EUMETSAT
H SAF
SUPPORT TO OPERATIONAL HYDROLOGY AND WATER MANAGEMENT

EUMETSAT Satellite Application Facility On Support to Operational Hydrology and Water Management (H SAF)

Product Validation Report (PVR) for product SE-D-SEVIRI (SE-D-SEVIRI (H34))

Reference Number:	SAF/HSAF/CDOP3/PVR-34
Issue/Revision Index:	Issue 1.1
Last Change:	30/06/2020

DOCUMENT SIGNATURE TABLE

	Name	Date	Signature
Prepared by :	H SAF Project Team	30/06/2020	
Approved by :	H SAF Project Manager	30/06/2020	

DOCUMENT CHANGE RECORD

Issue / Revision	Date	Description
1.0	30/04/2020	Version prepared for ORR
1.1	30/06/2020	Updated version which acknowledges RIDs dispositions

TABLE OF CONTENTS

1. Introduction: aim of this report

2. Validation methodology

3. Results

3.1 Validation for flat areas of Belgium

3.2 Validation for flat areas of Finland

3.3 Validation for mountainous areas of Italy

3.4 Validation for mountainous areas of Turkey

3.5 Validation using SENTINEL snow cover maps in Turkey

3.5.1 Case Study

3.6 Validation using Sentinel-2 satellite data for extra H-SAF areas

4. Conclusions: Product requirement compliance

Annex 1: Validation of H-SAF H34 Snow Product over Caucasus, Belarus, Mount Atlas, and Lebanon using Sentinel-2 Level-2 Scene Classifications

Annex 2: Validation methodology for H10 – Snow Detection

1. Introduction: aim of this report

This validation report is an Operational Readiness Review (ORR) of product H34 - Snow Detection.

H34 is an H-SAF snow product including daily snow-detection maps based on visible/near-infrared (VIS/NIR) data from the Spinning Enhanced Visible and Infrared Imager (SEVIRI) on board of Meteosat satellites. The algorithm of H34 largely builds on H-SAF products H10 and H31, which are both currently operational. Compared to H10, which provides binary map of snow/no-snow over the H-SAF area (25°N to 75°N latitude, 25°W to 45°E longitude), H34 is a second-generation product covering the full disk of SEVIRI. As such, H34 provides an opportunity to evaluate H-SAF snow products over larger areas than what is possible with H10.

In this report, product H34 is validated against in-situ observations from ground data networks of H-SAF member countries and high-resolution Satellite data from Copernicus Sentinel2 in extra H-SAF areas.

Ground data from Station networks for Snow Season October 1, 2018 till May 31, 2019 of four H-SAF members are taken into account: Finland, Turkey, Italy and Belgium.

In addition to this, validation with data from Finland for previous snow season 2017-18, and Sentinel2 validation for Turkey, compared with the ground data validation, are displayed.

For extra European (extra H-SAF) areas, validation is performed by CIMA using Sentinel2 data, specifically in the flat/forested areas of Belarus, and in the mountainous area of Caucasus, Lebanon and Atlas. This is the standard validation procedure for products H12 and H35 (Fractional Snow Cover) and should be used also for Snow Extension products validation, in areas outside Europe/H-SAF members (2) Piazzi *et al.*, 2019.

2. Validation methodology

The quantitative validation has been performed for Snow Season October 1, 2018 till May 31, 2019.

As H34 is the same product as H10, with the extended coverage to MSG Disk, for European areas the same validation methodology as H10 is applied, with ground data from observational networks of four selected H-SAF members: Finland, Turkey, Italy and Belgium.

The standard validation procedure and the same station network of the yearly Operational Review (OR) of H10 is performed. This methodology is described in Annex 2: Validation Methodology for H10 – Snow extension.

For extra-European areas (extra H-SAF areas), a different type validation, using high resolution Sentinel2 satellite data, is applied for selected areas in the northern hemisphere. The methodology is described briefly in Cap 3.5 and 3.6, and more extensively in Annex 2 and in *(2) Piazzi et al. (2019)*

The main scores to be evaluated for product compliance are the Probability Of Detection (POD) and the False Alarm Rate (FAR). However, in order to give a more complete idea of the product error structure, several other statistical scores have been evaluated and shown in this report.

The product requirements of product H34 are the same of product H10, and are displayed in Table 1 (requirements for flat/forested areas) and Table 2 (requirements for mountainous areas)Table .

Score	Threshold	Target	Optimal
POD	0.80	0.85	0.99
FAR	0.20	0.15	0.05

Table 1: product requirements for product SN-OBS-1 in Flat/Forest areas

Score	Threshold	Target	Optimal
POD	0.60	0.70	0.99
FAR	0.30	0.20	0.05

Table 2: product requirements for product SN-OBS-1 in Mountainous areas

3. Results

3.1 Validation for flat areas of Belgium

Validation of H34 is done with the same methodology as in H10. Validation is performed in the winter period of December to March 2018-19, even as Table 3 shows in Belgium there are only 2 months with good snow coverage, January and February.

Table 3: Validation results for 2018-2019 winter season in Belgium.

flat area	Dec 2018	Jan 2019	Feb 2019	Mar 2019	Total
hits	0	45	13	0	58
false alarms	4	33	0	0	37
misses	0	3	1	0	4
correct negatives	214	140	940	563	1857
numbers of obs	218	221	954	563	1956
pod	nd	0.94	0.93	nd	**0.94**
far	1.00	0.42	0.00	nd	**0.39**
csi	0.00	0.56	0.93	nd	0.59
pofd	0.02	0.19	0.00	0.00	0.02
acc	0.98	0.84	1.00	1.00	0.98
hss	0.00	0.61	0.96	nd	0.73

POD of 0.93 is very good and above target, close to optimal values. FAR of 0.39 is not as good and slightly below threshold: this is mainly due to the small number of observations. In fact, even in months without snow cover, product H34 retains a good Accuracy rate, of 0.9 and more.

3.2 Validation for flat areas of Finland

Validated months are as follows, 2017/12, 2018/01, 2018/02, 2018/03, 2018/04, 2018/05 and 2018/10, 2018/11, 2018/12, 2019/01, 2019/02, 2019/03, 2019/04, 2019/05, corresponding to partial winter season of 2017-2018 -with missing October and November- and full winter season of 2018-2019. Results of validations for these two winter seasons are shown in Table 4 and 5.

Table 4: Validation results for 2017-2018 winter season in Finland.

flat area (2017/18)	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Total
hits	-	-	0	9	703	1129	639	77	2557
false alarms	-	-	0	1	0	0	10	3	14
misses	-	-	0	0	1	15	75	35	126
corr negatives	-	-	0	0	0	0	151	1471	1622
numbers of obs	-	-	0	10	704	1144	875	1586	4319
pod	-	-	-1	1.000	0.999	0.987	0.895	0.688	**0.953**
far	-	-	-1	0.100	0.000	0.000	0.015	0.038	**0.005**
csi	-	-	-1	0.900	0.999	0.987	0.883	0.670	0.948
pofd	-	-	-1	1.000	-1	-1	0.062	0.002	0.009
acc	-	-	-1	0.900	0.999	0.987	0.903	0.976	0.968
hss	-	-	-1	0.000	0.000	0.000	0.720	0.790	0.932

Table 5: Validation results for 2018-2019 winter season in Finland

flat area (2018/19)	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Total
hits	63	0	0	30	959	1171	1144	50	3417
false alarms	284	5	0	0	0	13	57	4	363
misses	15	0	0	0	4	32	160	24	235
corr negatives	828	14	0	0	0	28	845	1382	3097
numbers of obs	1190	19	0	30	963	1244	2206	1460	7112
pod	0.807	-1	-1	1.000	0.995	0.973	0.877	0.675	**0.935**
far	0.818	1.000	-1	0.000	0.000	0.010	0.047	0.074	**0.096**
csi	0.174	0.000	-1	1.000	0.995	0.962	0.840	0.641	0.851
pofd	0.255	0.263	-1	-1	-1	0.317	0.063	0.002	0.104
acc	0.748	0.736	-1	1.000	0.995	0.963	0.901	0.980	0.915
hss	0.212	0.000	-1	-1	0.000	0.536	0.800	0.771	0.831

Both POD (0,95 and 0,94) and FAR (0,005 and 0,09) are very good and close to optimal values: product H34 has, as product H10, very good performances in flat and forested areas, especially in case of a homogenous and stable snow cover during the winter months, as in Nordic countries. Validation over Belarus with Sentinel2 data shows very similar and optimal results.

3.3 Validation for mountainous parts of Italy

Validation is performed in the winter period of December to February 2018-19, as in March there were too few observations.

Table 6: Validation results for 2018-2019 winter season in Italy.

mountaineous area	Dec 2018	Jan 2019	Feb 2019	Total
hits	495	184	490	1169
false alarms	23	11	26	60
misses	206	81	157	444
corrnegatives	329	146	350	825
numbers of obs	1053	422	1023	2498
pod	0.71	0.69	0.76	**0.72**
far	0.04	0.06	0.05	**0.05**
pofd	0.07	0.07	0.07	0.07
acc	0.78	0.78	0.82	0.80

POD of 0.72 is good and above threshold. FAR of 0.05 is very good and close to optimal, this is probably due to the selected cases, with a good number of cloud free observations. Product H34 retains a good Accuracy rate of 0.80 even in the demanding very complex topography of the Italian Alps.

3.4 Validation for mountainous parts of Turkey

MSG/SEVIRI product validation over mountainous areas of Turkey for the period of October 1. 2018 – May 31.2019 has been carried out using the snow depth measurements from meteorological stations. The distribution of all stations over DEM of Turkey is shown in **Errore. L'origine riferimento non è stata trovata.**. Snow depth measurement data have been retrieved from AWOS stations located within the validation area over Turkey. Validation of the H34 product is being carried out using the H34 Validation program in the MATLAB environment.

All the available station observations have been revised, corrected and filtered, in order to minimize noise generated from the observation data. For the period of the validation, 361 synoptic stations, 16 SPA stations and 133 AWOS stations with ground observations were available. 5913 measurements collected from 3 synoptic, 4 SPA and 94 AWOS stations have been utilized in the validation study for mountainous areas while 517 measurements collected from 16 synoptic and 2 AWOS stations have been utilized in the validation study for flat areas.

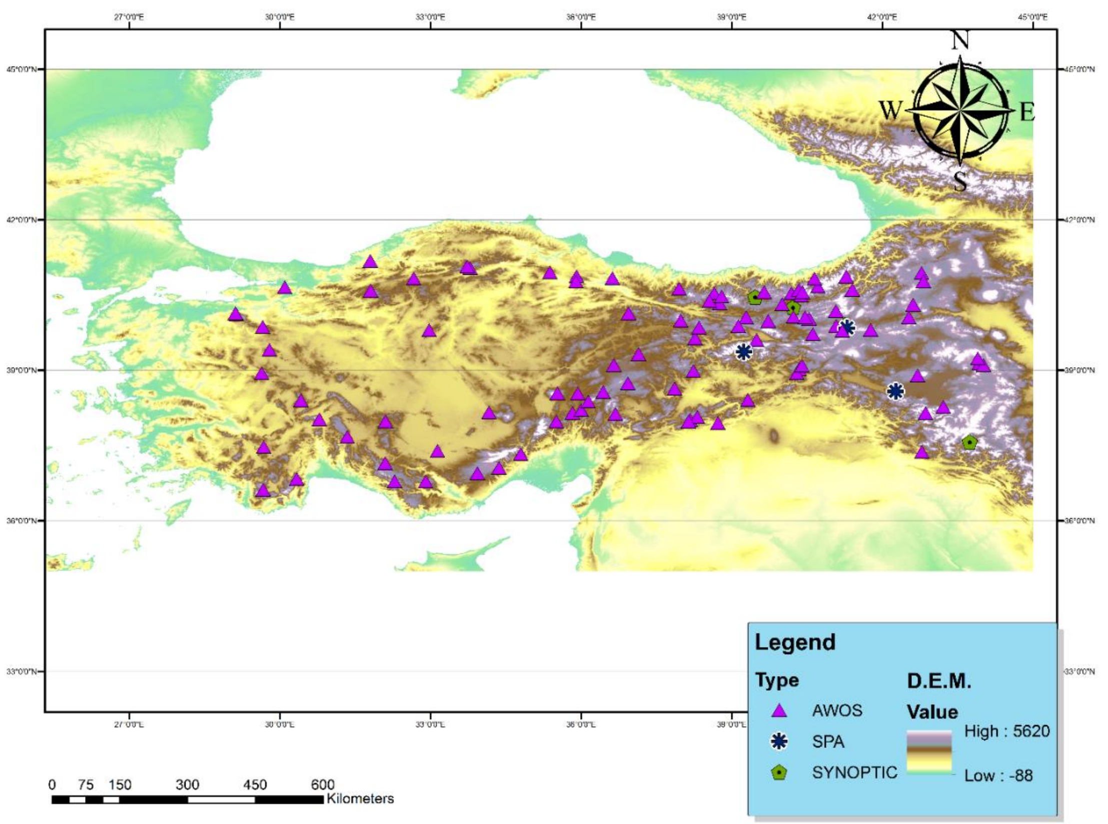


Figure 1. All the stations that are used in the validation after the quality control filtering has been applied (101 Stations)

The validation results for mountainous areas are presented in Table 7.

Table 7. H34 validation results by using ground observations on mountainous areas in Turkey.

mountainous	**Oct 2018**	**Nov 2018**	**Dec 2018**	**Jan 2019**	**Feb 2019**	**Mar 2019**	**Apr 2019**	**May 2019**	**Total**
hits	1	151	395	1706	1337	682	237	52	4561
false alarms	0	1	1	4	10	2	2	0	20
misses	20	274	237	61	329	516	395	509	2341
corr negatives	10	61	1	0	2	1	0	0	75
numbers of obs	31	487	634	1771	1678	1201	634	561	6997
pod	0.05	0.36	0.63	0.97	0.80	0.57	0.38	0.09	**0.66**
far	0.00	0.01	0.00	0.00	0.01	0.00	0.01	0.00	**0.00**
csi	0.05	0.35	0.62	0.96	0.80	0.57	0.37	0.09	0.66
pofd	0.00	0.02	0.50	1.00	0.83	0.67	1.00		0.21
acc	0.35	0.44	0.62	0.96	0.80	0.57	0.37	0.09	0.66
hss	0.03	0.12	0.00	0.00	0.00	0.00	-0.01	0.00	0.03

Validation cluster retains that in Turkey only data in mountainous areas is to be considered for validation of H34. In the mountainous areas, POD (0.66) and FAR (0.00) are both above the threshold, with FAR figures well at optimal levels.

In flat areas (not displayed), except for the month of January, data is not sufficient for good statics: in these cases, a high number of misses are observed.

3.5 Validation using SENTINEL-2 snow cover maps in Turkey

Validation of H34 product is performed by using higher spatial resolution Sentinel 2 data. In total. 106 Sentinel 2 images with cloud cover less than 30% taken between November 2018 and April 2019 are employed in the validation. The Sentinel 2 imagery were downloaded from the dedicated web page of USGS (https://earthexplorer.usgs.gov/). Location of the corresponding Sentinel 2 tiles are shown in Figure . As a first step, reference binary snow maps are obtained at 20 m spatial resolution by processing of the associated Sentinel 2 images via Sentinel 2's own scene classification processor Sen2Cor v2.8(1) ESA. 2018). In our previous study(2) Piazzi *et al.*. 2019), Sentinel 2 binary snow maps were tested against ground-based snow depth measurements and the results indicated that binary snow maps obtained by Sen2Cor processor were in good agreement with in-situ snow depth data with POD = 0.82, FAR = 0.08 and ACC = 0.79.

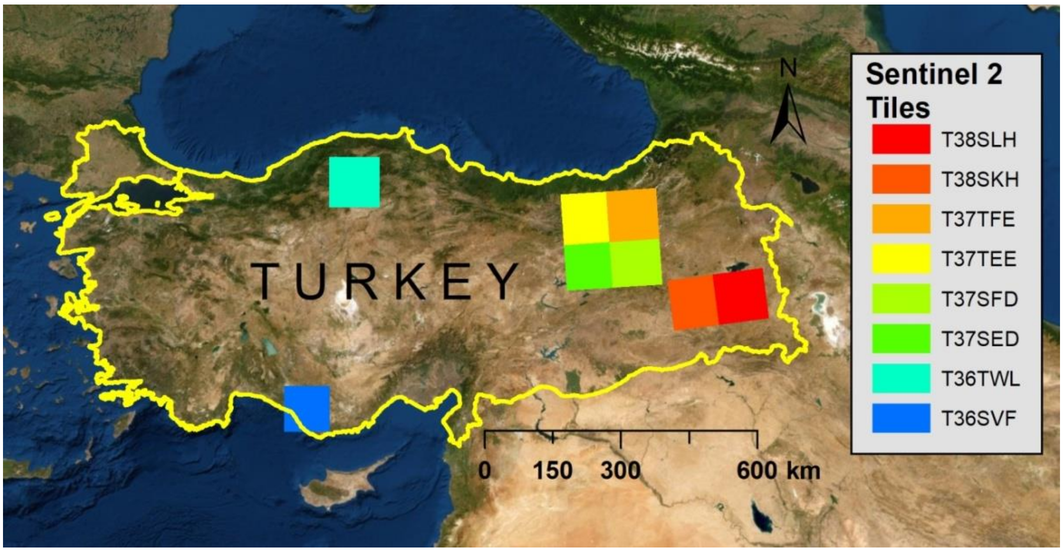


Figure 2. Locations of Sentinel 2 tiles over Turkey

All H34 images are projected to a common WGS84/UTM projected coordinate system in order to match with the projection of the corresponding Sentinel 2 tiles. Then, the mean of snow pixels that fall inside the exact footprint of an H34pixel is calculated, to find the fractional snow cover (FSC) value of that H34pixel. If the FSC of an H34 pixel is greater than 50% it is labeled as snow. The cloud threshold during the analysis is set to 50%, meaning that if the cloud cover and cloud shadow fraction calculated from the Sentinel 2 imagery in an H34 pixel exceed 50%, it is excluded from the validation. Additionally, H34 pixels including No-data, Dark area and saturated/defective Sentinel 2 pixels are excluded from further analysis. The results are given in Table 8.

Table 8. H34 validation results by using Sentinel data in Turkey area

mountainous	Nov 2018	Dec 2018	Jan 2019	Feb 2019	Mar 2019	Apr 2019	Total
hits	693	1416	3220	4040	2896	1072	13337
false alarms	207	196	162	393	312	335	1605
misses	322	250	67	228	486	267	1620
corr negatives	6393	1019	376	1393	3621	2546	15348
numbers of obs	7615	2881	3825	6054	7315	4220	31910
pod	0.683	0.850	0.980	0.947	0.856	0.801	**0.892**
far	0.230	0.122	0.048	0.089	0.097	0.238	**0.107**
csi	0.567	0.760	0.934	0.867	0.784	0.640	0.805
pofd	0.031	0.161	0.301	0.220	0.079	0.116	0.095
acc	0.931	0.845	0.940	0.897	0.891	0.857	0.899
hss	0.684	0.685	0.733	0.747	0.780	0.675	0.797

Results obtained by using Sentinel 2 data are better than the ones obtained by using ground data in terms of POD (0.89), and worse in terms of FAR (0.11). Both figures are above the thresholds and above target levels.

3.5.1 Case Study: Influence of topography in validation by using SENTINEL for two different areas

Ground observations have the limitation of pixel area representation. Tile-based POD and FAR graphs are given in Figures 3 and 4, respectively.

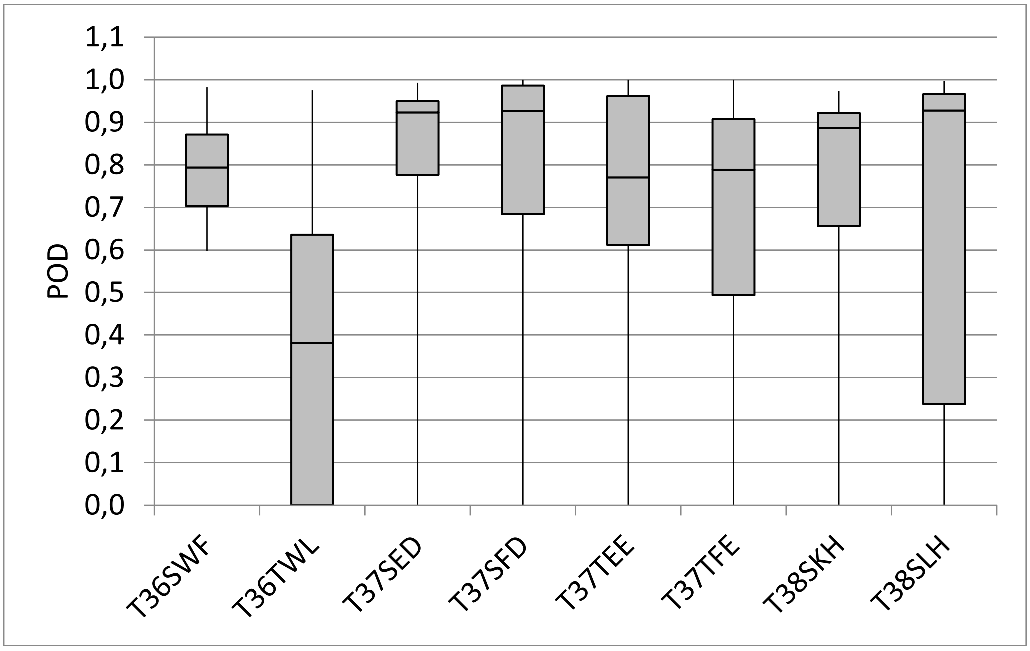


Figure 3. Tile-based POD values.

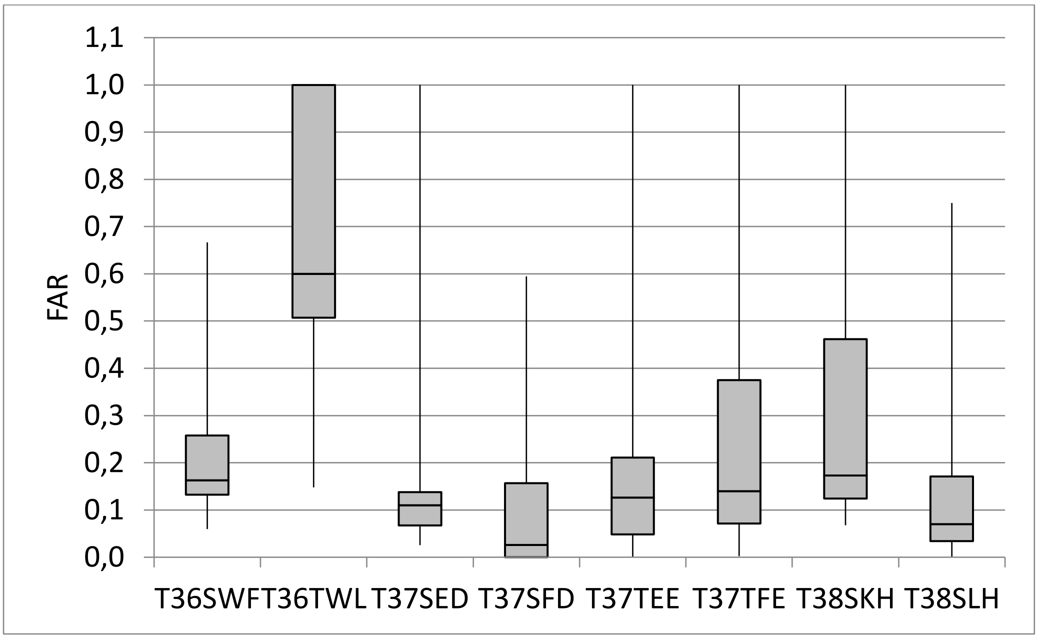


Figure 4. Tile-based FAR values.

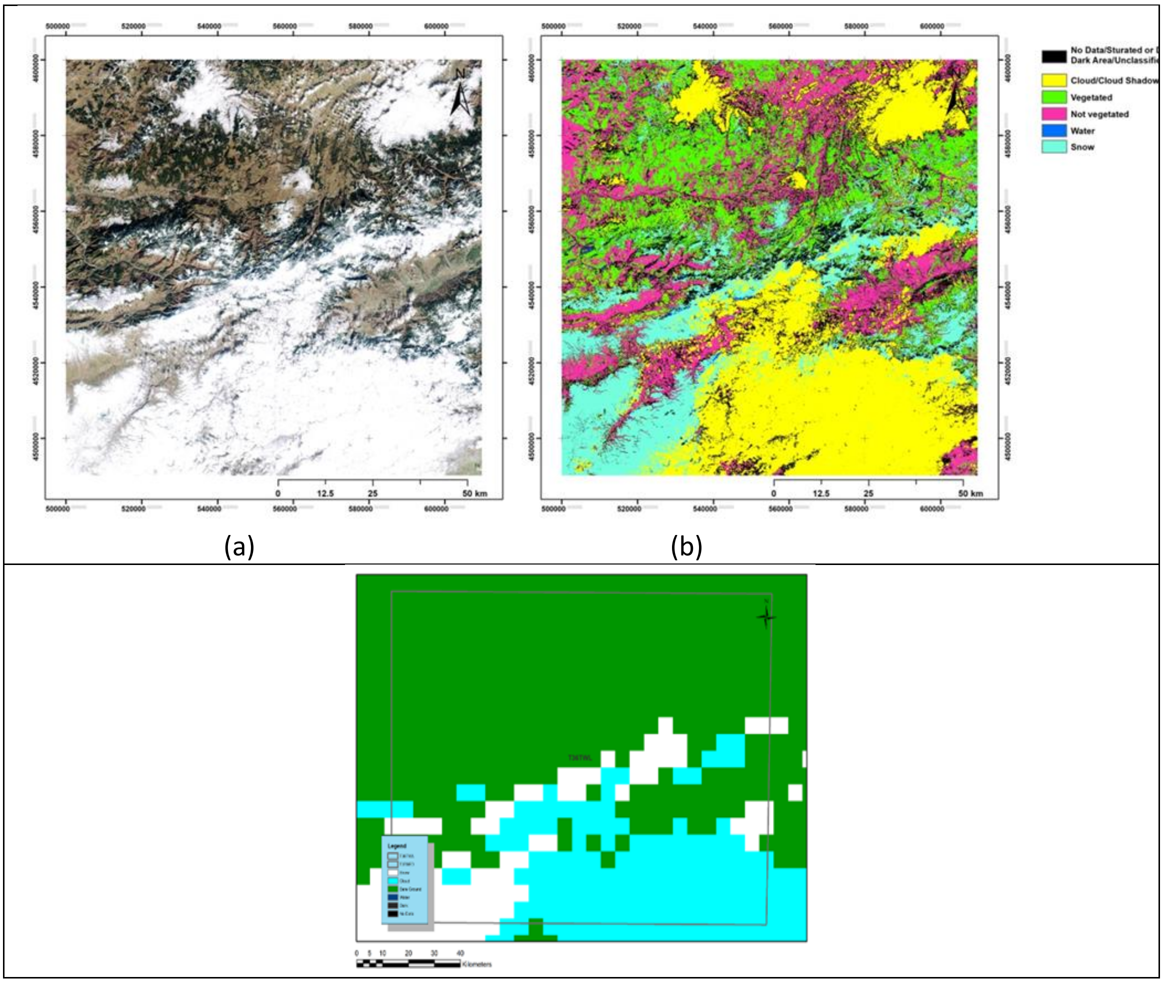
As shown in fig. 3 and 4, in different areas of Turkey product H34 performs differently: especially tiles T36TWL and T37SED show big differences in POD and FAR. This is mainly due to differences in topography as the two cases below enlighten.

In Figures 5 and 6 a comparison of the two areas is shown.

Fig. 5 shows Real color RGB composite, Sen2Cor scene classification images and H34 products for the date 11 Feb 2019 of T36TWL tile, Fig. 6 the same images for28 Jan 2019 in T37SED tile.

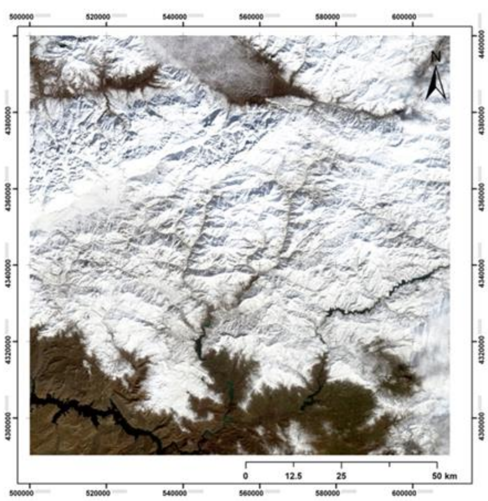
The low POD and high FAR values obtained for the tile T36TWL (Fig. 5) is due to low spatial resolution of H34 product, compared to very high resolution of Sentinel2, which can resolve the steep valleys without snow cover, in a very complex topography. When terrain and snow cover are more homogeneous, as in tile T37SED (Fig. 6), product H34 has better performances for snow detection.

Validation with Sentinel2 products in Middle Eastern and north-African mountainous areas, where snow cover is even less homogeneous and climate very dry, does not allow either for good scores, especially in terms of FAR.

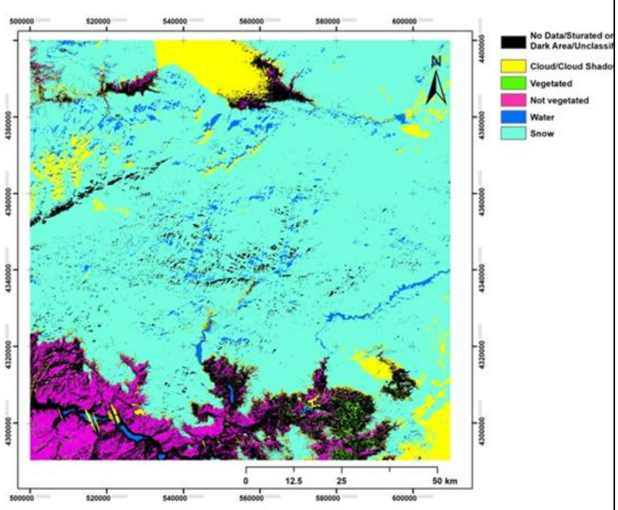


(c)

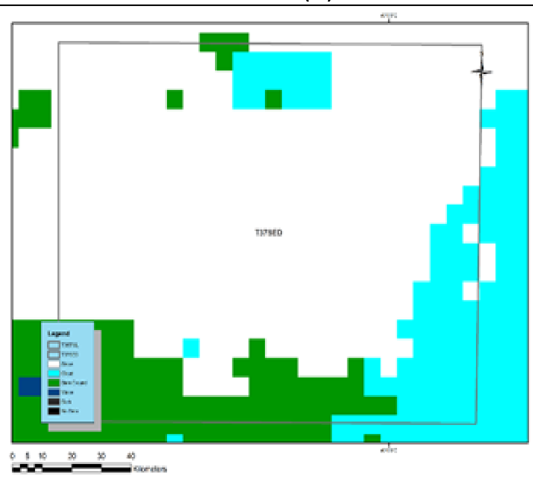
Figure 5.11 Feb 2019 (a) T36TWL real color RGB and (b) Sen2Cor scene classification and (c) H34 images.



(a)



(b)



(c)

Figure 6. 28 Jan 2019 (a) T37SED real color RGB. (b) Sen2Cor scene classification and (c) H34 product images.

3.6 Validation using SENTINEL-2 satellite data for extra H-SAF areas

This section is a summary of the report "Validation of H-SAF H34 Snow Product over Caucasus, Belarus, Mount Atlas, and Lebanon using Sentinel-2 Level-2 Scene Classifications (Oct. 2018 - May 2019)", written by Francesco Avanzi and Simone Gabellani, of CIMA research centre.
The original Report is in Annex 1, here is a brief summary of the methods and the results of the validation.

Evaluation of H-SAF snow products has been usually performed using ground-based snow data, such as snow-depth sensors.
In this report, we followed the validation methodology for H10 that was applied in Piazzi et al. (2019) and we compared SCA maps derived from H34 vs. SCA maps derived from Sentinel-2. The validation period was the 2018-2019 snow season**;** areas of interest were the Caucasus mountain range, Belarus, Mount Atlas in Morocco, and Mount Lebanon. These areas of interest include a mixture of predominantly flat areas (Belarus) and predominantly mountainous areas (Caucasus, Mount Atlas, and Mount Lebanon). Climates are also markedly different, with Belarus, Caucasus, and both Mount Atlas and Mount Lebanon falling in the Boreal, Boreal-to-Polar, and Warm-Temperature-to-Arid regions according to the Köppen- Geiger climate map (Kottek et al., 2006).

The validation follows the methodology discussed in Piazzi et al. (2019), which we briefly summarise below. In this regard, it should be noted that Belarus and the Caucasus region are the most significant testing regions here and should be considered as the two reference areas of interest for flat and mountain performance of H34. This is because we considered a significantly smaller number of tiles in Lebanon and the Atlas region; the arid climate of these regions also makes snow-cover patterns highly variable in space and time, an inherent challenge for a relatively large-scale product like H34.

The first step was the selection of the Sentinel-2 tiles over each area of interest to be compared with H34. This selection was performed by seeking a trade-off between extensively assessing H34 performance and limiting computational times. As a result, tile selection in Caucasus and the Atlas region focused on the most mountainous areas, while in Belarus we limited ourselves to tiles that were fully included in the country and reasonably far from the Baltic Sea, which may reduce snow-cover duration due to mitigation effects. In Lebanon, we selected one tile covering a central portion of the local mountain range and not including any portion of the Mediterranean Sea. Overall, we considered 10 tiles in Belarus, 10 tiles in the Caucasus region, 3 tiles in the Atlas region, and 1 tile in Lebanon (Figure 7).

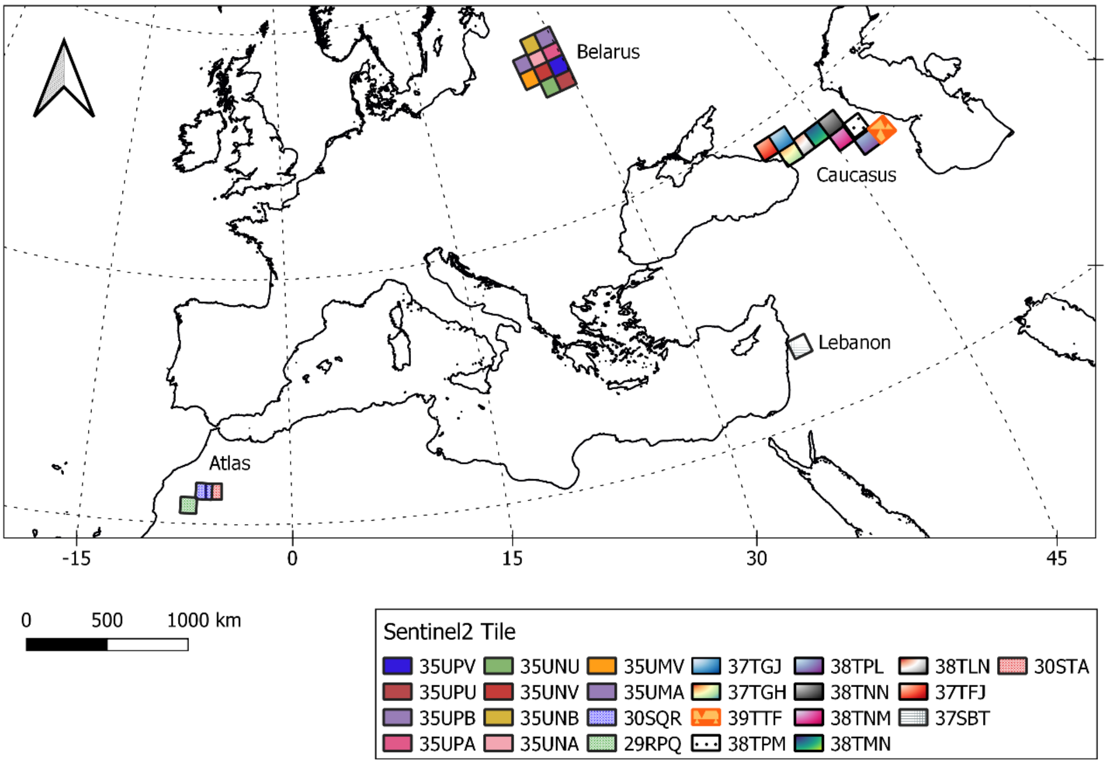


Figure 7: Location of validation regions with selected tiles

Then we performed an image-to-image comparison between Sentinel 2 and H34 with the goal of assessing H34 performance in estimating Sentinel-2 SCA. This assessment was limited to Sentinel images with overall cloud cover below 20%. The impact of clouds was particularly high in Belarus and the Caucasus region, where the number of usable images per tile was sometimes less than 5. Cloud obstruction was much less significant in Lebanon and the Atlas region, in agreement with a drier climate.

Agreement between H34 and Sentinel-based SCA maps was assessed using multiple performance metrics, as usual with Snow Extension products.

Between target and optimal	Between threshold and target	Threshold exceeded by < 50 %	Threshold exceeded by ≥ 50 %

Table 10: Accuracy metrics over the four areas of interest, aggregated across all tiles and images available.

Validation area	Caucasus	Belarus	Lebanon	Atlas
hits	4132	773	729	725
false alarms	747	0	1040	14914
misses	1355	13	28	269
corr negatives	6068	8682	6670	37657
numbers of obs	12302	9468	8467	53565
pod	0.75	0.98	0.96	0.73
far	0.15	0.00	0.59	0.95
csi	0.66	0.98	0.41	0.04
pofd	0.11	0.00	0.13	0.28
acc	0.83	0.99	0.87	0.72
hss	0.65	0.99	0.52	0.05

Accuracy metrics in Table 10 confirm the generally good performance of H34 in detecting snow conditions over the Caucasus range (POD = 0.75), Belarus (POD = 0.98), Lebanon (POD = 0.96), and the Atlas mountains (POD = 0.73). Indeed, all PODs are above the threshold (0,60 in mountainous areas and 0,80 in flat areas) and even above target in all areas, and close to optimal in 2 out of 4 areas.
These results are in line with those in Piazzi et al. (2019), with Belarus exceeding the already high scores reported in Finland. We also noted that PODs tend to decrease in tiles that are likely more vegetated, which is again in agreement with previous validations

The FAR (ratio of false alarms) is expectedly low in Caucasus and Belarus: in these regions, FARs are below threshold and even target (20% and 15%, respectively). FARs are higher in the Atlas area and in Lebanon, where a significantly smaller number of tiles were considered; in this areas FAR does not reach thresholds, this could be target of future research.
Overall, accuracy scores (ACC) are systematically higher than 70% in all areas: thus, fractional-snow-cover statistics by Sentinel 2 and snow-covered areas by H34 are spatially consistent.

In sum, results show a good agreement between H34 and Sentinel 2, especially in flat areas like Belarus, where snow-cover distribution is spatially more homogeneous. In mountain areas like Caucasus, scores are lower but still very good, small-scale snow-topographic interactions are more important in these landscapes and the scale of these interactions are sometimes smaller than the resolution of H34.

For Atlas and Lebanon region, POD are good, whereas FAR are high and doesn't meet thresholds; future users of H34 in very dry areas with mountains and therefore a inhomogeneous snow cover should be aware of this. Validation group retains that other areas in the hemisphere covered by H34 product (with a good seasonal snow cover) may be tested in the future, comparing Sentinel data and also ground data, if available, to better assess the product's performance.

4. Conclusions: Product requirement compliance

The main scores to be evaluated are the Probability Of Detection (POD) and the False Alarm Rate (FAR). However, in order to give a more complete idea of the product error structure, also CSI and ACC indices are shown.

Averaging the figures over all the countries and terrains, with ground station data validation over H-SAF areas, a good capability of the product to detect snow on the ground with a total POD 0.85, and an associated good low false alarm FAR 0.11, is observed (Table 11); this value lies above the thresholds of even the most demanding Flat/Forested areas requirements.

As expected problems arise in the complex orography of mountainous areas (see POD in Turkey and Italy), and in terms of FAR in the flat areas of Belgium, where probably snow cover is not stable or homogeneous. This may be due to the validation procedure with ground data and point observation, compared to observations of H34 satellite products.

Between target and optimal	Between threshold and target	Threshold exceeded by < 50 %	Threshold exceeded by ≥ 50 %

Table 11: Statistical scores for H34 over mountainous and flat areas with ground station data relatively period 1.10.2018-31.5.2019 (except Finland)

validation area	Belgium	Finland (2017-18)	Finland	Italy	Turkey	Average
pod	0.94	0.95	0.94	0.72	0.66	**0.85**
far	0.39	0.01	0.09	0.05	0.00	**0.11**
csi	0.59	0.95	0.86	0.07	0.66	0.66
acc	0.98	0.97	0.92	0.80	0.66	0.87

Table 12: Simplified compliance analysis for product H34 (Flat/Forest areas)

H-SAF Accuracy requirements for H34 in Flat/Forest areas				
Product requirements				H10
Score	threshold	target	optimal	total
POD	0.80	0.85	0.99	**0.94**
FAR	0.20	0.15	0.05	**0.16**

Table 13: Simplified compliance analysis for product H34 (mountainous areas).

H-SAF Accuracy requirements for H34 in Mountainous areas				
Product requirements				H10
Score	threshold	target	optimal	total
POD	0.60	0.70	0.99	**0.69**
FAR	0.30	0.20	0.05	**0.03**

In Tables 12 and 13, the results are separated for flat and mountainous areas, as thresholds are different. Flat areas are average of Belgium and Finland, mountainous areas of Italy and Turkey.

In average, on H-SAF areas and with standard ground data validation, product H34 is above thresholds for all terrains, and mostly above or near target levels.

Sentinel 2 high-resolution satellite data validation was performed by CIMA and Turkey, and overall results are displayed in Table 14.

Table 14: Statistical scores for H34 over mountainous and flat areas with Sentinel2 data relatively period 1.10.2018-31.5.2019

Validation area	**Caucasus**	**Belarus**	**Turkey**	**Lebanon**	**Atlas**
pod	0.75	0.98	0.89	0.96	0.73
far	0.15	0.00	0.11	0.59	0.95
csi	0.66	0.98	0.81	0.41	0.04
acc	0.83	0.99	0.90	0.87	0.72

In all areas of validation, POD is above threshold and above target. FAR is above threshold and target in the three main target areas, whereas in the dry zones of Lebanon and Atlas mountains it is below threshold.

In conclusion, product H34 satisfies requirements in almost all cases, even in the most complex and demanding areas. Further studies in dry mountainous environments, in near desert areas and in African areas, may be useful to investigate better product behaviour and performances.

References

1) ESA. (2018). Sen2Cor Configuration and User Manual v2.8. from http://step.esa.int/thirdparties/sen2cor/2.8.0/docs/S2-PDGS-MPC-L2A-SUM-V2.8.pdf

2) Piazzi. G.. Tanis. C. M..Kuter. S.. Simsek. B.. Puca. S..Toniazzo. A.. Takala. M.. Akyürek.Z.. Gabellani. S. & Arslan. A. N.. "Cross-Country Assessment of H-SAF Snow Products by Sentinel-2 Imagery Validated against In-Situ Observations and Webcam Photography".*Geosciences*.9(3). pp. 129. (2019).

ANNEX 1

Dipartimento della Protezione Civile

Università degli Studi di Genova

RESEARCH FOUNDATION cima

Regione Liguria

Provincia di Savona

Francesco Avanzi & Simone Gabellani

Validation of H-SAF H34 Snow Product over Caucasus, Belarus, Mount Atlas, and Lebanon using Sentinel-2 Level-2 Scene Classifications (Oct. 2018 - May 2019)

Savona, April 22, 2020

Contents

Contents 20

1 Context: aim of this report 21

2 Methods 22

3 Results 26

3.1 Confusion matrices 26
3.2 Accuracy metrics 31
3.3 H34 SCA vs. Sentinel-2 FSC and FGC 34

4 Conclusions 35

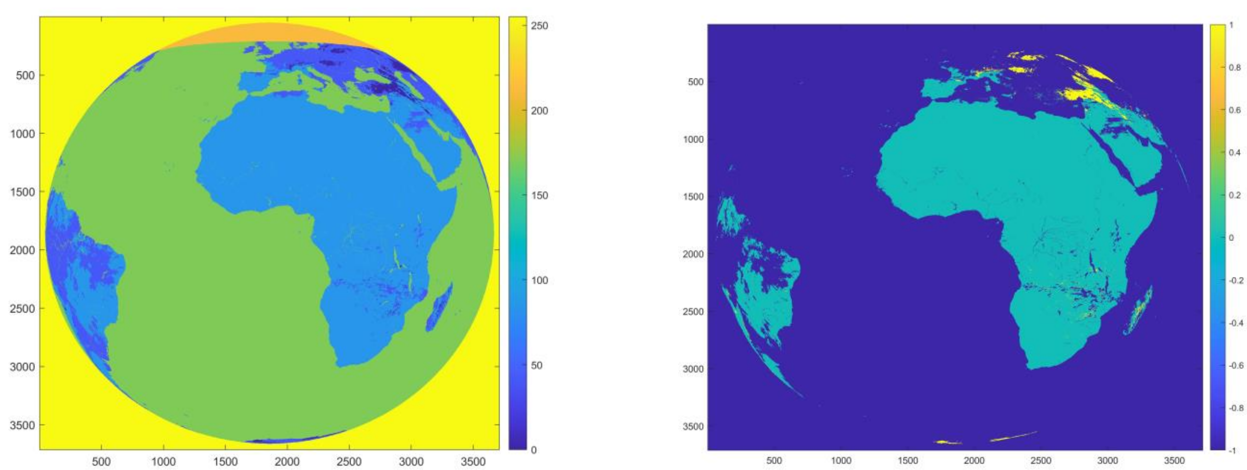
Bibliography 44

A Appendix 36

Context: aim of this report

This report discusses validation results for the H-SAF snow product H34 vs. Sentinel-2 Level-2 snow-covered-area maps (SCA). H34 is an H-SAF snow product including daily snow-detection maps based on visible/near-infrared (VIS/NIR) data from the Spinning Enhanced Visible and Infrared Imager (SEVIRI) on board of Meteosat satellites. The algorithm of H34 largely builds on H-SAF products H10 and H31, which are both currently operational after successful validation (see e.g. Piazzi et al., 2019). Compared to H10, which provides binary map of snow/no-snow over the H-SAF area (25°N to 75°N latitude, 25°W to 45°E longitude), H34 is a second-generation product covering the full disk of SEVIRI (see Figure 1). As such, H34 provides an opportunity to evaluate H-SAF snow products over larger areas than what is possible with H10.

Evaluation of H-SAF snow products has been usually performed using ground-based snow data, such as snow-depth sensors, or Sentinel-2 Scene-Classification maps (Piazzi et al., 2019). In this report, we followed the validation methodology for H10 that was applied in Piazzi et al. (2019) and we compared SCA maps derived from H34 vs. SCA maps derived from Sentinel-2. The validation period was the **2019-2019 snow season** and areas of interest were the Caucasus mountain range, Belarus, Mount Atlas in Morocco, and Mount Lebanon. These areas of interest include a mixture of predominantly flat areas (Belarus) and predominantly mountainous areas (Caucasus, Mount Atlas, and Mount Lebanon). Climates are also markedly different, with Belarus, Caucasus, and both Mount Atlas and Mount Lebanon falling in the Boreal, Boreal-to-Polar, and Warm-Temperature-to-Arid regions according to the KöppenGeiger climate map (Kottek et al., 2006).



(a) (b)

Figure 1: *Example of H-SAF H34 original classification map for January 10 2019 (Figure 1(a)) and snow-covered-area map for the same day derived during the validation process (Figure 1(b)), where snow is in yellow.*

Methods

Our validation approach followed the methodology discussed in Piazzi et al. (2019), which we briefly summarized below. In this regard, it should be noted that Belarus and the Caucasus region are the most significant testing regions here and should be considered as the two reference areas of interest for flat and mountain performance of H34. This is because we considered a significantly smaller number of tiles in Lebanon and the Atlas region; the arid climate of these regions also makes snow-cover patterns highly variable in space and time, an inherent challenge for a relatively large-scale product like H34.

The first step was the selection of the Sentinel-2 tiles over each area of interest to be compared with H34. This selection was performed by seeking a tradeoff between extensively assessing H34 performance and limiting computational times. As a result, tile selection in Caucasus and the Atlas region focused on the most mountainous areas, while in Belarus we limited ourselves to tiles that were fully included in the country and reasonably far from the Baltic Sea, which may reduce snow-cover duration due to mitigation effects. In Lebanon, we selected one tile covering a central portion of the local mountain range and not including any portion of the Mediterranean Sea. Overall, we considered 10 tiles in Belarus, 10 tiles in the Caucasus region, 3 tiles in the Atlas region, and 1 tile in Lebanon. For each tile, we then selected an orbit providing a full-square map by visually inspecting representative images on https://scihub.copernicus.eu/dhus/#/home. Figure 2 reports the location of each tile, while Table 1 summarizes how many maps for each tile we considered (after filtering images with cloud cover above 20%) and the corresponding orbit.

After tile selection, we downloaded from https://scihub.copernicus.eu/dhus/#/home all available S2MSI2A Sentinel-2 Level2A images between October 1 2018 and June 1 2019, including both Sentinel 2A and 2B. For each map, we then extracted Scene Classification maps (SCL) at 20 m (note that these Scene Classifications are directly available from https: //scihub.copernicus.eu/dhus/#/home and so no processing with Sen2Cor was performed). Scene-Classification maps provide indication of predominant land/cloud cover for each pixel in the image, including ground (ID 5), vegetation (ID 4), snow (ID 11), water (6), and various IDs for cloud and thin-cirrus cover[1]. These Scene-Classification maps represent the validation source against which to assess the performance of H34.

After downloading from the *meteoam.it* ftp all available H34 images between October 1 2018 and June 1 2019, we performed an image-to-image comparison between Sentinel 2 and H34 with the goal of assessing H34 performance in estimating Sentinel-2 SCA. This assessment was limited to Sentinel images with overall cloud cover below 20%, as already done by Piazzi et al. (2019). The impact of clouds is particularly high in Belarus and the Caucasus region, where the number of usable images per tile was sometimes less than 5 (see Table 1). Cloud obstruction was much less significant in Lebanon and the Atlas region, in agreement with a drier climate.

[1] see details at https://earth.esa.int/web/sentinel/technical-guides/sentinel-2-msi/level-2a/algorithm

The comparison procedure worked as follows (see also Piazzi et al., 2019):

1. We firstly determined the H34 scene corresponding to each Sentinel-2 tile-map by clipping the H34 full-disk raster according to the bounding box of each Sentinel-2 image;
2. We then estimated a fractional-snow-cover (FSC) and fractional-ground-cover (FGC) map from Sentinel-2 at the coarser resolution of H34 by counting the number of Sentinel's snow and ground pixels in each coarser pixel of H34. Any pixel where the percentage of non-snow and non-ground pixels was above 50% was set to NaN and was not considered in this evaluation (following again Piazzi et al., 2019);
3. We finally derived a Sentinel-based snow-covered-area map by assuming any pixel with FSC >50% as snow-covered, and as ground-covered otherwise. This SCA map was the final comparison source for H34.

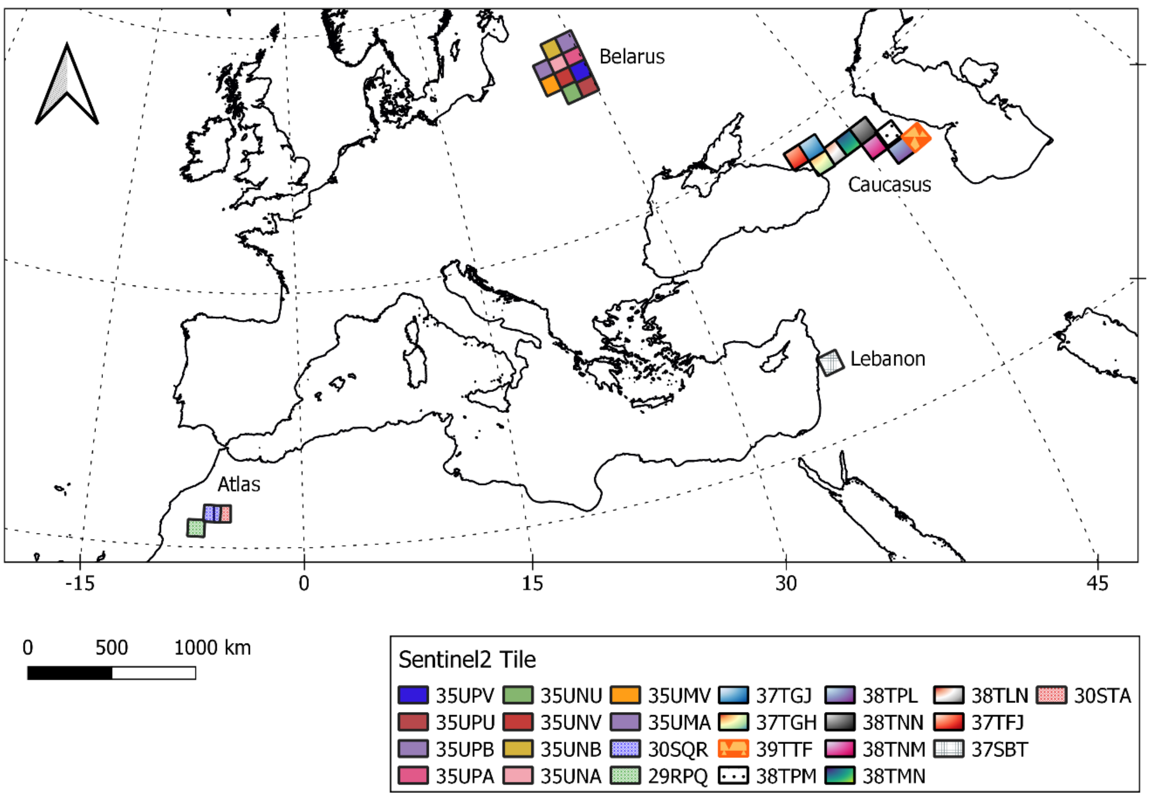


Figure 2: *Considered Sentinel-2 tiles.*

Agreement between H34 and Sentinel-based SCA maps was assessed using multiple performance metrics. The first one is the confusion matrix, which quantifies the frequency of agreements and disagreements between the two products. Agreements can result from both products assigning a given pixel to snow (hits, n_{11}) or to bare ground (correct negatives, n_{00}). On the other hand, disagreement may results from Sentinel and H34 assigning a pixel to snow and ground, respectively (misses, n_{10}), or Sentinel and H34 assigning a pixel to ground and snow, respectively (false alarms, n_{01}). Confusion matrices are generally summarized as follows:

$$\begin{bmatrix} n_{11} & n_{01} \\ n_{10} & n_{00} \end{bmatrix}$$

Table 1: *Considered Sentinel-2 tiles, satellite orbit, and number of images for each area of interest. Images cover the period October 1 2018 through June 1 2019 and were filtered to discard those with overall cloud cover above 20% (Piazzi et al., 2019).*

Area of interest	S2 Tile	Orbit	Number of images
Caucasus	37TFJ	78	8
	37TGJ	78	6
	37TGH	78	8
	38TLN	35	5
	38TMN	35	11
	38TNN	135	3
	38TNM	135	6
	38TPM	135	2
	38TPL	92	4
	38TTF	92	4
Belarus	35UMA	136	5
	35UMV	136	5
	35UNB	93	7
	35UNA	93	8
	35UNV	93	10
	35UNU	93	9
	35UPB	93	5
	35UPA	93	8

	35UPV	93	7
	35UPU	50	4
Lebanon	37SBT	121	13
Atlas	29RPQ	137	28
	29SQR	94	18
	30STA	94	17
Total			**201**

Frequencies of the contingency matrix were also used to compute a number of other metrics:

$$POD = \frac{n_{11}}{n_{11} + n_{10}} \tag{1}$$

$$FAR = \frac{n_{01}}{n_{11} + n_{01}} \tag{2}$$

$$POFD = \frac{n_{01}}{n_{01} + n_{00}} \tag{3}$$

$$ACC = \frac{n_{11} + n_{00}}{n_{11} + n_{00} + n_{10} + n_{01}} \tag{4}$$

$$CSI = \frac{n_{11}}{n_{11} + n_{10} + +n_{01}} \tag{5}$$

$$HSS = \frac{2 \times (n_{11} \times n_{00} - n_{01} \times n_{10})}{[(n_{11} + n_{10}) \times (n_{10} + n_{00}) + (n_{11} + n_{01}) \times (n_{01} + n_{00})]} \tag{6}$$

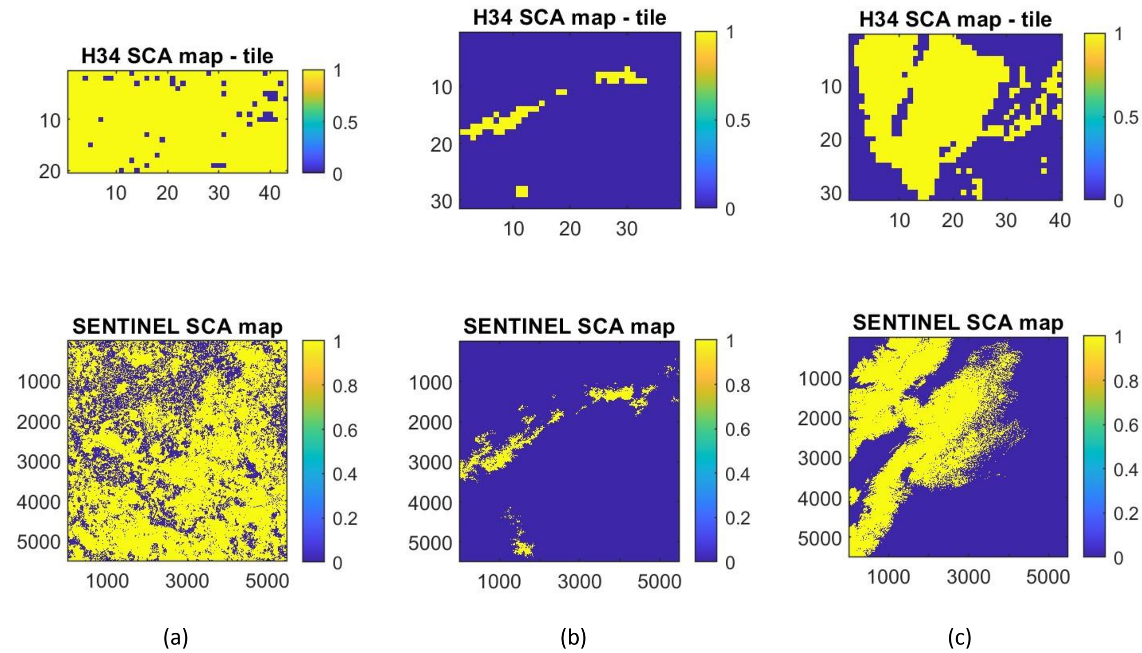


Figure 3: *Comparison between H34 and Sentinel-2 snow-covered area in Belarus (February 18, 2019, tile 35UPB), Mount Atlas (November 28, 2018, tile 29RPQ), and Lebanon (January 21, 2019, tile 37SBT).*

Following the methodology applied in Piazzi et al. (2019), we assumed as threshold, target, and optimal PODs values of 70%, 80%, and 90%, respectively. In terms of FAR, we assumed as threshold, target, and optimal scores values of 20%, 10%, and 5%, respectively. All these metrics were computed by aggregating results of all tiles belonging to the same area of interest (Belarus, Caucasus, Atlas, and Lebanon). Additionally, we also computed mean monthly performance scores and mean POD and FAR by tile.

Results

Confusion matrices

Figure 3 shows three examples of H34 vs. Sentinel-2 SCA for Belarus, Mount Atlas, and Lebanon. The overall spatial pattern of SCA is generally well reproduced in these cloud-free images, despite this pattern being significantly different across the three landscapes. In the boreal winter of this flat region of Belarus, for example, snow cover is homogeneous, with only sporadic cloud cover in Figure 3(a). Mount Atlas (Figure 3(b)) and Lebanon (Figure 3(c)) are both mountainous regions, but the latter presents a much larger snow-cover extent than the former. H34 successfully captures this difference in snow-cover extent, even though small-scale transitions between snow-covered peaks and snow-free valleys are only partially reconstructed by H34 due to its natively large-scale resolution compared to Sentinel 2 (see for example Figure 3(c)).

Confusion matrices for the four areas of interest under study read as follows:

$$\begin{bmatrix} 4,132\ (34\%) & 747\ (6\%) \\ 1,355\ (11\%) & 6,068\ (49\%) \end{bmatrix}_{Caucasus}$$

$$\begin{bmatrix} 773\ (8\%) & 0\ (0\%) \\ 13\ (0\%) & 8,682\ (92\%) \end{bmatrix}_{Belarus}$$

$$\begin{bmatrix} 725\ (1\%) & 14,914\ (28\%) \\ 269\ (1\%) & 37,657\ (70\%) \end{bmatrix}_{Atlas}$$

$$\begin{bmatrix} 729\ (9\%) & 1,040\ (12\%) \\ 28\ (0\%) & 6,670\ (79\%) \end{bmatrix}_{Lebanon}$$

Correct negatives (n_{00}) were generally high in all areas of interest, with peaks in Belarus (92%) and Lebanon (79%). In Caucasus, which was the region with the most persistent snow cover in our dataset, correct negatives were obviously smaller than elsewhere (49%); however, Caucasus was also the region with the highest rate of correct hits (n_{11}, 34%), which is consistent with this being the snowiest areas among the four we considered here.

False alarms (n_{01}) were minimal in Caucasus and Belarus (6% and 0%, respectively), but they increased in Lebanon (12%) and the Atlas area (28%). Misses (n_{10}) were a small number as well, with an increase in snowier regions (e.g., Caucasus). We interpret this higher rate of false alarms and misses in Lebanon and Atlas as possibly due to snow-cover in these regions being more ephemeral than in Caucasus due to lower latitudes (see Sturm et al., 1995, for a classification). In such conditions, small-scale interactions between snow and the landscape play a more important role than in areas with a more persistent snow cover. However, we again stress that we considered a significantly smaller number of tiles in these regions compared to Belarus and the Caucasus region.

Figure 4 shows monthly contingency scores over the Caucasus area (upper panel). In this area of interest, hits (n_{11}) and correct negatives (n_{00}) are higher than misses (n_{10}) and false positives (n_{01}) for all months. Misses are generally higher than false positives, with a peak in March – a period of likely peak accumulation for the area. Importantly, the number of snow and ground pixels in this area is comparable to each other every month (see bottom panel in Figure 4), which undercores that this validation area and winter season was not biased in favor of snow or ground pixels.

Temporal patterns of confusion scores in Belarus (Figure 5) are significantly different from the Caucasus area. In Belarus, hits are close to 1 in January and February, but quickly decrease to 0 from March on. This is consistent with a decline in the number of snow pixels in the area from March and a concurrent rise in ground pixels. Here again, it is worth noting that snow patterns in the boreal, flat

Belarus are markedly different from the mountainous Caucasus: while snow cover is either continuous or absent in Belarus, snow-topographic interactions dominate in Caucasus. This leads to coexistence of snow-covered and snow-free slopes, as well as snow-covered peaks and snow-free plains.

Results for Mount Atlas and Lebanon are reported in the Appendix (Figure A.1 and A.2, respectively) and are consistent with expectation. In the Atlas region, the number of ground pixels is significantly higher than the number of snow pixels for all months (see again Figure 3(b)) and as a result correct negatives (n_{00}) are generally close to 1. The only exception is February, when the frequency of false positives (n_{01}) significantly increases. This response could be further investigated; nonetheless, it represents an occasional response in an otherwise well classified scenario. In Lebanon, seasonal patterns in both snow-ground pixels and confusion scores are consistent with expectations and in line with other mountainous regions.

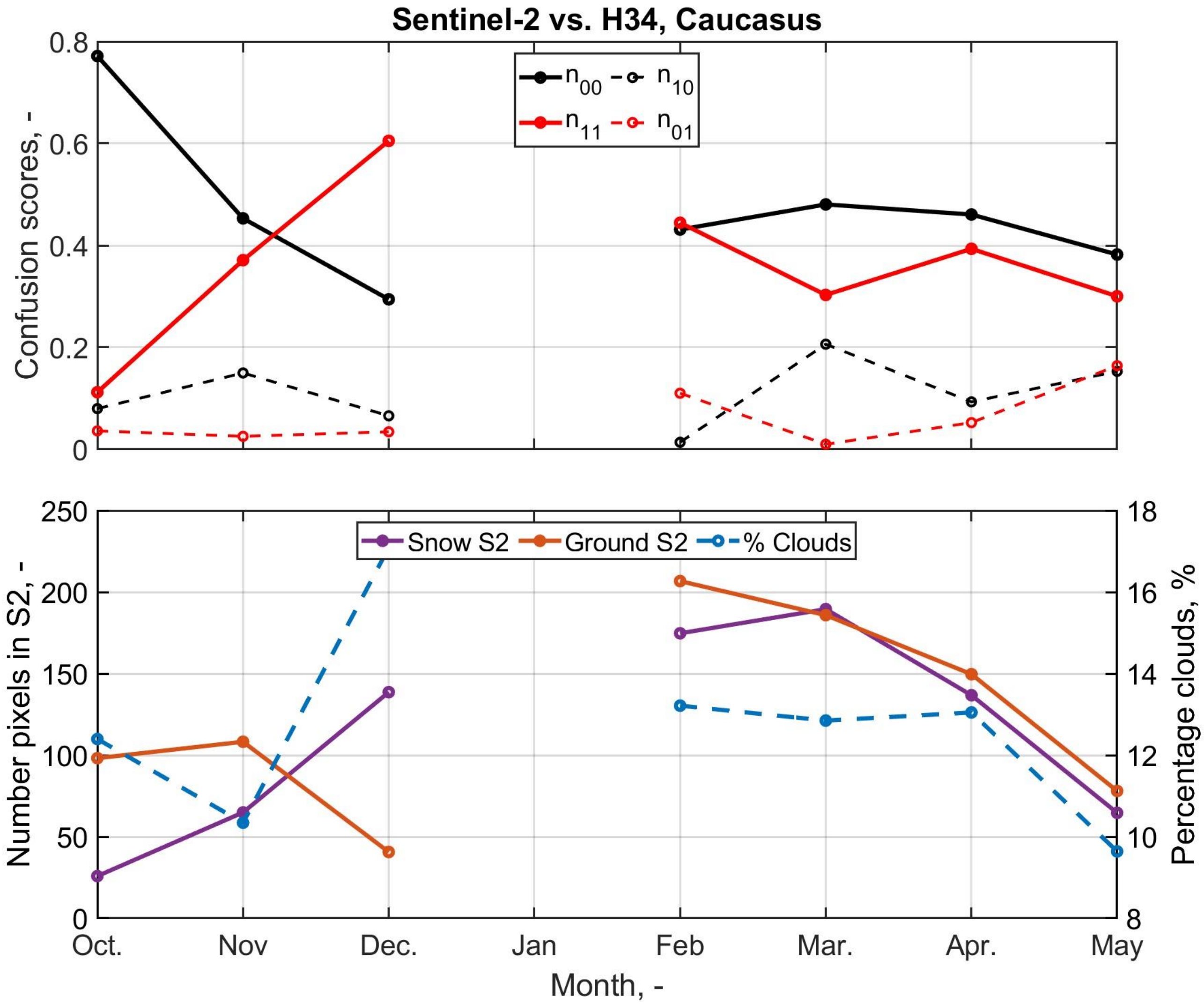


Figure 4: *Caucasus: monthly contingency scores (top), monthly number of snow and ground pixels in Sentinel 2 (bottom, left y-axis), and monthly percentage of clouds in the Sentinel-2 image (bottom, right y-axis)*

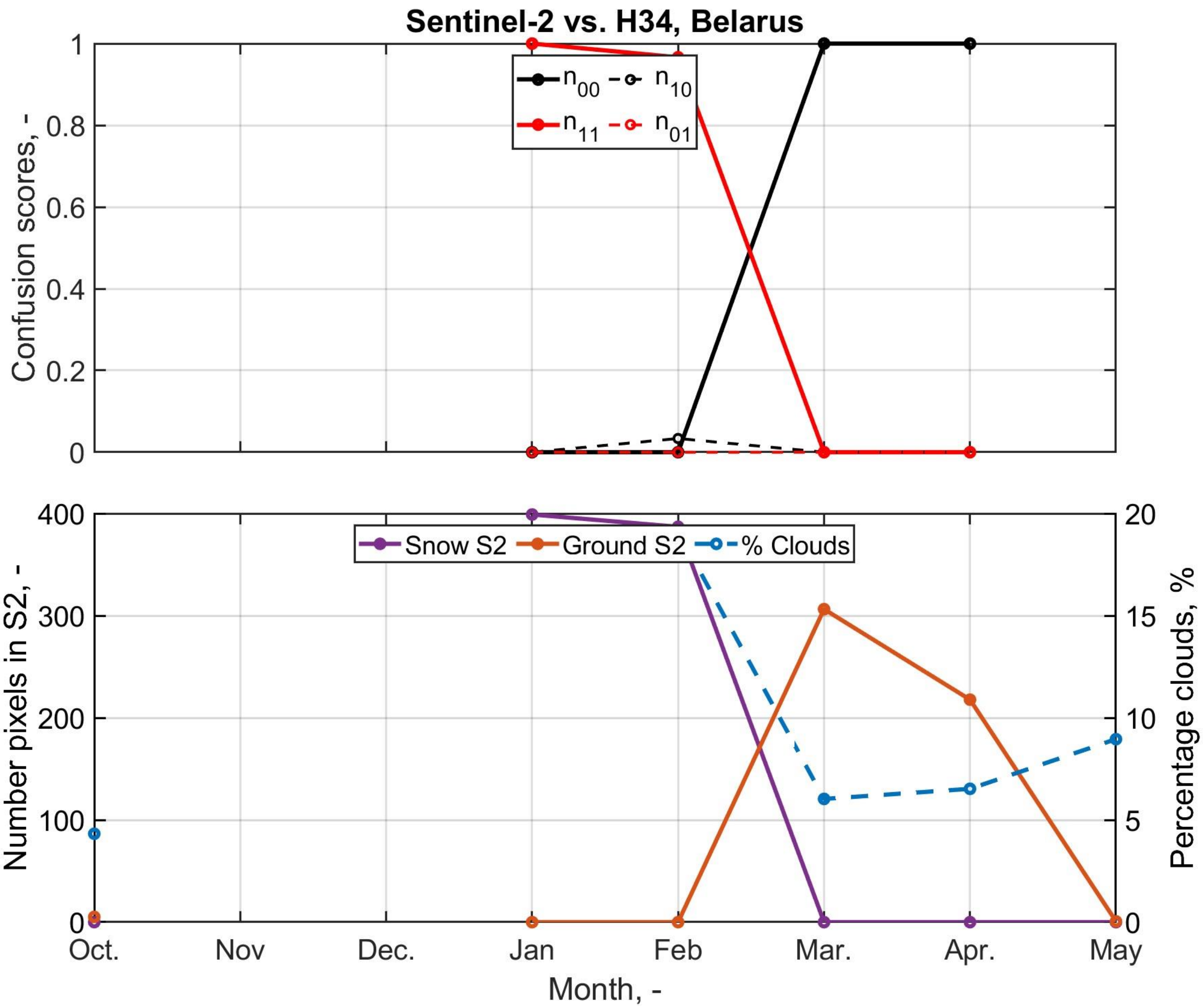


Figure 5: *Belarus: monthly contingency scores (top), monthly number of snow and ground pixels in Sentinel 2 (bottom, left y-axis), and monthly percentage of clouds in the Sentinel-2 image (bottom, right y-axis)*

Table 2: *Accuracy metrics over the four areas of interest under study, aggregated across all tiles and images available. Following the methodology applied in Piazzi et al. (2019), we assumed as threshold, target, and optimal PODs values of 70%, 80%, and 90%, respectively. In terms of FAR, we assumed as threshold, target, and optimal PODs values of 20%, 10%, and 5%, respectively.*

Area of Interest	POD	FAR	POFD	ACC	CSI	HSS
Caucasus	0.75	0.15	0.11	0.83	0.66	0.65
Belarus	0.98	0.00	0.00	0.99	0.98	0.99
Lebanon	0.96	0.59	0.13	0.87	0.41	0.52
Atlas	0.73	0.95	0.28	0.72	0.04	0.05

Accuracy metrics

Accuracy metrics in Table 2 confirm the generally good performance of H34 in detecting snow conditions over the Caucasus range (POD = 0.75), Belarus (POD = 0.98), Lebanon (POD = 0.96), and the Atlas mountains (POD = 0.73). Indeed, all PODs are above the threshold score (70%) and even above the optimal score in 2 out of 4 areas.

Given confusion scores in Section 3.1, the ratio of false alarms is expectedly low in Caucasus and Belarus; in these regions, FARs are below the threshold of 20%. FARs are higher in the Atlas area and in Lebanon, where we considered a significantly smaller number of tiles.

Overall, accuracy scores (ACC) are systematically higher than 70% in all areas.

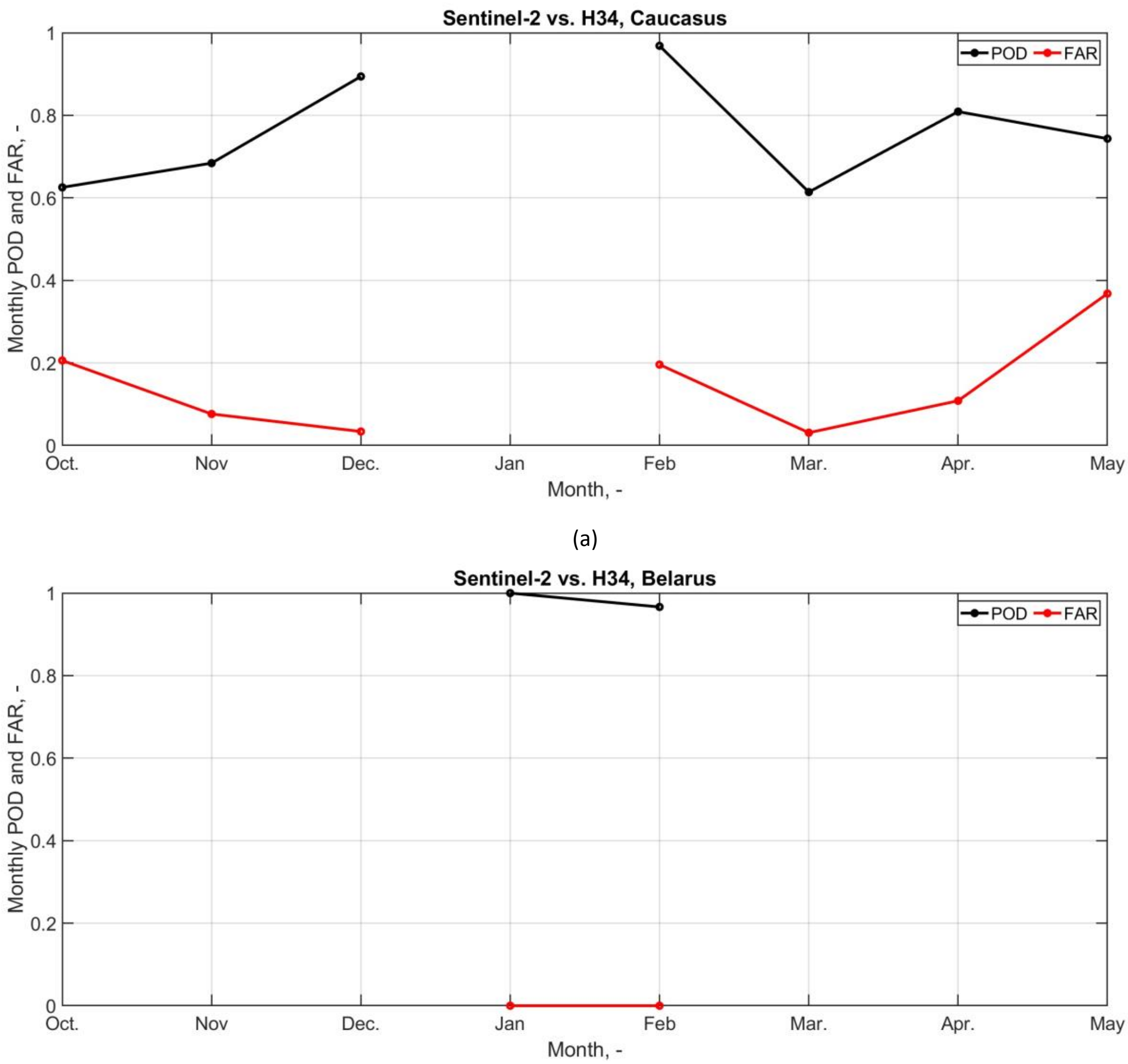
Metrics reported here for Belarus, Caucasus, and Lebanon are in line with results in Piazzi et al. (2019), with Belarus exceeding the already high scores reported by Piazzi et al. (2019) in Finland (see their Table 11). Results for the Atlas region are likely site-specific and could be target of future research.

Monthly PODs for Caucasus (Figure 6(a)) are generally very high, up to ~100% in February and always higher than 60% for all months. FARs are very low in mid-winter (e.g., December and February) and slightly increase at the beginning and at the end of the snow season, when snow becomes patchy and topographic effects increase in importance. In Belarus, on the other hand, FARs and PODs in January and February, the two months with snow on the ground according to either product in our dataset, are close to 0 and 100 %, respectively.

Results for the Atlas region and Lebanon are reported in the Appendix (Figure A.3) and are consistent with all findings so far. In the Atlas region, in particular, PODs are generally above 80% in the winter months (January to March) and decline afterwards, when snow in this area of interest becomes very patchy. During winter months, FARs are also relatively high, as already discussed in terms of confusion scores. Monthly PODs in Lebanon are close to 100% in winter and lower in late spring, with lower(higher) FARs than in the Atlas(Caucasus) region.

PODs and FARs are highly variable across tiles in Caucasus (Figure 7), a behavior that was already highlighted in other mountain regions of Europe by Piazzi et al. (2019), see their Figure 14. Piazzi et al. (2019) point to vegetation cover being a potential controlling factor for POD and FAR variability across contiguous tiles. While we have not analyzed any vegetation product here, we note that the tiles with the lowest PODs in the Caucasus region (37TFJ, 37TGJ, and 38TLN) are located in the western portion of the mountain range and are therefore closer to the Black Sea than other tiles. Being a potential source of moisture for this area, proximity to the Black Sea may imply higher vegetation cover in those tiles than elsewhere, in agreement with the arguments by Piazzi et al. (2019). As a further piece of evidence here, we also note that tiles in Figure 7 are ordered from west (left, i.e., closer to the Black Sea) to right (east) and indeed PODs show an overall increasing trend in that direction.

Results by tile in Belarus, the Atlas region, and Lebanon are reported in the Appendix (Figures A.4, A.5, and A.6, respectively) and confirm all findings discussed above.



(a)

(b)

Figure 6: *Monthly POD and FAR in Caucasus (top) and Belarus (bottom).*

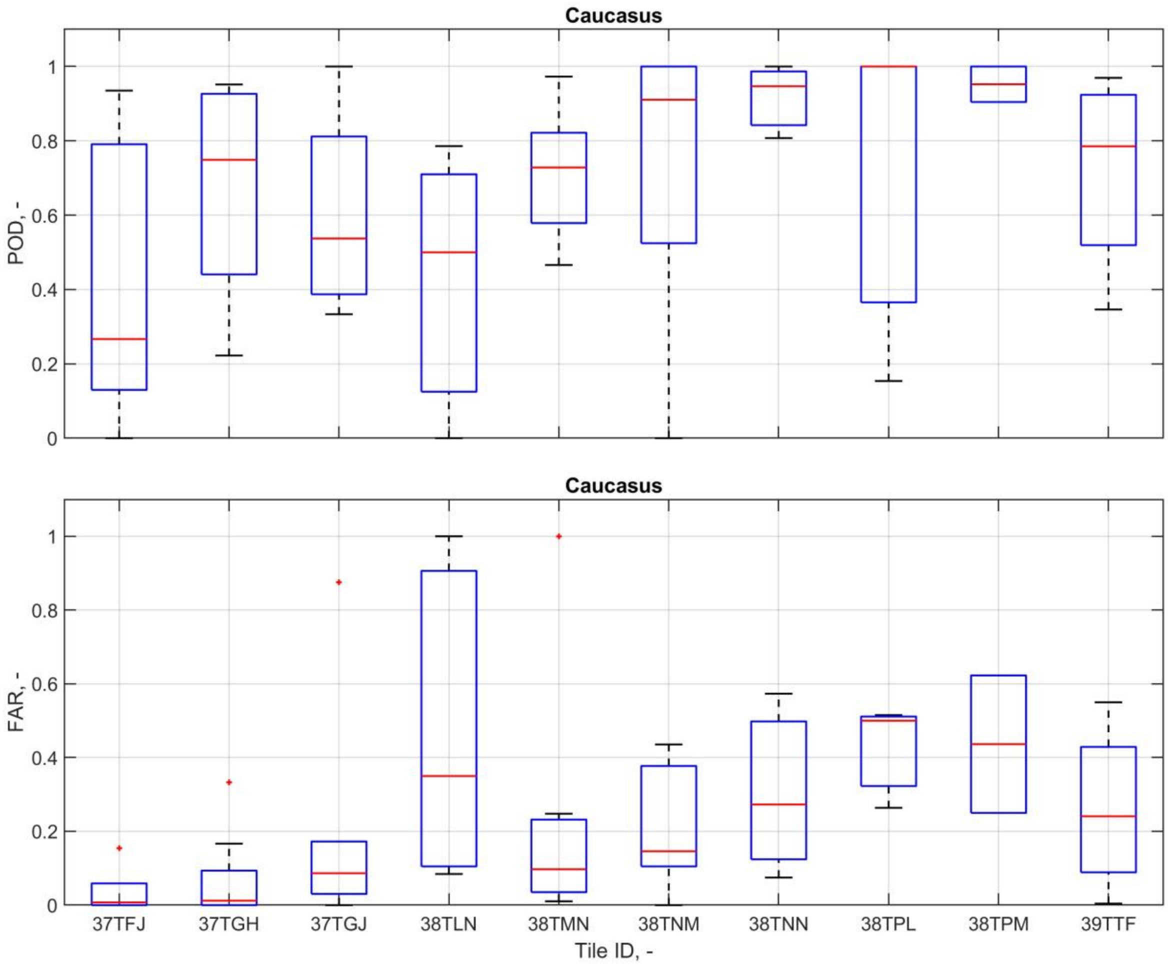


Figure 7: *Caucasus: PODs and FARs by tile.*

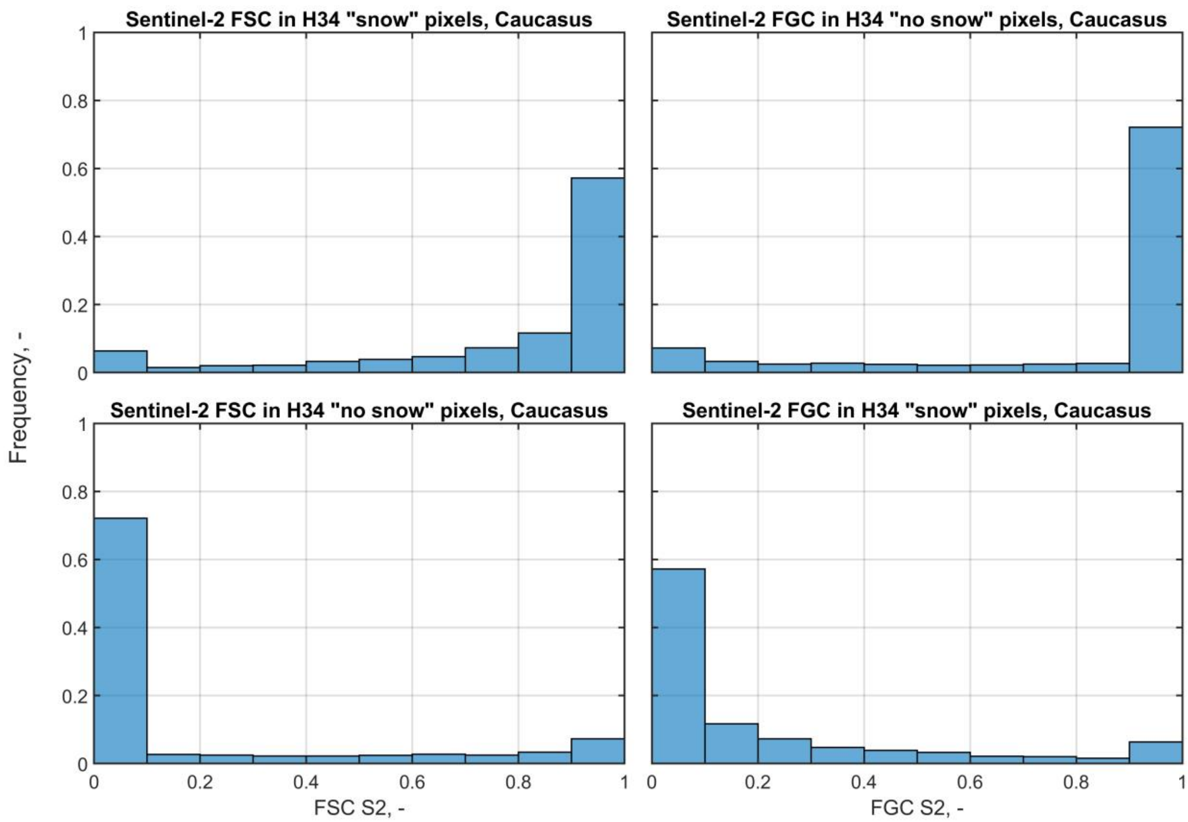


Figure 8: *Caucasus: frequency distribution of Sentinel-2 Fractional Snow Cover (FSC) and Fractional Ground Cover (FGC) in H34 pixels classified as snow and bare ground.*

H34 SCA vs. Sentinel-2 FSC and FGC

Figure 8 compares the fractional snow and ground cover (FSC and FGC, respectively) of resampled Sentinel-2 images within pixels of H34 maps that were classified by H34 as snow and ground. In conditions of perfect agreement, one would expect (1) the frequency of Sentinel-2 FSC = 1 in H34 snow pixels (upper left panel) to be close to 100%, (2) the frequency of Sentinel-2 FGC = 1 in H34 no-snow pixels (upper right panel) to be close to 100%, (3) the frequency of Sentinel-2 FSC = 0 in H34 no-snow pixels (lower left panel) to be close te 100%, and (4) the frequency of Sentinel-2 FGC = 0 in H34 snow pixels (lower right panel) to be close to 100%. In other words, one would expect Sentinel-2 FSC to increase in pixels that H34 classified as snow and Sentinel-2 FGC to increase in pixels that H34 classified as ground.

In Caucasus as well as in all our areas of interest (see also Figures A.7, A.8, and A.9), the comparison between Sentinel-2 FSC-FGC and H34 generally follows these expectations. Results are particularly good in Belarus, which again points to this region being ideal for snow detection by H34. In the Atlas region, we note a fairly large amount of H34 snow pixels for which Sentinel-2 FGC is close to 1 and FSC is close to 0 (Figure A.8). This is in agreement with this region being the one yielding the lowest PODs among the four areas considered here (Table 2). The area with the largest spread in FSC-FGC is Lebanon, where we note again that only one tile was considered.

Conclusions

We compared Sentinel-2 ad H34 snow-covered maps for winter season 2018-2019 in Caucasus,

Belarus, the Mount-Atlas region, and Lebanon. To this end, we downloaded and processed

418 Sentinel-2 Level-2A S2MSI2A images between October 1, 2018 and June 1, 2019 across 20 tiles over the four areas of interest. These images were filtered to retain those with less than 20% cloud cover, which resulted in 201 Scene-Classification maps available for comparison with H34. The comparison was performed in terms of confusion matrices and the standard accuracy metrics proposed by Piazzi et al. (2019), both in terms of average values across all tiles and images and in terms of average results by month and tile.

Results showed a good agreement between H34 and Sentinel 2, especially in flat areas like Belarus where snow-cover distribution is spatially homogeneous. In mountain areas like Caucasus, PODs are lower but still reasonably high despite small-scale snow-topographic interactions being more important in these landscapes and the scale of these interactions being sometimes smaller than the resolution of H34. We also noted that PODs tend to decrease in tiles that are likely more vegetated, which is again in agreement with previous validations (Piazzi et al., 2019). Overall, fractional-snow-cover statistics by Sentinel 2 and snow-covered areas by H34 are spatially consistent as PODs are above the threshold value of 70% in all areas (Table 2).

Additional Data and Case Studies

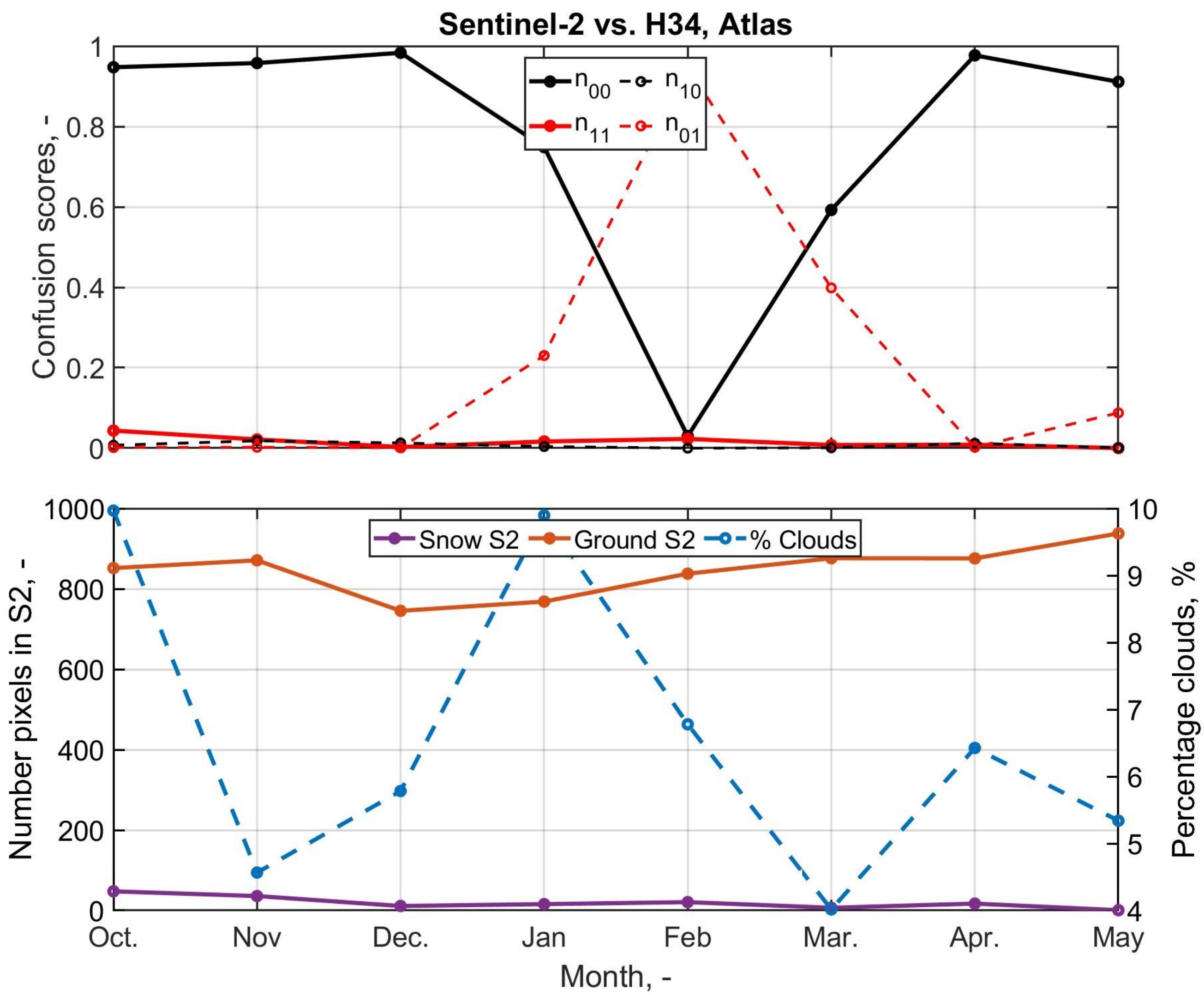


Figure A.1: *Atlas: monthly contingency scores (top), monthly number of snow and ground pixels in Sentinel 2 (bottom, left y-axis), and monthly percentage of clouds in the Sentinel-2 image (bottom, right y-axis)*

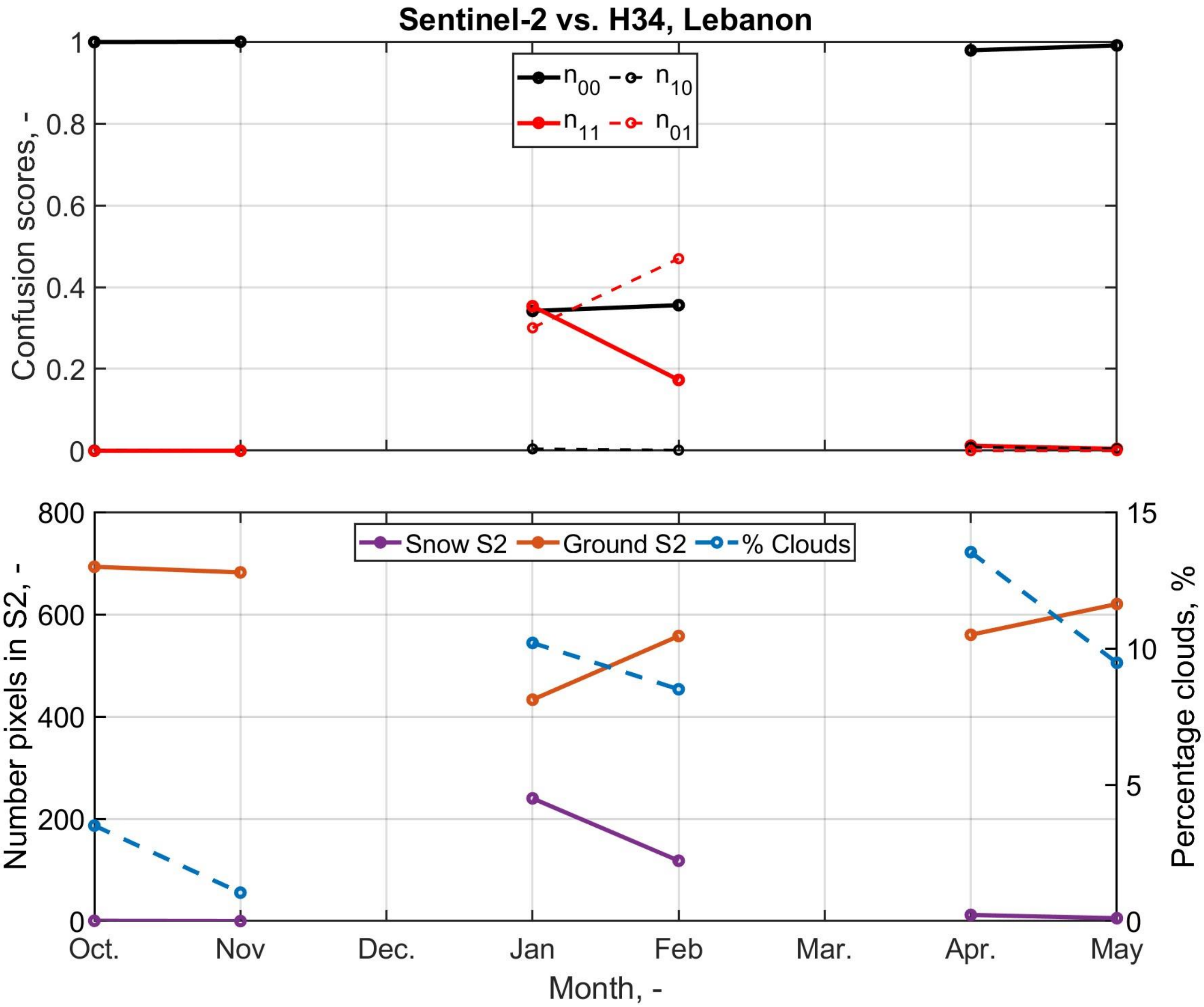
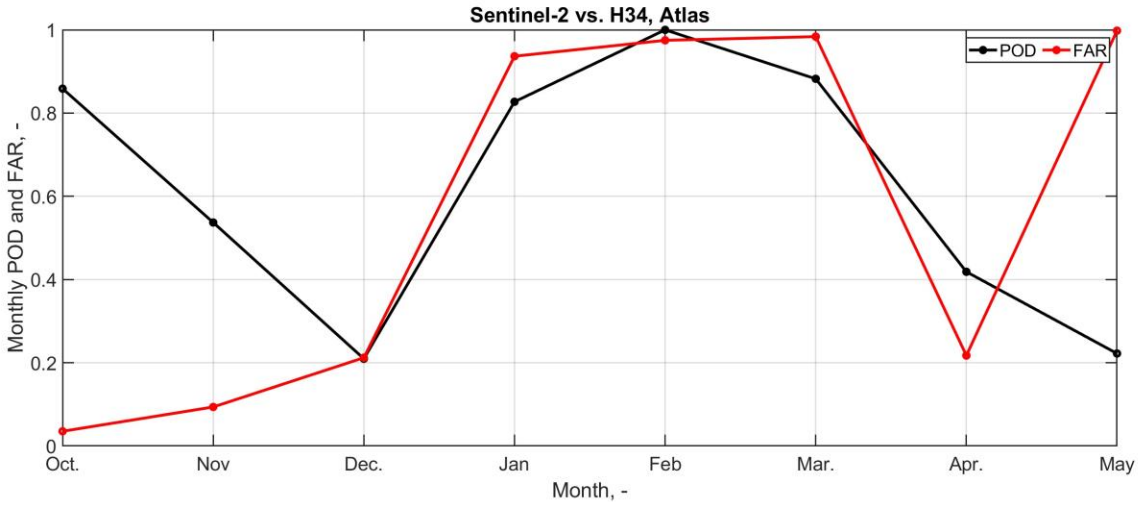
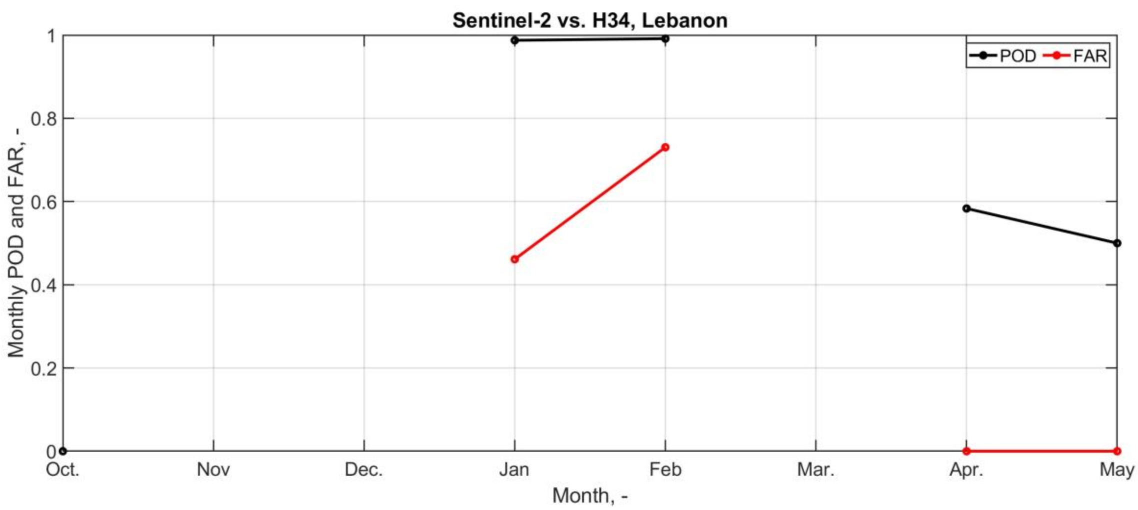


Figure A.2: *Lebanon: monthly contingency scores (top), monthly number of snow and ground pixels in Sentinel 2 (bottom, left y-axis), and monthly percentage of clouds in the Sentinel-2 image (bottom, right y-axis)*



(a)



(b)

Figure A.3: *Monthly POD and FAR in the Atlas region (top) and Lebanon (bottom).*

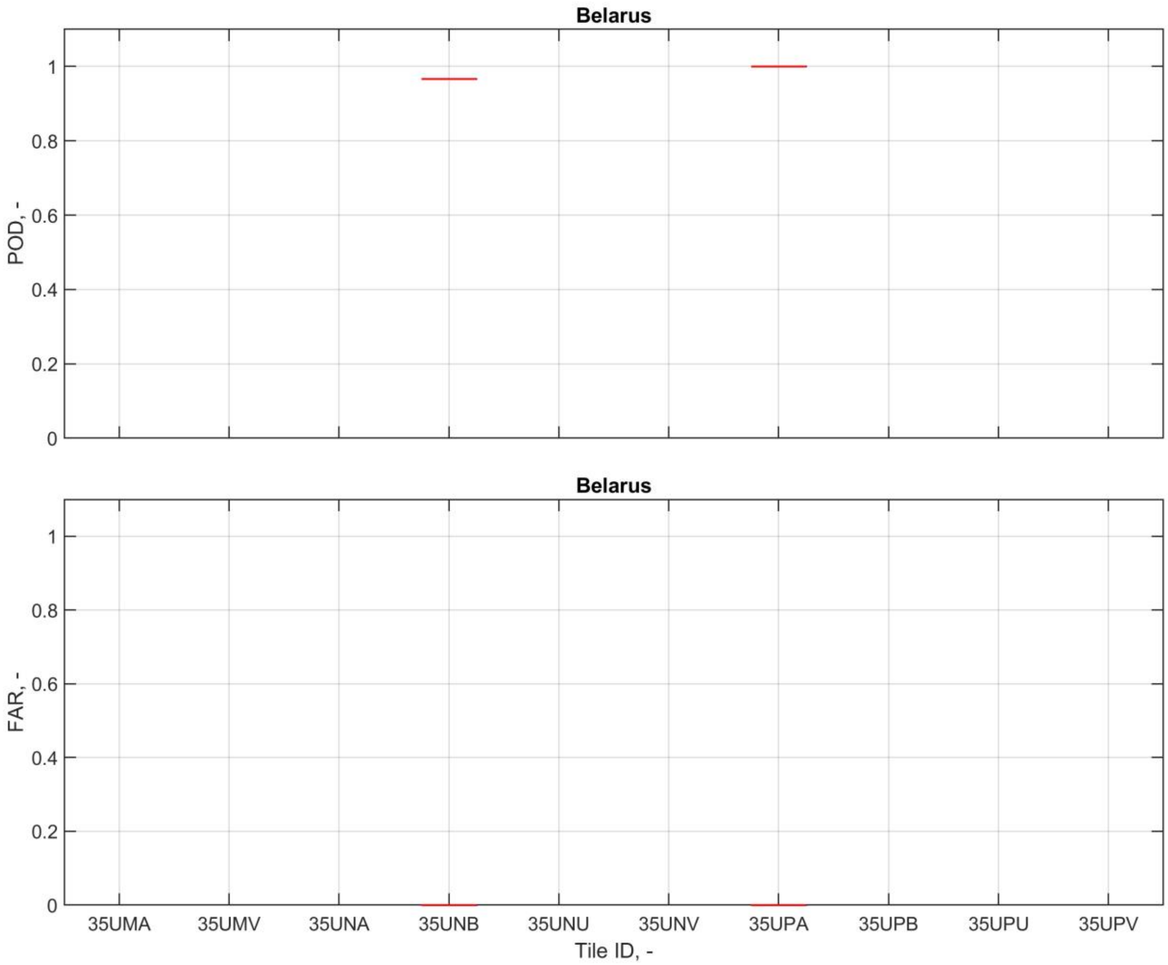


Figure A.4: *Belarus: PODs and FARs by tile.*

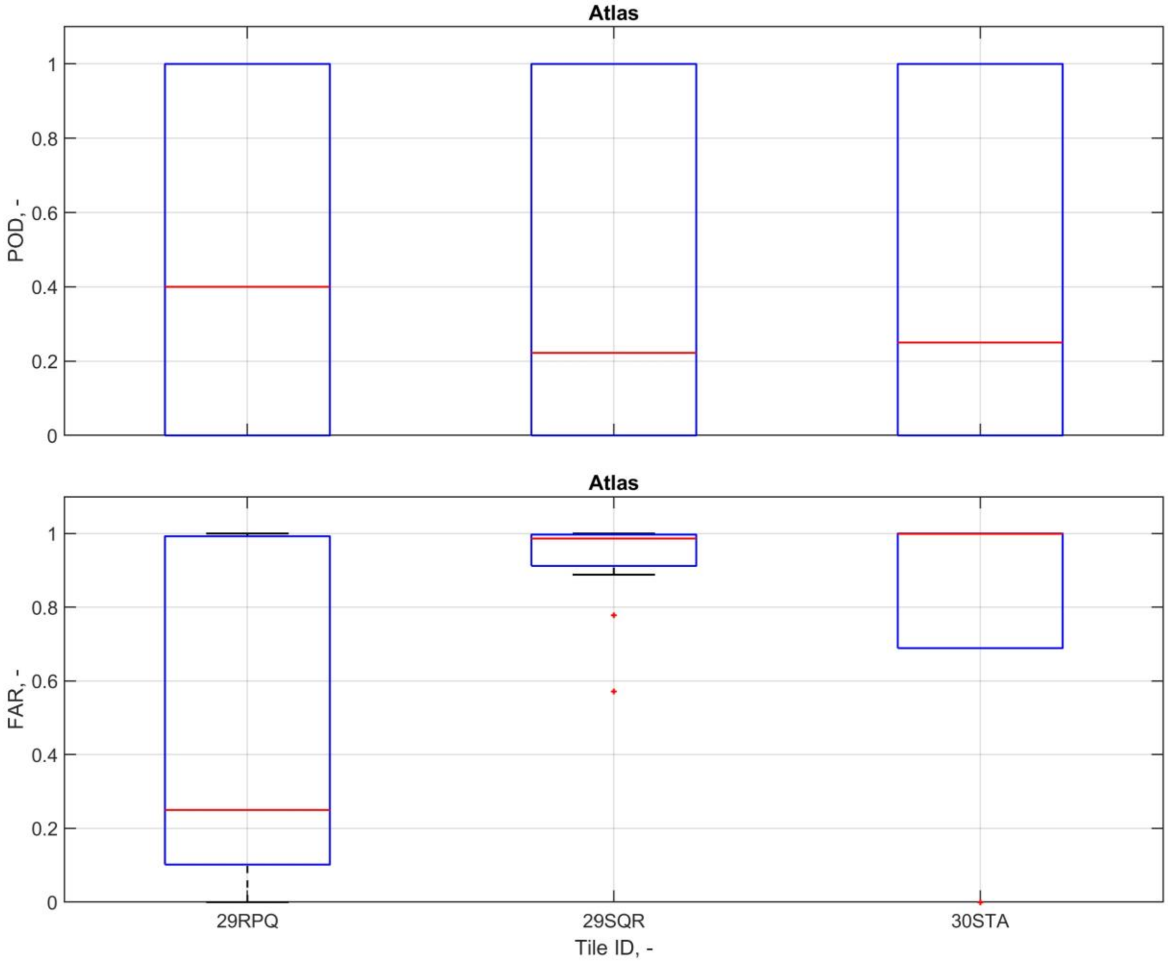


Figure A.5: *Atlas: PODs and FARs by tile.*

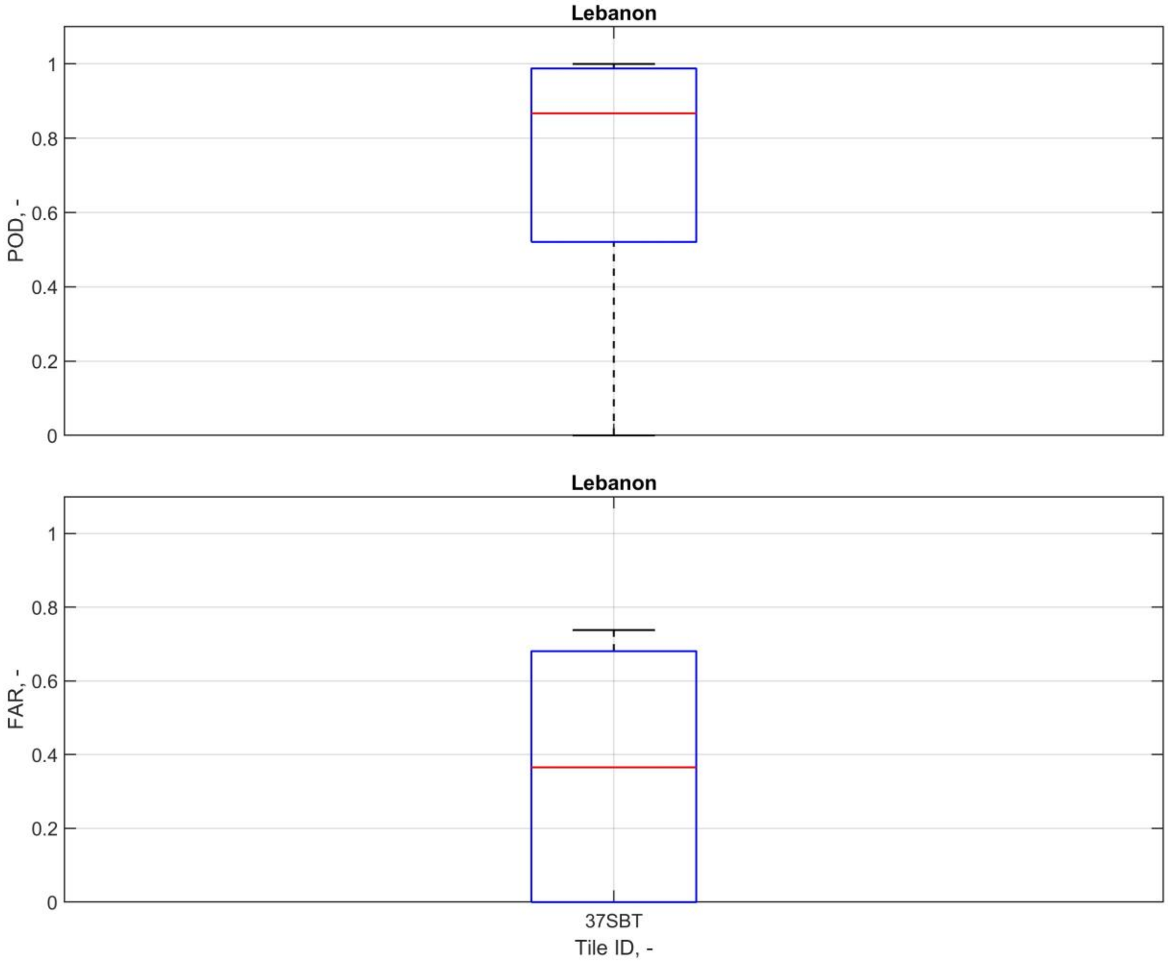


Figure A.6: *Lebanon: PODs and FARs by tile.*

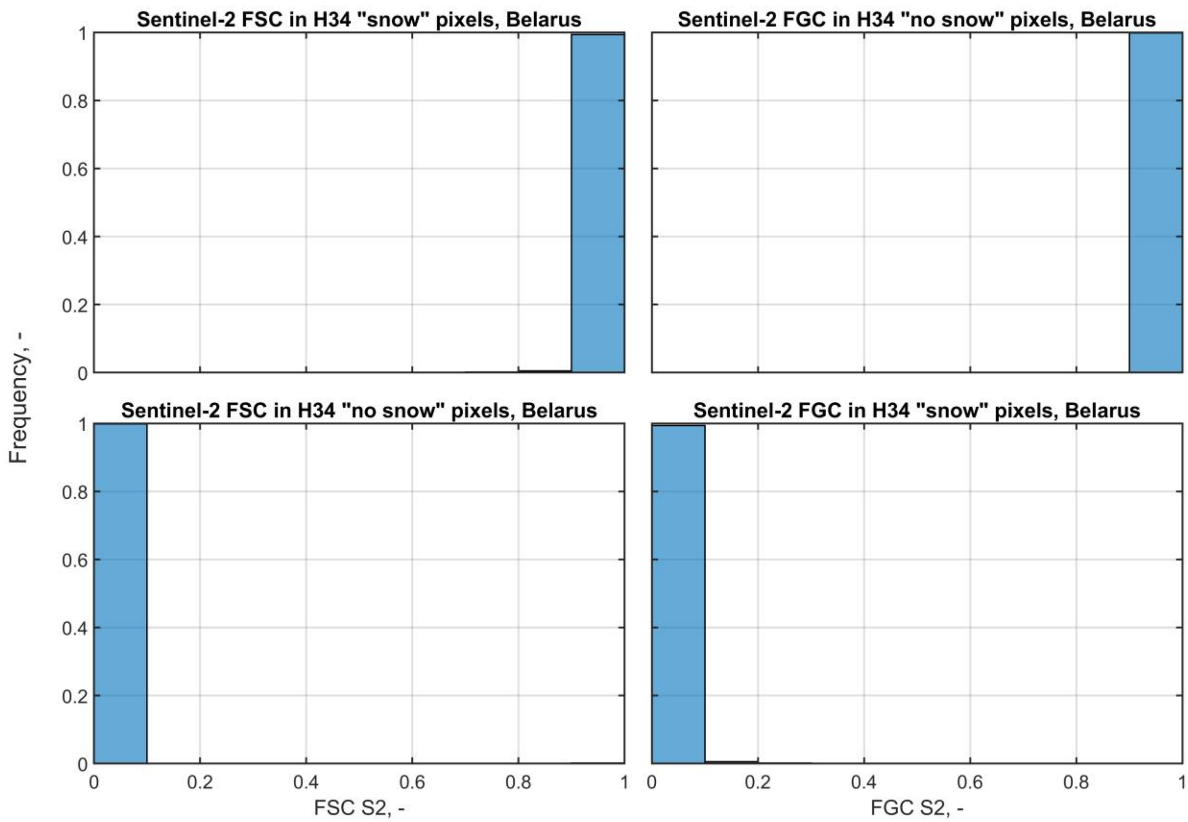


Figure A.7: *Belarus: frequency distribution of Sentinel-2 Fractional Snow Cover (FSC) and Fractional Ground Cover (FGC) in H34 pixels classified as snow and bare ground.*

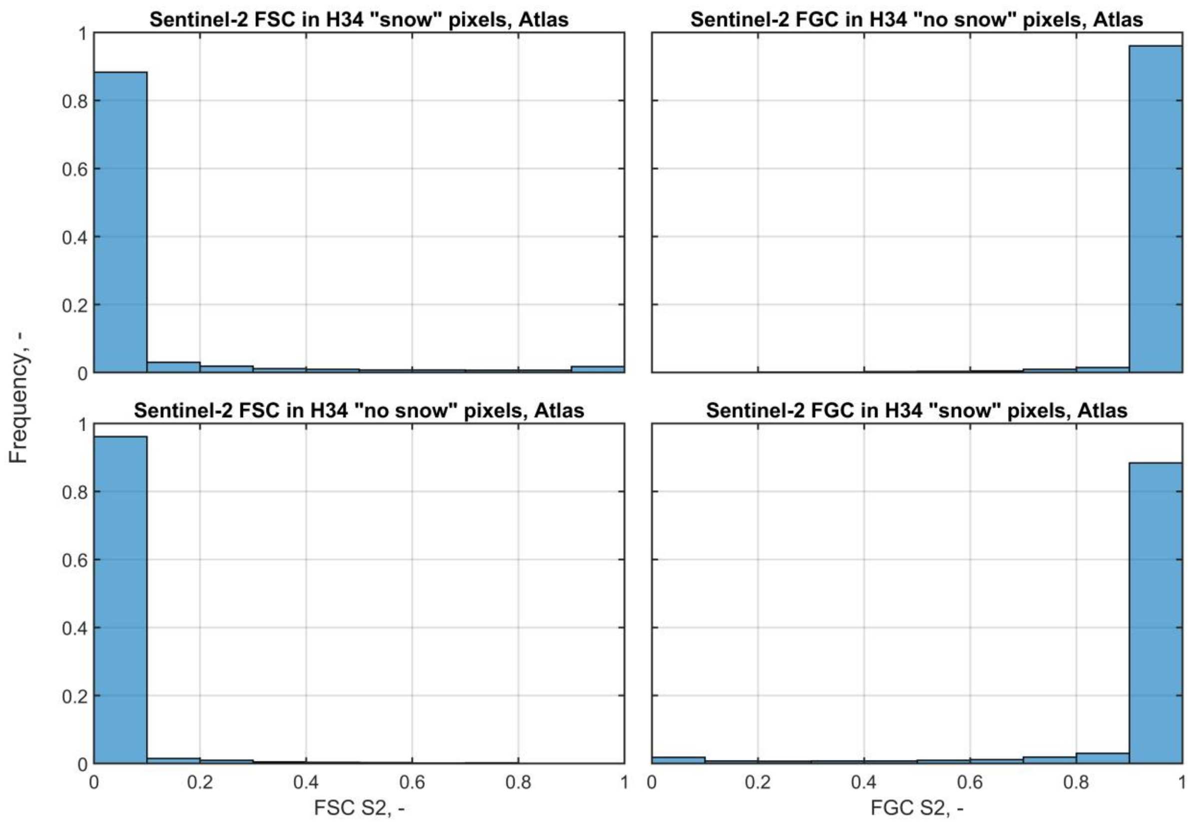


Figure A.8: *Atlas: frequency distribution of Sentinel-2 Fractional Snow Cover (FSC) and Fractional Ground Cover (FGC) in H34 pixels classified as snow and bare ground.*

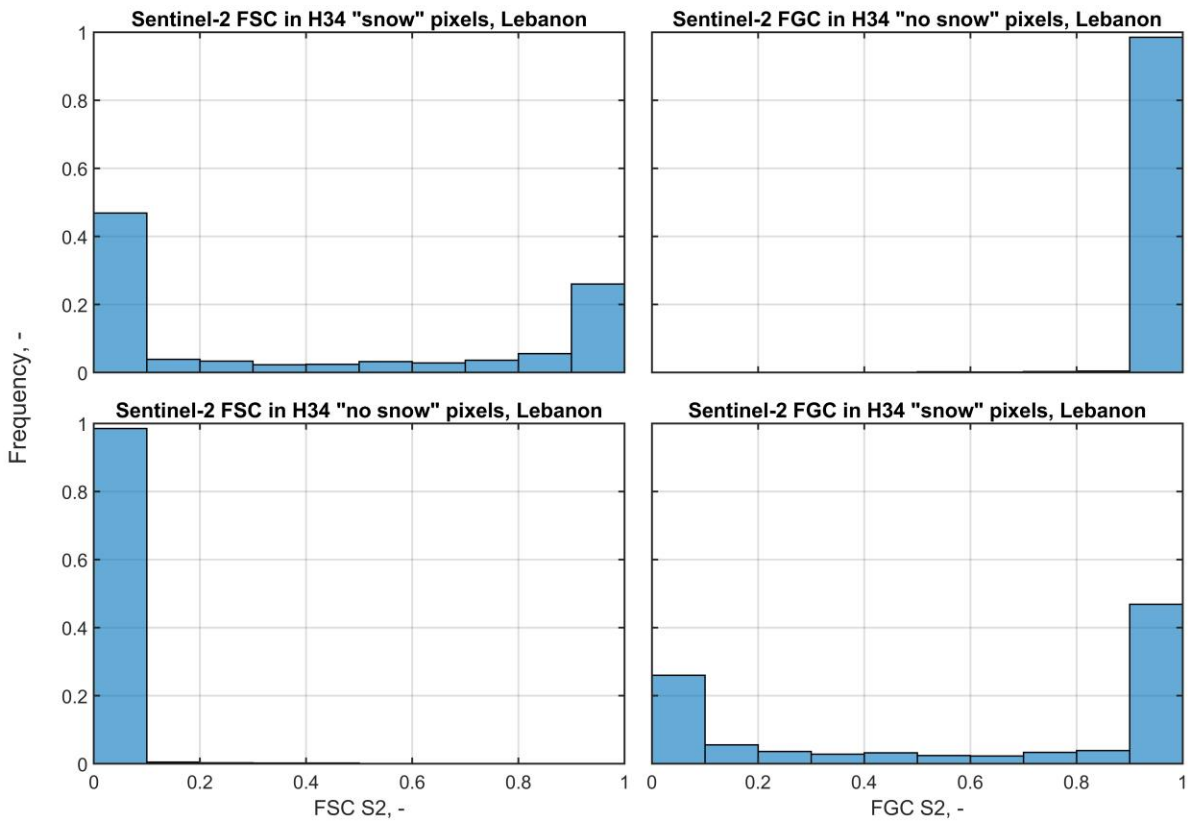


Figure A.9: *Lebanon: frequency distribution of Sentinel-2 Fractional Snow Cover (FSC) and Fractional Ground Cover (FGC) in H34 pixels classified as snow and bare ground.*

Bibliography

M. Kottek, J. Grieser, C. Beck, B. Rudolf, and F. Rubel. World Map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift*, 15(3):259–263, 07 2006. doi: 10.1127/0941-2948/2006/0130. URL

G. Piazzi, C. Tanis, S. Kuter, B. Simsek, S. Puca, A. Toniazzo, M. Takala, Z. Akyürek,

S. Gabellani, and A. Arslan. Cross-Country Assessment of H-SAF Snow Products by Sentinel-2 Imagery Validated against In-Situ Observations and Webcam Photography. *Geosciences*, 9(3):129, Mar 2019. ISSN 2076-3263. doi: 10.3390/geosciences9030129. URL http://dx.doi.org/10.3390/geosciences9030129.

M. Sturm, J. Holmgren, and G. E. Liston. A seasonal snow cover classification system for local to global applications. *Journal of Climate*, 8(5):1261–1283, 1995. doi: 10.1175/1520-0442(1995)008<1261:ASSCCS>2.0.CO;2. URL https://doi.org/10.1175/1520-0442(1995)008<1261:ASSCCS>2.0.CO;2.

Annex 2

Validation methodology for H10 – Snow detection

Pawel Przeniczny, pawel.przeniczny@imgw.pl

Latest update: January 31, 2011

This document describes the methodology applied when validating H-SAF snow product H10 – Snow detection (Snow mask) by VIS/IR radiometry.

Validation procedure

To properly validate H10 product, the following steps have to be taken:

1. Observation data containing snow cover measurements have to be gathered.
2. Satellite product needs to be acquired.
3. Both observation and satellite data series need to be checked for consistency.
4. Comparison between the observation data and the product has to be performed.
5. Results of the comparison need to be presented.

1. Observation data

Measurements at ground stations (SYNOP and other lower level posts) are made on a daily basis at 0600 UTC. Metadata concerning the method and instrument used for snow measurement as well as accuracy and frequency of measurements should be included.

From the data collected by ground network, a subset containing snow cover depth (SD) for the *reference season* (1.10.2009 – 31.09.2010) is extracted and a local database is created.

The data is stored in plain text. Each file contains the data from all reporting stations for one day of the reference season. For each station the following columns (separated by whitespace) are assigned:

– date and time of measurement,
– number and name of the station as well as it's coordinates: latitude (degrees), longitude (degrees) and height (m asl.),
– a flag indicating whether the station is located in mountainous or flat/forested area. The masking is performed by applying the mountain mask. The file "mountainmask_sr.h5" is available at the TSMS ftp site at /OUT/h10/mountainmask

address: ftp://hsaf.meteoroloji.gov.tr

username: *snowtur* password: *rs37kar* snow cover depth (in cm).

2. Satellite product

H-SAF H10 product is available at the FMI ftp server, ftp://ftp.fmi.fi.

An ftp client (e.g. FileZilla, WinSCP) is required to log in and retrieve the product, which is stored in the binary HDF5 files.

Conversion to ASCII format is recommended to simplify algorithms used in validation software.

To this end, one can apply a H10 converter prepared by FMI and available at the FMI ftp server (*snobs1_hdf5_to_ascii.tar.gz*).

Alternatively one can use a freeware *h5dump.exe* tool on the binary product files. *h5dump.exe* can extract a subset containing latitude, longitude and the snow cover parameter from the full satellite product.

Each partner performs validation only on a local domain, thus the computations may be speed up by extracting from a full product a subset limited to local coordinates.

3. Data consistency check

To guarantee high quality of the validation it is advised to check if both the observation data and the satellite product are available for all days of the reference season.

We think it should be recommended to make a 'sanity check' both on the satellite product downloaded from the ftp server and the observational data – a quick look on the filename format and file modification dates can prevent making validation on wrong (e.g. old version of the product) or incomplete datasets (e.g. missing observation data).

4. Comparison between the observation data and the product

The comparison between the observation data and the satellite product (point to pixel) needs to be performed.

To compare the satellite product with observation data, the measurement from the station that is the nearest to the satellite pixel is used.

From satellite product, only pixels with code 0 (snow) and 85 (ground) are taken into consideration. Cloudy and data-missing pixels are discarded from comparison.

It has been stipulated[1] to treat the measurement as snow occurrence if the snow depth parameter value is equal or greater than 2 cm,

$$SD \geq 2 \text{ cm} \quad (1)$$

Thus:

1. HITS counter is increased if (1) is true and the satellite pixel value is 0,
2. MISSES counter is increased if (1) is true and the satellite pixel value is 85,
3. FALSE ALARMS counter is increased if (1) is false and the satellite pixel value is 0,
4. CORRECT NEGATIVES counter is increased if (1) is false and the satellite pixel value is 85.

These relations can be presented in a table:

		Observation data		
		SD ≥ 2 cm	**SD < 2cm**	
Satellite product	**Snow code = 0**	**HITS**	**FALSE ALARMS**	HITS + FALSE ALARMS
	Snow code = 85	**MISSES**	**CORRECT NEGATIVES**	MISSES+ CORRECT NEGATIVES
		HITS + MISSES	FALSE ALARMS + CORRECT NEGATIVES	

From these classification results, different scores for dichotomous statistics can be calculated. To simplify the formulae, the following notation is used:

- A = number of HITS
- B = number of FALSE ALARMS
- C = number of MISSES
- D = number of CORRECT NEGATIVES

Probability of detection:

- POD = A/(A+C)

[1] Antalya Snow Validation Meeting, 1-2 December 2010; more detailed explanation by Panu Lahtinen can be found in "Validation methodology for H11 – Snow status" document.

False alarm ratio:

- FAR = B/(A+B)

Probability of false detection:

- POFD = B/(B+D)

Accuracy:

- ACC = (A+D)/(A+B+C+D)

Critical success index:

- CSI = A/(A+B+C)

Heidke skill score:

- HSS = 2(AD-BC) / [(A+C)(C+D) + (A+B)(B+D)]

5. Results of the comparison

Above-mentioned statistics need to be calculated both for flat/forested and for mountainous areas for each month of the reference season as well as for the whole season.

Also a merged product needs to be validated in the similar manner[2].

Results should be presented in the form of contingency tables and statistical scores.

To complement the validation, 3 case studies for the reference season should be presented.

For each case study quantitative analysis in the same manner like for the longer period (explained above) should be performed. Additionally, qualitative analysis by comparing pictures of H10 product with different satellite product (e.g. Meteosat 9 RGB composition of channels 1,3 and 9-inverted) should be performed.

For each case study teams are welcome to introduce their own additional analysis or algorithms.

2 Merged product validation can be achieved by combining the HITS, FALSE ALARMS, MISSES and CORRECT NEGATIVES obtained for flat/forest and mountainous areas separately.

References

WMO Guide to Meteorological Instruments and Methods of Observation. WMO-No. 8, Seventh edition. 2008.

Minutes of Antalya Snow Validation Meeting, 1-2 December 2010

Daniel S. Wilks *Statistical Methods in the Atmospheric Sciences,* Academic Press 2006.