire contribution from water bodies in the study area is expected to be minor, and affects all comparisons in a systematic way not influencing the relative analysis.

5.2 Conclusions and Indications for Users

This work presented the validation activity performed on the updated SEVIRI active fire monitoring product (FIR). The updates were designed to improve the overall detection performance, with a special focus on coastline areas.

The main outcome of the analysis is that the updated algorithm performs generally better than the previous implementation. Specifically, the inclusion in the product of a new fire probability indicator gives users the important possibility of tuning the detection behaviour according to their applications and geographical areas.

Following indications for users were identified:

- In the updated product, the FIRProb threshold for classifying „possible" fires in the full-disk fire mask is set to 40%. Compared to the fire mask of the old product, this leads to a lower amount of false alarms, but also to a lower amount of correct detections. Users seeking results comparable to the old algorithm are advised to set the minimum FIRProb threshold to 30%.
- The "probable"-only fire mask of the new product has a slightly worse precision (higher false alarms ratio), but a higher probability of detection compared to the old product.

- The analysis has confirmed the importance and usefulness of tuning the FIRProb threshold according to specific use cases. Generally, a lower FIRProb is recommended for applications were a high probability of detection is preferred, at the cost of increased false alarm detections.
- In terms of F1 measure (see section 3.4 for the definition), the best performance has been observed for a FIRProb value of 30%, with good results in the range 10-40%. FIRProb values lower than 10% lead to a drop in performance due to the very high false alarm rate.
- The detection behaviour obtained using a specific FIRProb threshold is dependent on the geographical area of application due to the changing sensing geometry. The indications above have been obtained analysing detections focused over central Africa and may be fully valid only for this region. Users applying the FIR product over different areas are invited to tune the results accordingly using different FIRProb thresholds. Feedback about these activities would be appreciated.