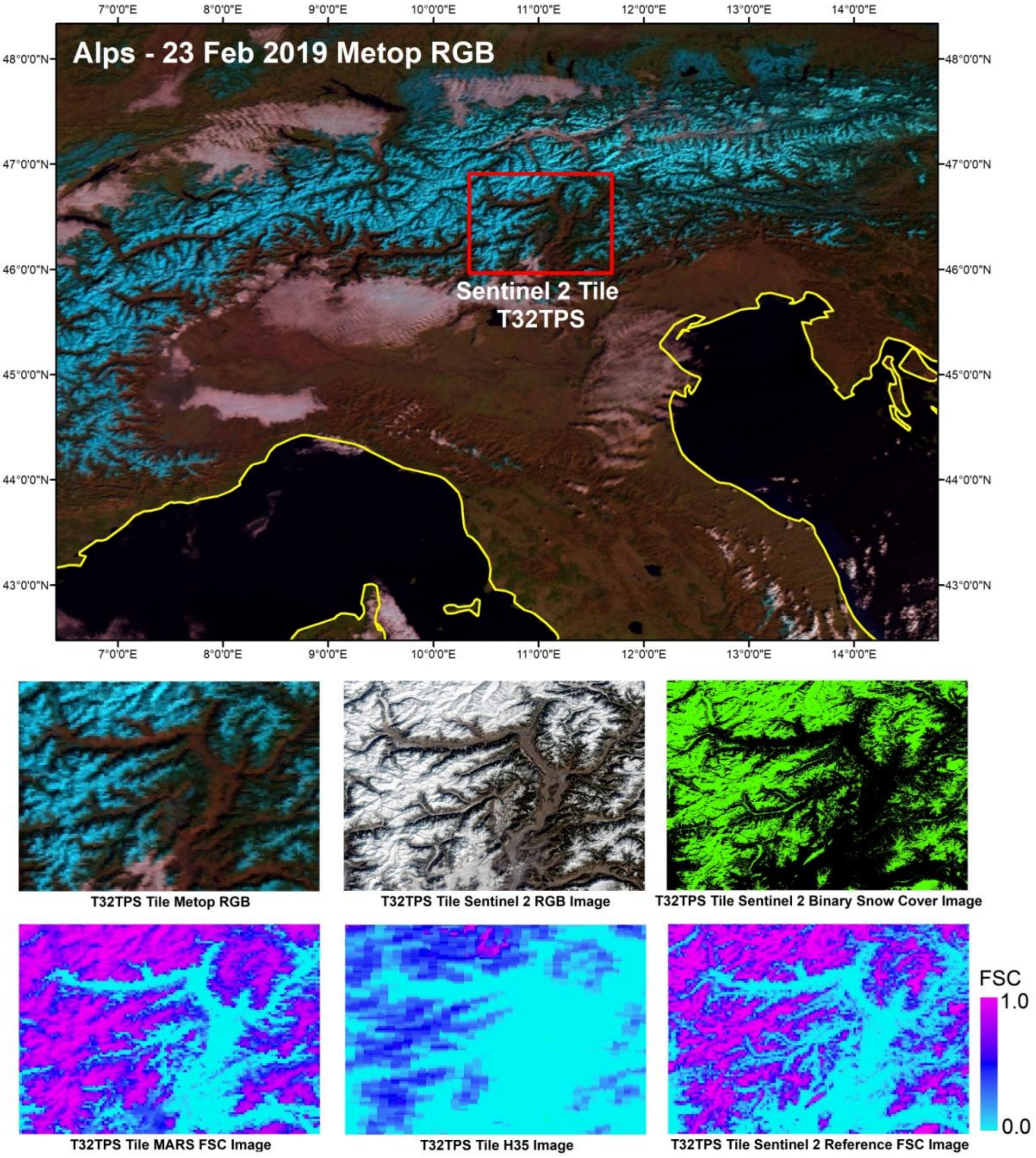
Support to Development of H-SAF Snow Product for the Mountainous Areas by using Multivariate Adaptive Regression Splines

H_AVS18_03

Final Report - November 2020 Turkey

Semih KUTER (PhD)



T32TPS Tile Metop RGB

T32TPS Tile Sentinel 2 RGB Image

T32TPS Tile Sentinel 2 Binary Snow Cover Image

T32TPS Tile MARS FSC Image

T32TPS Tile H35 Image

T32TPS Tile Sentinel 2 Reference FSC Image

TABLE OF CONTENTS

1. INTRODUCTION

Snow is an important physical element of the Earth's surface and it has various distinctive features that make the continuous monitoring of its spatial and temporal extent highly significant for fundamental environmental studies. Regarding climatology, snow cover has a direct impact on the energy exchange process between the Earth and its atmosphere since: (*i*) it reflects considerable amount of incoming solar radiation due to its high albedo in the visible and the near-infrared regions, and (*ii*) it prevents the release of heat from the Earth's surface during winter by acting as an insulator with poor heat conducting characteristic (Hall and Martinec 1985). Snow also plays an important role in hydrology of the world's mountainous regions as well as mid- to high-latitude alpine environments (Cline et al. 1998; Hartman et al. 1999; Lehning et al. 2006). Water accumulated in the snowpack acts as a high-volume frozen-water reservoir during winter season and keeps precipitation until snowmelt runoff begins, which holds a potential risk of flooding, yet also constitutes a valuable source of energy and water supply (Akyürek et al. 2010; Romanov et al. 2003).

Spatial extent of snow has been declared as an essential climate variable (Hüsler et al. 2012) and, thus, accurate modeling of snow cover is crucial for the better prediction of snow water equivalent and, consequently, for the success of general circulation and weather forecasting models as well as climate change and hydrological studies (Brown et al. 2003; Clark et al. 2006; Dozier 1992; Foster et al. 1996; Hosaka et al. 2005). Although snow cover analysis based on in situ measurements (Beniston 2003; Brown and Goodison 1996; Laternser and Schneebeli 2003; Marty 2008) provides high-quality long time series data, snow cover mapping through field surveys is not economically and logistically practical (Gafurov and Bárdossy 2009). Remote sensing (RS) has been offering a powerful alternative for consistent monitoring of snow cover extent with timely and multispectral data acquired by various kinds of coarse and medium spatial resolution instruments since the mid-60s, when the first operational snow mapping was done by National Oceanic and Atmospheric Administration (NOAA) (Hall and Martinec 1985).

The fractional snow cover products (H12 and H35) are daily operational products based on multi-channel analysis of the Advanced Very High Resolution Radiometer (AVHRR) on board National Oceanic and Atmospheric Administration (NOAA) and Meteorological Operational Satellite (Metop) satellites. They are supplied by the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT) Satellite

Application Facility on Support to Operational Hydrology and Water Management (HSAF). Fractional snow cover (FSC) is generated at pixel resolution by exploiting the brightness intensity, which is the convolution of the snow signal and the fraction of snow within the pixel and the sampling is carried out at 1-km intervals. The product for flat/forested regions is generated by Finnish Meteorological Institute (FMI) and the product for mountainous areas is generated by Turkish State Meteorological Service (TSMS). Both products, thereafter, are merged at FMI. The pixel codes for H35 product is given in Table 1.

Table 1. H35 pixel codes.

Pixel Code	Explanation
0-100	FSC
250	Land
251	Cloud
252	Sea/Water
253	NoClass
254	Dark
255	NoData/Space

1.1 Main Objectives of the AVS Study

In the recent study (Kuter et al. 2018), an alternative approach for FSC mapping from MODIS data (MOD02HKM) is proposed in RS by using *Multivariate Adaptive Regression Splines* (MARS) for the first time. MARS is a nonparametric regression technique developed by Friedman (1991) and it is widely used in data mining and estimation theory in order to build flexible regression models for complex and high-dimensional nonlinear data. The main advantage of MARS is its ability to define the underlying functional relationships between dependent and independent variables by simply and smoothly connecting piecewise linear polynomial pieces, i.e., linear splines, resulting in a flexible model that can handle both linear and nonlinear behavior.

The main goal of this AVS activity is to develop an alternative methodology to improve/refine the current H12 and H35 FSC products in use by implementing the new MARS-based approach, as well as *Random Forests* (RF) methodology. In the original AS activity proposal, the RF methodology was not initially considered; however, due to its recent successful implementations in the field of RS (Pelletier et al. 2016; Timm and McGarigal 2012; Tramontana et al. 2015; Yu et al. 2011), it was also included as the second machine learning method in the analysis. The following items can be enlisted as the main objectives of this AS activity:

- to investigate the use of MARS and RF methods in FSC mapping from AVHRR data,
- to use Sentinel 2 MSI data in generating the reference FSC maps by using Sen2Cor v2.8 (Mueller-Wilm 2019):
    - We already made a validation study for the assessment of Sentinel 2 binary snow maps generated by the Sentinel 2's own processing module Sen2Cor in our previous study (Piazzi et al. 2019) and Sentinel 2 binary snow maps were tested against ground-based snow depth measurements. 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.
- to assess and compare the FSC mapping accuracies of the developed MARS and RF models,
- and finally, to evaluate the applicability of MARS and RF algorithms for an operational FSC product generation.

2. MATERIALS AND METHODS

2.1 Satellite Data and Basic Image Pre-processing Steps

In total, 332 Sentinel 2 images over Alps, Tatra Mountains and Turkey acquired between November 2018 and April 2019 are used in order to generate the necessary data for the training of the models. An independent test dataset consisting of 25 Sentinel 2 scenes acquired between November 2018 and April 2019 over Alps, Tatra Mountains, Turkey, Greenland and Norway is reserved for accuracy assessment. Please note that initially 16 test images over Alps, Tatra and Turkey were used and their results were reported in the interim report. However, in the second phase of the study, 9 extra Sentinel 2 images over Greenland and Norway are added to the initial test dataset, resulting in 25 test scenes in total. The details of the Sentinel 2 images employed in the training and the testing are given in Tables 2 and 3, respectively. Each Sentinel 2 image with scene cloud contamination less than 30% was downloaded from the dedicated web page of the United States Geological Survey (USGS) (https://earthexplorer.usgs.gov/). Location of the corresponding Sentinel 2 tiles belonging to Alps, Tatra Mountains, Turkey, Greenland and Norway are shown in Figures 1-5, respectively.

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. Class labels used in Sen2Cor are given in Table 4.

All Sentinel 2 binary snow maps are re-projected to a common GCS_WGS_1984 coordinate system in order to match with the projection of the corresponding AVHRR bands and H35 images. Since the same algorithm is used in H12 and H35 production, H35 product is employed in the analysis.

Table 2. Sentinel 2 tiles used in the training process.

Alps							
Sentinel 2 Tile No	**Nov 2018**	**Dec 2018**	**Jan 2019**	**Feb 2019**	**Mar 2019**	**Apr 2019**	**Total No. of Tiles**
T31TGK	-	2	3	3	2	4	14
T31TGL	-	2	1	3	2	2	10
T32TLS	-	1	2	3	3	3	12
T32TMS	1	2	2	3	1	1	10
T32TNS	-	1	-	1	1	1	4
T32TPS	1	4	4	2	5	1	16
T32TPT	1	1	1	3	3	1	10
T32TQS	1	2	2	3	4	1	13
T32TQT	1	-	-	1	-	-	2
T33TUM	-	3	5	2	-	-	10
T33TUN	1	1	-	1	-	-	3
Alps Total	6	19	20	25	21	14	104
Tatra							
T33UYQ	1	-	1	2	1	2	7
T34TGT	3	-	-	2	4	2	11
T34UCU	1	1	2	4	4	5	17
T34UCV	1	-	1	3	2	4	11
T34UDU	-	1	2	2	2	3	10
T34UDV	1	-	-	3	1	5	10
T34UFV	1	-	-	2	1	1	5
T34UGU	3	-	-	2	2	3	10
T35TLL	1	2	1	3	5	4	16
T35TLM	2	-	-	3	2	2	9
T35TLN	2	-	-	2	2	1	7
T35TML	-	1	1	1	2	2	7
T35TMM	2	2	-	4	2	2	12
T35ULP	1	-	-	2	3	3	9
Tatra Total	19	7	8	35	33	39	141
Turkey							
T36SVF	3	5	1	4	4	5	22
T36TWL	4	2	3	2	8	3	22
T37SED	1	-	1	2	2	1	7
T37SFD	3	-	1	2	1	2	9
T37TEE	1	1	-	2	1	1	6
T37TFE	4	2	1	1	-	1	9
T38SKH	2	1	-	2	1	-	6
T38SLH	2	1	1	-	1	1	6
Turkey Total	20	12	8	15	18	14	87
Grand Total	**45**	**38**	**36**	**75**	**72**	**67**	**332**

Table 3. Sentinel 2 images reserved as independent test dataset (please note that the results of the first 16 test images over Alps, Tatra and Turkey were reported in the interim report).

Alps		
Test Image No.	Tile No	Date
Test 1	T31TGK	20 Jan 2019
Test 2	T32TLS	1 Nov 2018
Test 3	T32TLS	21 Mar 2019
Test 4	T32TPS	23 Feb 2019
Test 5	T32TPT	18 Dec 2018
Test 6	T32TPT	19 Apr 2019
Tatra		
Test 7	T34UCU	19 Dec 2018
Test 8	T34UGU	30 Mar 2019
Test 9	T35TLL	18 Feb 2019
Test 10	T35TLL	24 Apr 2019
Turkey		
Test 11	T36SVF	18 Mar 2019
Test 12	T36TWL	3 Feb 2019
Test 13	T37SFD	9 Nov 2018
Test 14	T37TEE	9 Dec 2018
Test 15	T38SKH	20 Jan 2019
Test 16	T38SKH	25 Apr 2019
Greenland		
Test 17	T22VEQ	11 Apr 2019
Test 18	T22VER	4 Nov 2018
Test 19	T22VER	15 Feb 2019
Test 20	T23VNG	24 Jan 2019
Test 21	T23VPK	15 Mar 2019
Norway		
Test 22	T32VLL	21 Jan 2019
Test 23	T32VLN	10 Mar 2019
Test 24	T32VMP	5 Feb 2019
Test 25	T32VNN	11 Apr 2019

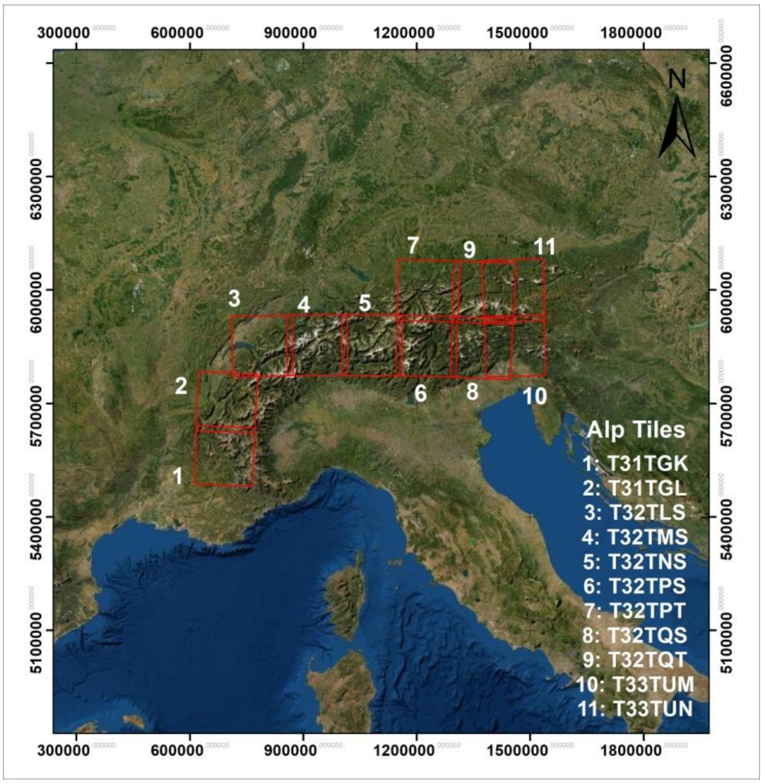


Figure 1. Sentinel 2 tile locations over Alps.

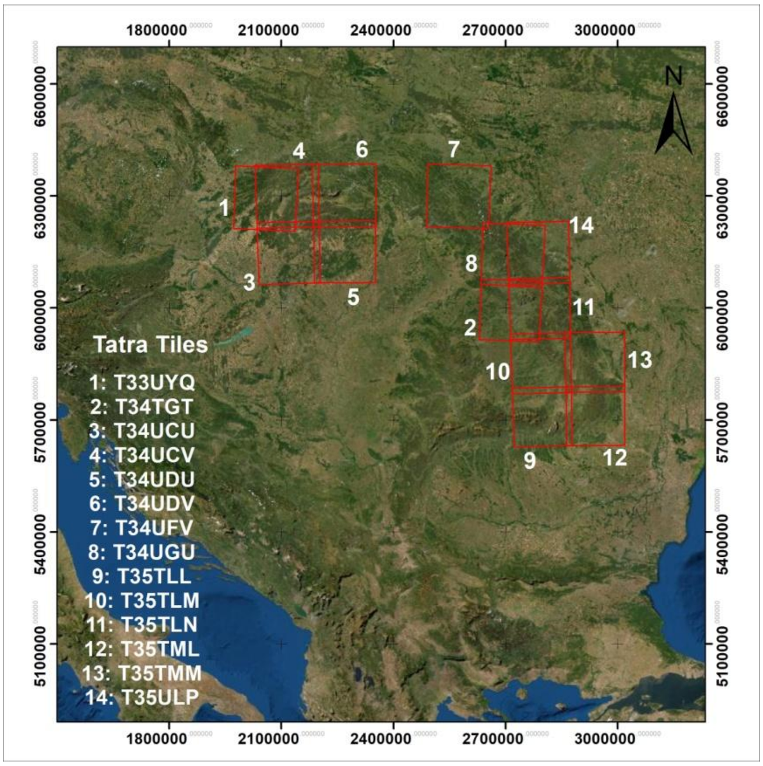


Figure 2. Sentinel 2 tile locations over Tatra.

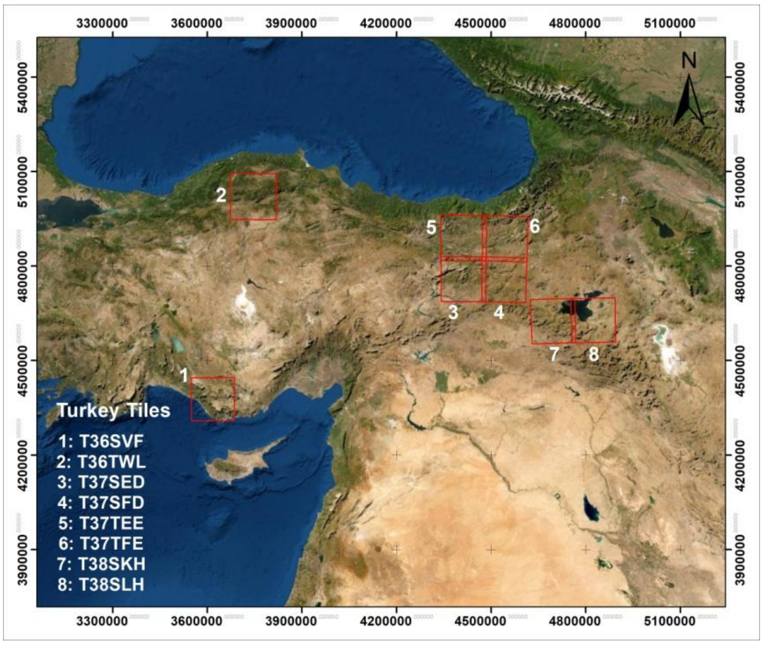


Figure 3. Sentinel 2 tile locations over Turkey.

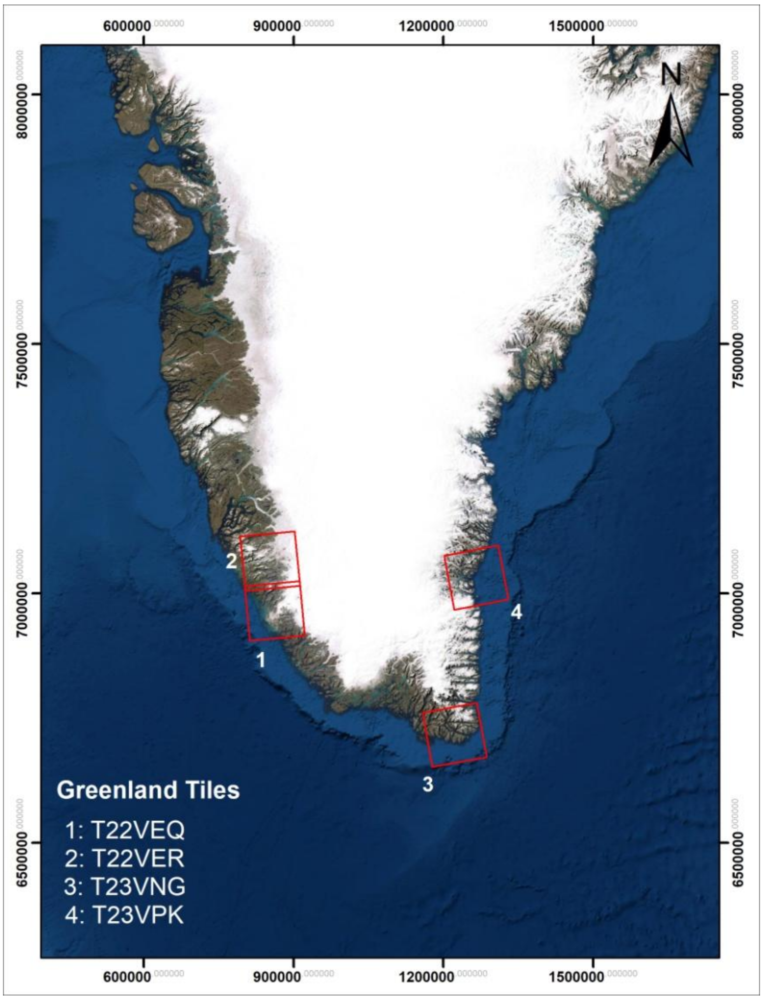


Figure 4. Sentinel 2 tile locations over Greenland.

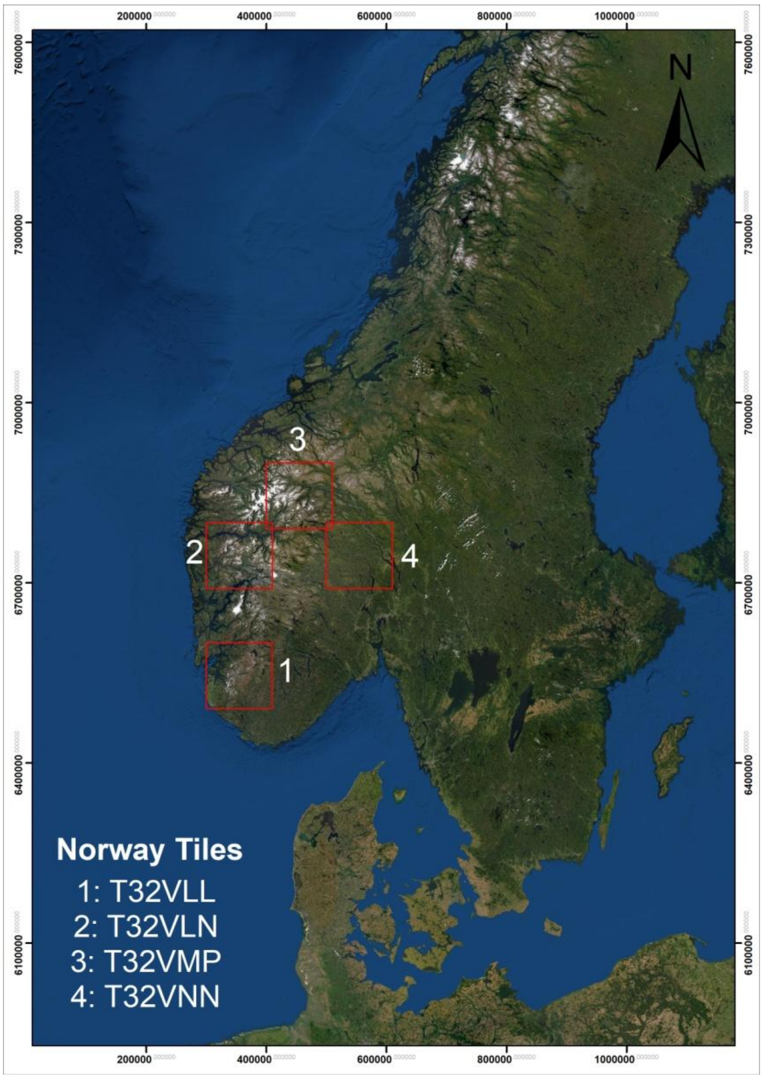


Figure 5. Sentinel 2 tile locations over Norway.

Table 4. Sen2Cor class labels and names.

Class Label	Class Name
0	No data
1	Saturated/defective
2	Dark area
3	Cloud shadows
4	Vegetation
5	Not vegetated
6	Water
7	Unclassified
8	Cloud (medium probability)
9	Cloud (high probability)
10	Thin cirrus
11	Snow

The associated Metop images are selected so that they temporally and spatially match with the corresponding Sentinel 2 images. The Metop – Sentinel 2 image pairs cover 110 individual dates spanning from Nov 2018 to Apr 2019. The Metop B / AVHRR L1B from

Global Data service (GDS) are read by using Python Satpy satellite data reader. The Satpy package is a python library for reading and manipulating meteorological remote sensing data and writing it to various image and data file formats. Satpy comes with the ability to make various RGB composites directly from satellite instrument channel data or higher level processing output. With the help of Satpy, the GDS L1B channels were calibrated accordingly. The L1B products for the AVHRR channels contain in-band radiances. By using the defined scale factors, channel radiances were converted to the reflectance and brightness temperature values, respectively. 21 Mar 2019 H35 image and the corresponding RGB image of Metop pass at 10:10 are shown in Figure 6. AVHHR band designation is given in Table 5.

Table 5. AVHRR bands.

Band No	**Wavelength Interval (µm)**
1	0.580-0.680
2	0.725-1.000
3a	1.580-1.640
3b	3.550-3.930
4	10.300-11.300
5	11.500-12.500

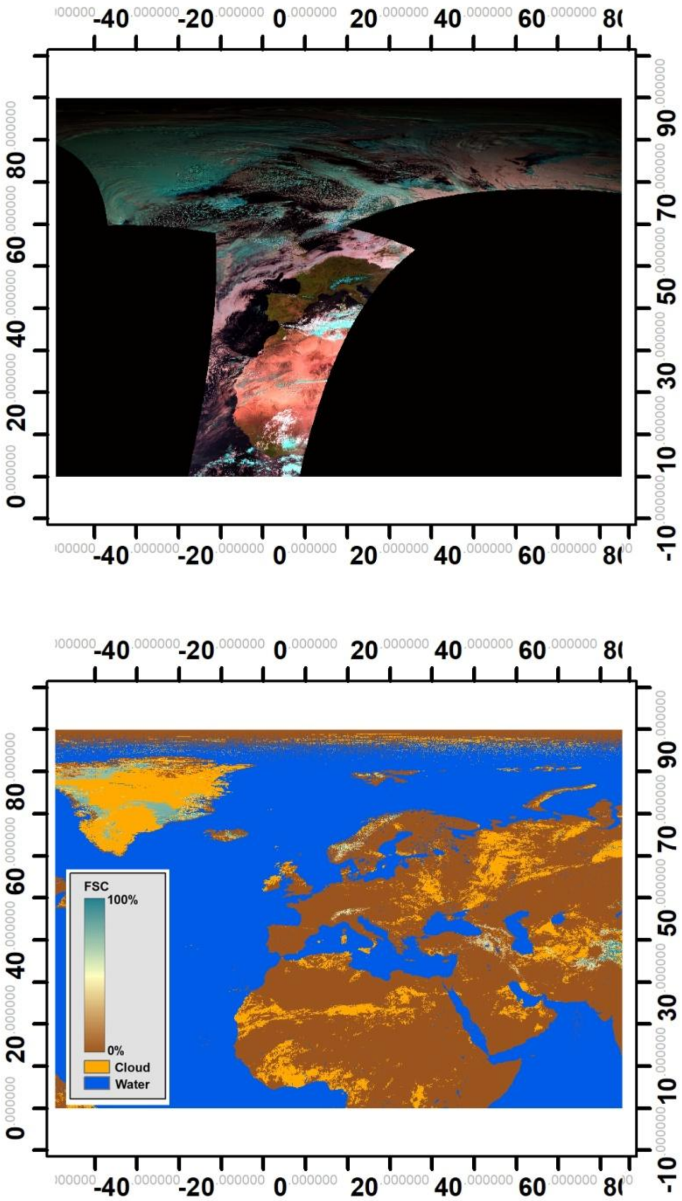


Figure 6. The lower image: H35 (TSMS mountain product) on 21 Mar 2019, and the upper image: The corresponding RGB image of Metop pass at 10:10.

2.2 Production of the Reference FSC Maps

Reference FSC snow maps are obtained by averaging the number of binary snow pixels inside the exact foot print of each H35 pixel (i.e., the number of binary snow pixels divided by the total number of pixels in each H35 pixel). Pixels labeled as cloud, water, no data etc. in the corresponding H35 product are excluded from the analysis. An additional cloud mask obtained from the Sentinel 2's scene classification images is also applied on each image in order to discard the H35 pixels with cloud cover > 10%. An example of reference FSC map is shown in Figure 7. Sentinel 2-derived reference FSC image has ~1 km spatial resolution. Figure 8 depicts the calculation of a reference FSC value for an H35 pixel from a Sentinel 2 binary snow map.

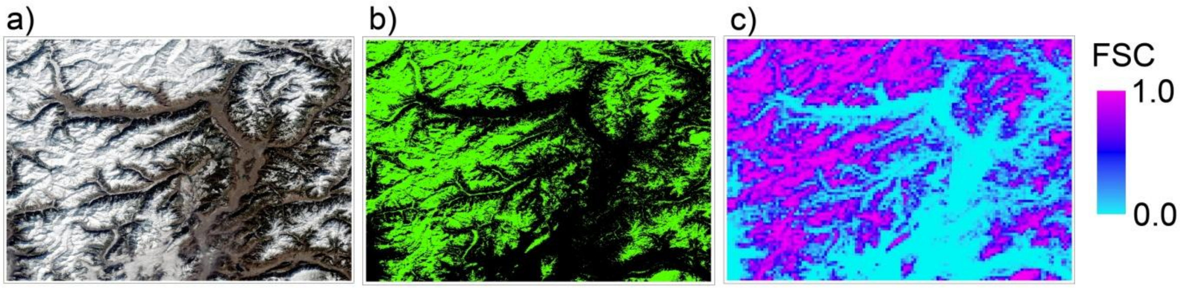


Figure 7. T32TPS Alps - Test image 4 - 23 Feb 2019: a) Sentinel 2 real color RGB image, b) Sentinel 2 binary snow map (Snow: 🟩 No snow: ⬛) and c) Sentinel 2-derived reference FSC map.

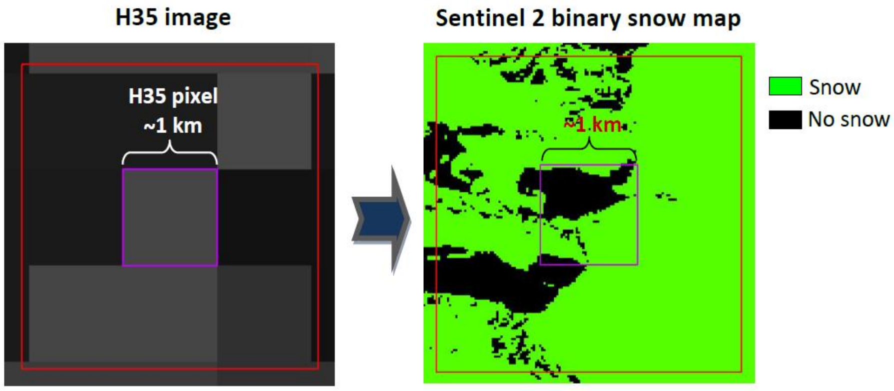


Figure 8. Reference FSC calculation in an H35 pixel from Sentinel 2 binary snow map by averaging the number of snow pixels (T32TPS Alps - Test image 4 - 23 Feb 2019).

2.3 Training Dataset

All of the available pixels from the training images (332 in total, cf. Table 2) are combined in a single dataset which is composed of 1,450,616 observations and it is used as the sampling frame. Then, a sample with size of 30% from the sampling frame is drawn by stratified random sampling. Stratification is done with respect to Sentinel 2-derived reference FSC values between 0.0 and 1.0 within 0.1 intervals in order to prevent the models being biased towards a certain FSC value. The number of sampled pixels within each 0.1 FSC interval is given in Table 6.

Table 6. Number of pixels within each 10% FSC range.

FSC interval	No. of sampled pixels	Percentage (%)
0.0-0.1	258,255	59.3
0.1-0.2	4,589	1.1
0.2-0.3	3,860	0.9
0.3-0.4	4,110	0.9
0.4-0.5	4,691	1.1
0.5-0.6	5,780	1.3
0.6-0.7	7,663	1.8
0.7-0.8	11,156	2.6
0.8-0.9	20,989	4.8
0.9-1.0	114,092	26.2
Total	435,185	100

Training data for each H35 pixel consists of a response variable, i.e., FSC value derived from higher resolution Sentinel 2 images, and 7 predictor variables, namely, top-of-atmosphere (TOA) reflectance values from AVHRR bands 1, 2, 3a, 4, 5, NDSI and NDVI. For AVHRR, NDSI and NDVI are calculated by using the following equations (for AVHRR band designation, cf. Table 6):

$$\mathrm{NDSI}_{AVHRR} = \frac{(Band1 - Band3a)}{(Band1 + Band3a)},$$
$$\mathrm{NDVI}_{AVHRR} = \frac{(Band2 - Band1)}{(Band2 + Band1)}. \tag{1}$$

70% of the sampled observations are used as training, 30% as validation dataset. The performance of the models during the training stage is evaluated by using the 30% validation dataset. The independent test dataset (cf. Table 3), which are unseen by the trained models, is employed to assess the accuracy (i.e., the generalization ability) of the models. The details of the training dataset are given in Table 7.

Table 7. Training dataset used in building of MARS and RF FSC models.

Total number of available pixels	1,450,616
Sampling Type	Stratified random sampling with respect to FSC
Sample size	30%
Number of sampled pixels	435,185
70% training set	305,338
30% validation set	129,847

2.4 Machine Learning Algorithms

2.4.1 *Multivariate Adaptive Regression Splines (MARS)*

In this subsection, the details of MARS algorithm are presented based on Friedman (1991) and Hastie et al. (2009).

In MARS, one-dimensional piecewise linear basis functions (BFs) are used to define relationships between a response variable and a set of predictors. The range of each predictor variable is cut into subsets of the full range by using knots which defines an inflection point along the range of a predictor. The slope of the linear segments between each consecutive pair of knots varies which ensures that the fully fitted function has no breaks or sudden steps. Selection of BFs is data-based and specific to the problem in MARS, which makes it a powerful adaptive regression procedure, suitable for solving high-dimensional and complex problems. The method is nonparametric, since it does not make any specific assumptions on the relations between the variables.

During the model building, BFs are fitted in such a way that additive and interactive effects of the predictors are taken into account to determine the response variable. The one-dimensional truncated piecewise linear BFs of MARS have the following form:

$$\begin{aligned} \left(x-\tau\right)_{+} &= \begin{cases} x-\tau, & \text{if } x>\tau, \\ 0, & \text{otherwise,} \end{cases} \\ \left(x-\tau\right)_{-} &= \begin{cases} \tau-x, & \text{if } x<\tau, \\ 0, & \text{otherwise,} \end{cases} \end{aligned} \tag{2}$$

where τ is a univariate knot ($x, \tau \in \mathbb{R}$). These two functions are called “reflected pair”, and the symbol “+” and “-” indicate that only the positive and the negative parts are taken, respectively, and zero otherwise.

The general model on the relation between the predictor variables and their response is defined by the following equation:

$$Y = f\left(\boldsymbol{X}\right) + \varepsilon, \tag{3}$$

where *Y* is the response variable, $\boldsymbol{X} = \left(x_1, x_2, \ldots, x_p\right)^T$ is the vector of predictors and ε is an additive stochastic component with zero mean and finite variance. The logic in MARS is to generate reflected pairs for each input $X_j\ (j=1,2,\ldots,p)$ with p-dimensional knots $\boldsymbol{\tau}_i=(\tau_{i,1},\tau_{i,2},\ldots,\tau_{i,p})^T$ at or just nearby each input data vectors $\boldsymbol{X}_i=\left(x_{i,1},x_{i,2},\ldots,x_{i,p}\right)^T$ of that input $(i=1,2,\ldots,N)$.

The set of 1-dimensional BFs of MARS can be expressed as follows:

$$C:=\left\{\left(x_j-\tau\right)_+,\ \left(x_j-\tau\right)_-,\ \middle|\ \tau\in\left\{x_{1,j},x_{2,j},\ldots,x_{N,j}\right\},\ j\in\left\{1,2,\ldots,p\right\}\right\}, \tag{4}$$

where N is the total number of observations, p is the dimension of the input space. Here, $f(\boldsymbol{X})$ in Equation (3) can be represented as a linear combination, which is successively constructed by the set C and with the intercept β_0, and it is of the following form:

$$Y=\beta_0+\sum_{m=1}^{M}\beta_m B_m\left(\boldsymbol{X}^m\right)+\varepsilon, \tag{5}$$

where B_m is a BF or product of two or more BFs from the set C, and it is taken from a set of M linearly independent BFs. Here, $\boldsymbol{X}^m$ is a subvector of $\boldsymbol{X}$ contributing to the function B_m, and β_m denotes an unknown coefficient of the mth BF, or the constant $1\left(m=0\right)$. By multiplying an existing BF with another reflected pair including another variable, a new BF is generated which represents a certain degree of interaction between different predictor variables, and both the existing BFs and the newly created BFs are included in the model. In this way, spline fitting in higher dimensions is achieved, leading to multivariate spline BFs with the following form:

$$B_m\left(\boldsymbol{X}^m\right):=\prod_{j=1}^{K_m}\left(s_{\kappa_j^m}\cdot\left(x_{\kappa_j^m}-\tau_{\kappa_j^m}\right)\right)_+, \tag{6}$$

where the total number of truncated linear functions multiplied in the mth BF is denoted by K_m, $x_{\kappa_j^m}$ indicates the input variable corresponding to the kth truncated linear function in the mth BF, $\tau_{\kappa_j^m}$ is the knot location for $x_{\kappa_j^m}$ and, finally, $s_{\kappa_j^m}\in\left\{\pm1\right\}$.

MARS uses two stages when building up a regression model, namely, the “forward pass” and the “backward pass” algorithms. In the forward pass, the algorithm chooses the knot and its corresponding pair of BFs that result in the largest decrease in residual error, and the products satisfying the above mentioned condition are successively added to the model until a predefined value M_{max} is reached.

Since the forward pass creates an over-fit model, the backward pass is applied in order to prevent the model obtained in the forward pass from over-fitting by decreasing the complexity of the model without degrading the fit to the data. Those BFs which give the smallest increase in the residual sum of squares are removed at each step during the backward pass. Consequently, a predictor variable can be completely excluded from the model unless any of its BFs has a meaningful contribution to the predictive performance of the model. This iterative procedure continues until an optimal number of effective terms is represented in the final model.

Among the sequence of models obtained from the above mentioned process, an estimated best model, $\hat{f}_\alpha$, with an optimum number of terms α, which gives the best predictive fit, is chosen through a lack-of-fit (LOF) criterion defined by generalized cross validation (GCV):

$$\mathrm{LOF}(\hat{f}_\alpha) = \mathrm{GCV}(\alpha) := \frac{\sum_{i=1}^{N} (y_i - \hat{f}_\alpha(\boldsymbol{X}_i))^2}{(1 - Q(\alpha)/N)^2}. \tag{7}$$

Here, N is the number of sample observations, $Q(\alpha) = u + dK$ with K representing the number of knots which are selected in the forward pass, u is the number of linearly independent functions in the model, and d denotes a cost for each BF optimization.

2.4.2 *Random Forests (RF)*

RF introduced by Breiman (2001) is an ensemble of classification and regression trees (CARTs) and fundamentally based on the Breiman's bagging idea (Breiman 1996). RF can be employed for either classification or regression tasks, and in both versions, the predictor variables can be either categorical or continuous (Cutler et al. 2012).

The basic functioning principles of RF can be given as follows (Boulesteix et al. 2012; Cutler et al. 2012):

- The predictor space is partitioned by the trees using a sequence of binary partitions (i.e., splits) and the root node of the tree encompasses the whole feature space.

- Each tree is a standard CART using a splitting criterion and choose the splitting predictor from a randomly selected subset of predictors (the subset is different at each split).

- Each node goes into a splitting process generating two descendant nodes, one on the left and one on the right, depending on the value of one of the predictors. The splitting process is ceased at the terminal nodes which decide the final partition of the predictor space.

- In order to partition a node into two descendant nodes, every possible split on each predictor variable is searched and the best one is chosen according to some *lack-of-fit* (LOF) criterion. For a regression case, a split is often decided by the mean squared residual at the node for which $y_1,\ldots,y_n$ are the response values:

$$\mathrm{LOF} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2, \tag{8}$$

where y_i is the estimated value at the node and the $\bar{y}$ is the mean of estimated values. For a classification case in which K class labels exist, i.e., $1,\ldots,K$, Gini index is used as the splitting criterion:

$$\mathrm{LOF} = \sum_{k \neq k'}^{K} \hat{p}_k \hat{p}_{k'}, \tag{9}$$

where $\hat{p}_k$ denotes the proportion of the kth class observations given by:

$$\hat{p}_k = \frac{1}{n}\sum_{i=1}^{n} I\left(y_i = k\right). \tag{10}$$

- Each tree is generated from a bootstrap sample drawn with replacement from the initial data set, and the predictions of all trees are finally combined through voting.

- Unweighted voting is used to combine the resulting trees if the response is categorical (i.e., classification), and unweighted averaging is used in case of a continuous response (i.e., regression).

A significant advantage of RF is the *out-of-bag* (OOB) error and it is often accepted as a good estimator of the error expected for unseen data. OOB data refers to the observations that are not involved in the construction process of some of the trees and this dataset is used internally by the algorithm as a validation set during the training process. The corresponding OOB error of an RF model is simply the mean error frequency calculated when the observations from the dataset are estimated by using the trees for which they are OOB (Boulesteix et al. 2012).

2.5 Performance Measures

Evaluation of MARS and RF models in both training and testing was accomplished by employing the following two basic statistical measures: Root-mean-square error (RMSE) and Pearson's correlation coefficient (*R*), and they are briefly explained in Table 8.

Table 8. Statistical performance measures employed in the assessment of MARS and RF FSC models.

Abbreviation	Measure	Explanation	Interpretation	Formula
RMSE	Root mean square error	Square root of the mean squared error	Smaller value is better.	$\mathrm{RMSE} := \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(y_k - \hat{y}_k\right)^2}$
R	Correlation coefficient	Linear relation between observed and predicted response	Value closer to one is better.	$R := \frac{\sum_{k=1}^{N}\left(y_k - \bar{y}\right)\left(\hat{y}_k - \bar{\hat{y}}\right) / (N-1)}{\sqrt{s(y)^2\, s(\hat{y})^2}}$
Note: *N*, number of observations; y_k, *k*th observed response value; $\hat{y}_k$, *k*th predicted (fitted) response value; $\bar{y}$, mean of observed response values; $\bar{\hat{y}}$, mean of predicted response values; $s(y)$, standard deviation of observed response values; $s(\hat{y})$, standard deviation of predicted response values.				

2.6 Initial Training of MARS and RF Models as Reported in the Interim Report

MARS model building procedure is simple and straightforward. There are two basic MARS parameters to control the "model tuning" process during the training stage. The first parameter is the "maximum allowed numbers of BFs in the forward pass" (max_BFs), and the second one is the "maximum allowed degree of interactions between predictor variables" (max_INT). Increasing max_BFs brings a rise in the amount of flexibility - therefore, complexity - of the resulting model, where increasing max_INT provides the ability to model nonlinearities and statistical dependencies between predictor variables.

Development of MARS models was performed with a basic trial-and-error procedure to decide the optimal values of these two parameters for each type of training samples. First, max_INT took the values from the set $\{1,2,3\}$, and then the value of max_BFs was incrementally varied taking within the set $\{50,100,\ldots,500\}$ for each value of max_INT.

The MARS model that gave maximum R on 30% validation data during model building was chosen for each max_INT. The details of the initial MARS models are represented in Table 9.

Table 9. Training performances of MARS FSC models.

max_INT	**max_BFs**	**Training (70%)**		**Validation (30%)**	
		RMSE	R	RMSE	R
1	50	0.1502	0.9403	0.1513	0.9396
2	50	0.1476	0.9424	0.1488	0.9416
3	100	0.1462	0.9435	0.1475	0.9426

RF algorithm uses three basic model tuning parameters which are:

- The number of the trees in the forest; N_{tree},
- The number of predictor variables to be randomly selected at each split; M_{try}, and
- The minimum number of observations per tree leaf; L_{size}.

M_{try} is generally considered as the only parameter to which RFs is somewhat sensitive (Cutler et al. 2012). The universally accepted default values of M_{try} for classification and regression in the literature are $\sqrt{p}$ and $\lceil p/3 \rceil$, respectively, where p is the dimension of the input parameter space, i.e., total number of predictor variables, and $\lceil \bullet \rceil$ indicates the ceiling operator. However, as shown in (Díaz-Uriarte and De Andres 2006), the overfitting effects due to the choice of M_{try} on the performance of the method are quite small and negligible. These default values of M_{try} are often employed in RS applications (Belgiu and Drăguţ 2016; Houborg and McCabe 2018).

The variance of an RF model decreases with increasing N_{tree} values. As shown in Breiman (2001), the generalization error of RFs eventually converges to a limit as N_{tree}

increases, which means that large choices of N_{tree} do not result in overfitting (Biau and Scornet 2016). In RS literature, wide range of N_{tree} values has been applied, 50 to 5000 (Colditz 2015; Guan et al. 2013; Houborg and McCabe 2018; Nitze et al. 2015; Pelletier et al. 2016); however, $N_{\text{tree}} = 500$ is accepted as a reasonable choice since the stability in the generalization error is often realized before reaching this value (Lawrence et al. 2006).

In this study, the default values for M_{try} and L_{size} for regression were used. L_{size} was taken as 5. M_{try} was set to 3 since the models were trained by using 7 predictor variables (i.e., AVHRR bands 1, 2, 3a, 4, 5, NDSI and NDVI). The N_{tree} took values between 200 and 2000 with 200 increments. The RF model with the best R value on 30% validation dataset during model building was chosen and applied on the independent test dataset. The training performance of the initial RF FSC model is given in Table 10.

Table 10. Training performance of RF FSC model.

N_{tree}	**Training (70%)**		**Validation (30%)**	
	RMSE	R	RMSE	R
1200	0.0938	0.9774	0.1401	0.9484

2.7 The Final MARS Model

It is decided to continue with MARS modelling approach for further model refinement and optimization purposes towards a potential operational product generation due to the following basic facts:

- The initial model assessment metrics on the test dataset indicate that the MARS and RF models perform almost equally.

- RF models are generated under Matlab. The final RF model (cf. Table 10) has a closed form which directly imposes that the FSC estimation on the new dataset has to be run under Matlab. On the other hand, MARS output is a simple regression-like equation which can be easily implemented in any programming environment.

- The required processing time for the estimation of FSC values in RF model under Matlab is far beyond the acceptable limits for a potential operational product generation.

The MARS model is further simplified by reducing the number of BFs to 10 by keeping the degree of interaction among the predictor variables as 1, i.e., additive model. This simplification is strictly needed to constrain the processing time of the new FSC product within required duration. The training performance of the final MARS model is given in Table 11.

Table 11. Training performance of the final MARS FSC model.

max_INT	**max_BFs**	**Training (70%)**		**Validation (30%)**	
		RMSE	*R*	RMSE	*R*
1	10	0.1630	0.9285	0.1648	0.9279

3. RESULTS

3.1 Initial MARS and RF FSC Models on the Independent Test Dataset

The results obtained from the assessment of initial MARS and RF FSC models on the independent test dataset including the first 16 test scenes over Alps Tatra and Turkey (cf. Table 3), as reported in the interim report, are given in Table 12. Since the MARS models with max_INT = 1, 2 and 3 exhibit almost the same performance on all test images, only the statistical metrics of the MARS model with max_INT = 1 and max_BF = 50 (cf. Table 9) are introduced.

Table 12. Performance metrics of initial MARS and RF FSC models together with H35 on the independent test dataset (Please note that H35 on Test 5 gives NaN for *R* since all H35 pixels are equal to 0).

Test Set		**MARS** **max_INT = 1** **max_BF = 50**		**RF** **N_{tree} = 1200**		**H35**	
		RMSE	*R*	RMSE	*R*	RMSE	*R*
Alps	**Test 1**	0.3578	0.4238	0.3378	0.4128	0.3363	0.5866
	Test 2	0.2428	0.8103	0.2399	0.8085	0.3567	0.1489
	Test 3	0.2474	0.8289	0.2449	0.8279	0.2887	0.7190
	Test 4	0.2890	0.6658	0.2834	0.6733	0.4755	0.6538
	Test 5	0.3384	0.2967	0.3339	0.3246	0.4311	NaN
	Test 6	0.2162	0.8492	0.2101	0.8623	0.4154	0.7685
Tatra	**Test 7**	0.2832	0.6781	0.3079	0.4788	0.0827	0.1813
	Test 8	0.1900	0.8180	0.1711	0.8411	0.2495	0.7814
	Test 9	0.2618	0.7800	0.2559	0.7617	0.5823	0.3577
	Test 10	0.1513	0.7811	0.1359	0.8253	0.1532	0.8549
Turkey	**Test 11**	0.1383	0.9378	0.1336	0.9427	0.4759	0.9088
	Test 12	0.1992	0.7964	0.1813	0.8189	0.3080	0.8704
	Test 13	0.1482	0.8705	0.1522	0.8620	0.3471	0.5413
	Test 14	0.3206	0.6208	0.3397	0.5919	0.2940	0.4501
	Test 15	0.1946	0.8988	0.1991	0.8851	0.2787	0.6289
	Test 16	0.2553	0.8075	0.2555	0.8088	0.3577	0.7500

Low RMSE and relatively high *R* values indicate that the MARS- and RF-derived FSC maps are in good agreement with Sentinel 2-based reference data. The scatter plots of the generated MARS and RF FSC models versus reference Sentinel 2-derived FSC values on the 16 independent test images for Alps, Tatra and Turkey are shown in Figures 9, 10 and 11, respectively. In these scatter plot diagrams, the mean value of the FSC estimated by the corresponding model within each incremental 10% FSC range is plotted against the mean of Sentinel 2-based reference FSC within that 10% FSC range.

As an overall conclusion drawn from the scatter plots; *i*) initial MARS and RF FSC models overestimate the FSC ranges from 0.0 to 0.6, and *ii*) a slight underestimation takes place, in general, for FSC ranges between 0.7 and 1.0.

In Alps Test 1 and Test 6 images, relatively lower *R* values are obtained, i.e., 0.413 and 0.325, respectively. Our visual analysis reveals that the reason for these low *R* values is due to the underestimation of snow cover in the corresponding Sentinel 2 scene classification images.

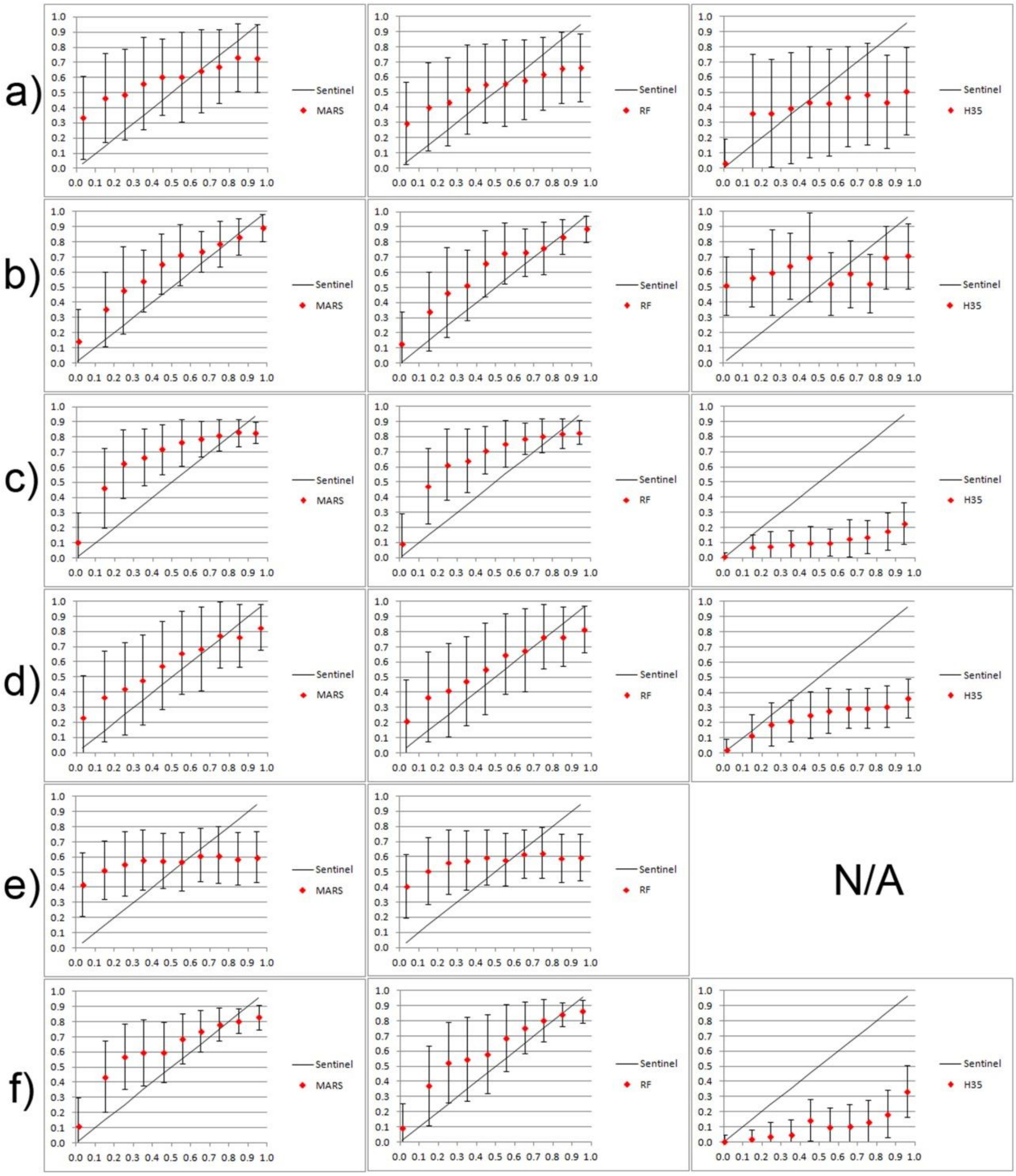


Figure 9. Scatter plots on the independent test images for Alps: a) Test 1, b) Test 2, c) Test 3, d) Test 4, e) Test 5 and f) Test 6 (Please note that the red dots indicate the mean estimated FSC by the model within each 10% FSC range and the vertical error bars indicate the corresponding standard deviations with respect to Sentinel 2-derived reference FSC values).

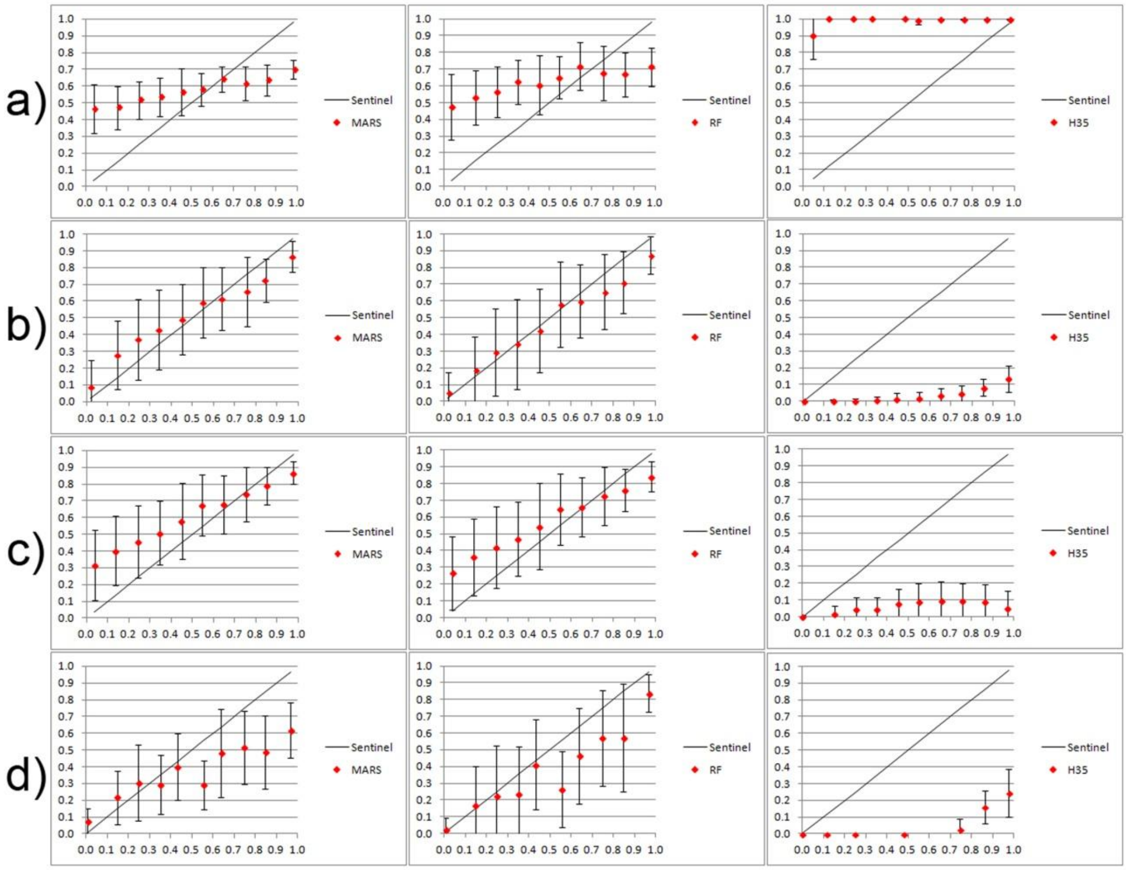


Figure 10. Scatter plots on the independent test images for Tatra: a) Test 7, b) Test 8, c) Test 9 and d) Test 10 (Please note that the red dots indicate the mean estimated FSC by the model within each 10% FSC range and the vertical error bars indicate the corresponding standard deviations with respect to Sentinel 2-derived reference FSC values).

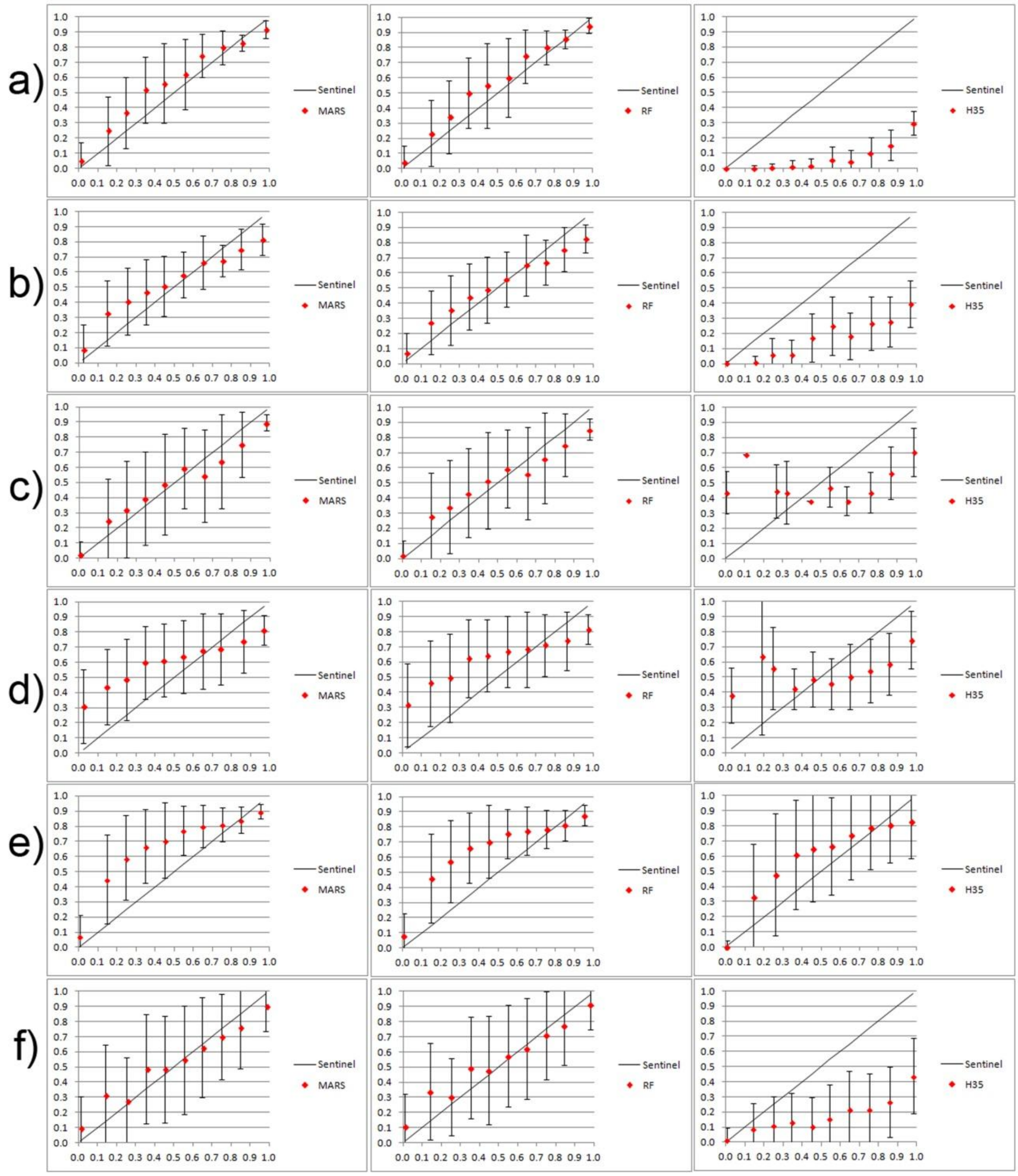


Figure 11. Scatter plots on the independent test images for Turkey: a) Test 11, b) Test 12, c) Test 13, d) Test 14, e) Test 15 and f) Test 16 (Please note that the red dots indicate the mean estimated FSC by the model within each 10% FSC range and the vertical error bars indicate the corresponding standard deviations with respect to Sentinel 2-derived reference FSC values).

The BFs and their coefficients of the initial additive MARS model with 50 BFs in the forward pass is given in Table 13.

Table 13. BFs and their associated coefficients in the additive MARS FSC model. The max_BFs in the forward pass is set as 50 and after the backward pass the remaining number of terms in the final model is 28 including the intercept (h denotes MARS hinge function, cf. Equation 2).

BFs	**Coefficient**
Intercept	-7.6720
h(Band_1-10.03)	-0.0282
h(19.32-Band_1)	-0.0525
h(Band_1-19.32)	0.0544
h(Band_1-47.11)	-0.0054
h(Band_2-5.01)	-0.0636
h(11.6-Band_2)	-0.0596
h(Band_2-11.6)	0.0573
h(Band_2-21.18)	-0.0112
h(Band_2-31.68)	-0.0026
h(Band_3a-0.58)	0.4118
h(17.4-Band_3a)	0.4284
h(Band_3a-17.4)	-0.4282
h(Band_4-255.06)	-0.0260
h(Band_4-279.94)	0.0055
h(Band_4-286.9)	0.0019
h(Band_5-219.19)	0.0356
h(265.54-Band_5)	0.0086
h(Band_5-265.54)	-0.0158
h(NDSI+0.140413)	1.2999
h(NDSI-0.248623)	-0.5772
h(NDSI-0.474725)	-0.8923
h(0.617874-NDSI)	0.4572
h(NDSI-0.617874)	-0.3463
h(NDVI+0.102639)	1.4271
h(NDVI-0.00170675)	-0.6935
h(NDVI-0.03108)	-1.4715
h(NDVI-0.180133)	0.9159

MARS and RF FSC maps of Alps test images 2 and 3; Tatra test images 7 and 9; Turkey test images 12 and 16 are shown in Figures 12-17, respectively. These images reveal that the MARS and RF FSC models are highly capable to capture the spatial pattern and the variability of the snow cover.

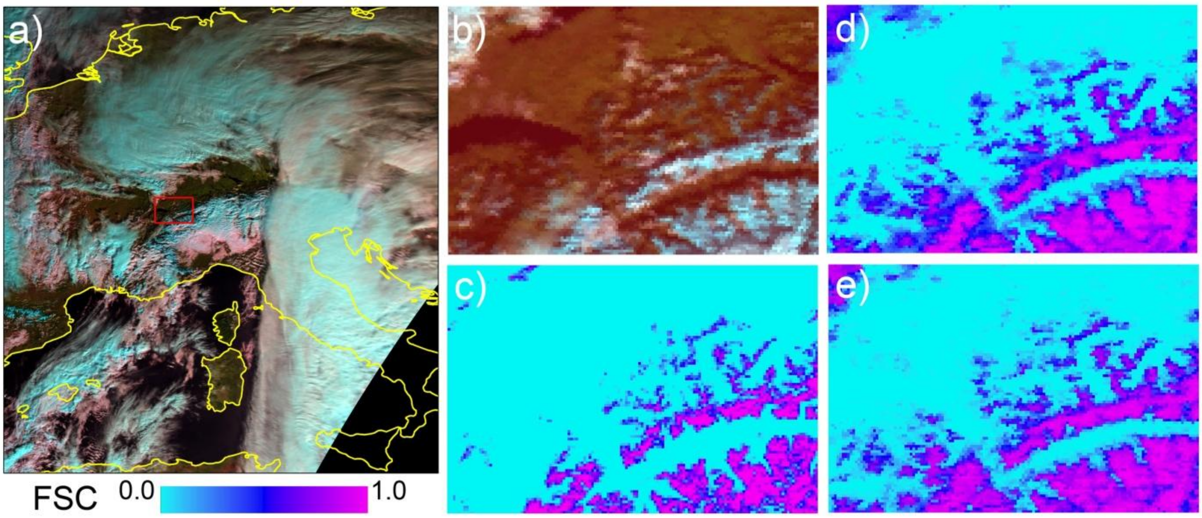


Figure 10. Alps Test Image 2: a) Metop RGB image (R: Band 3a, G: Band 2, B: Band 1), b) Metop RGB image of the test area, c) Sentinel 2-derived reference FSC map, d) MARS FSC image, and e) RF FSC image.

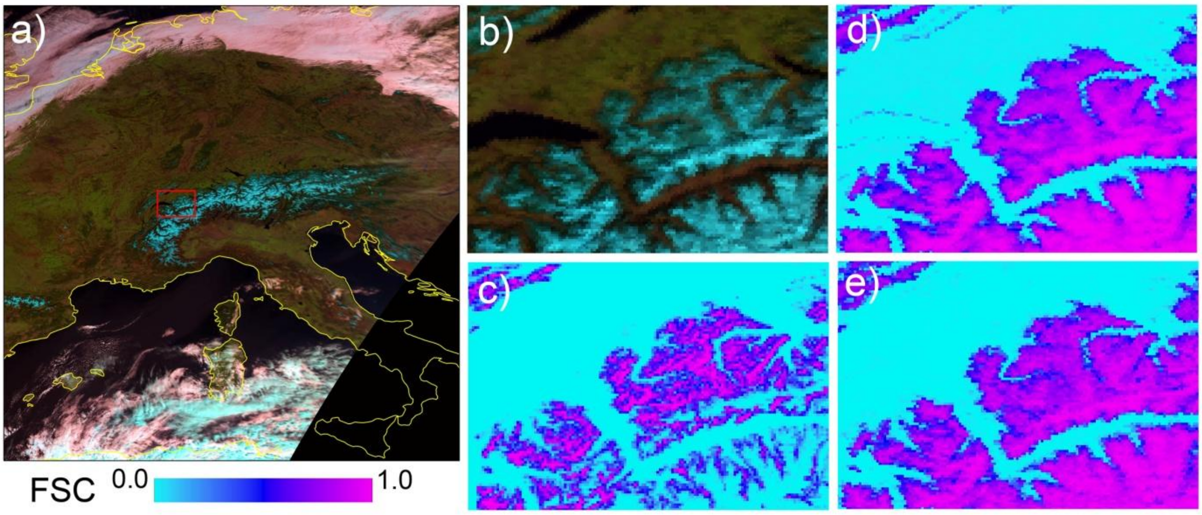


Figure 11. Alps Test Image 3: a) Metop RGB image (R: Band 3a, G: Band 2, B: Band 1), b) Metop RGB image of the test area, c) Sentinel 2-derived reference FSC map, d) MARS FSC image, and e) RF FSC image.

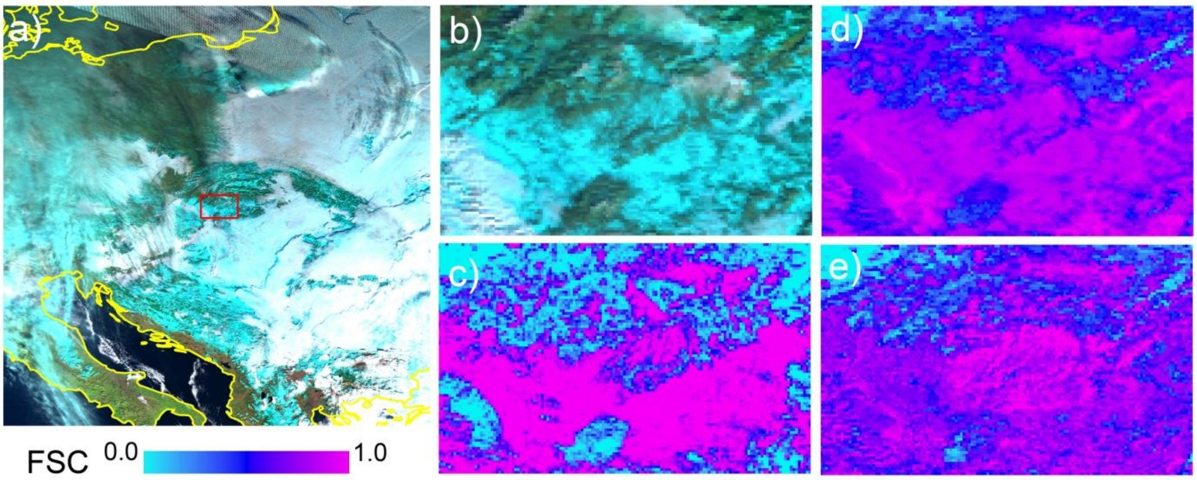


Figure 12. Tatra Test Image 7: a) Metop RGB image (R: Band 3a, G: Band 2, B: Band 1), b) Metop RGB image of the test area, c) Sentinel 2-derived reference FSC map, d) MARS FSC image, and e) RF FSC image.

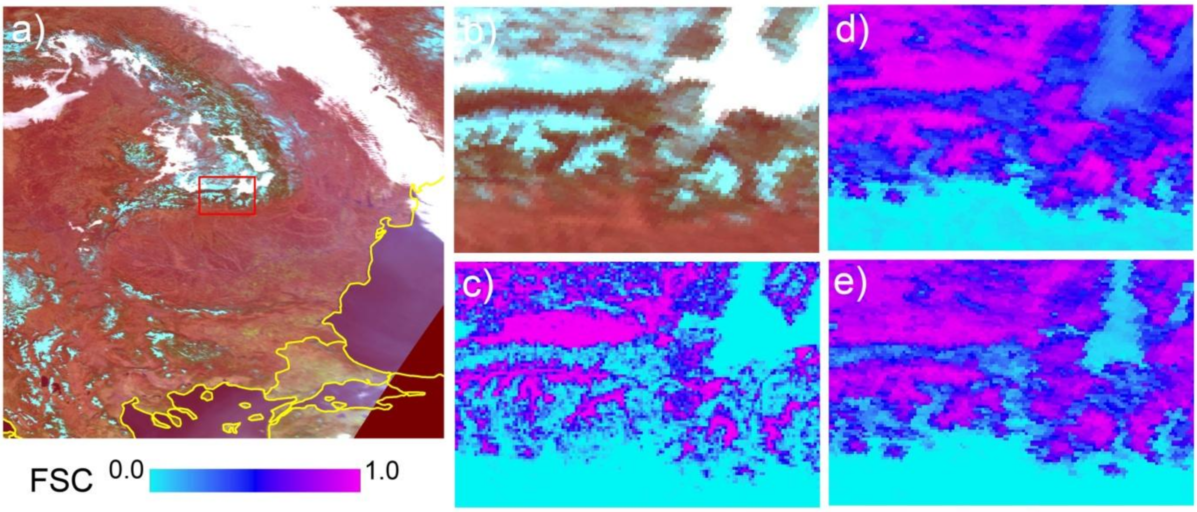


Figure 13. Tatra Test Image 10: a) Metop RGB image (R: Band 3a, G: Band 2, B: Band 1), b) Metop RGB image of the test area, c) Sentinel 2-derived reference FSC map, d) MARS FSC image, and e) RF FSC image.

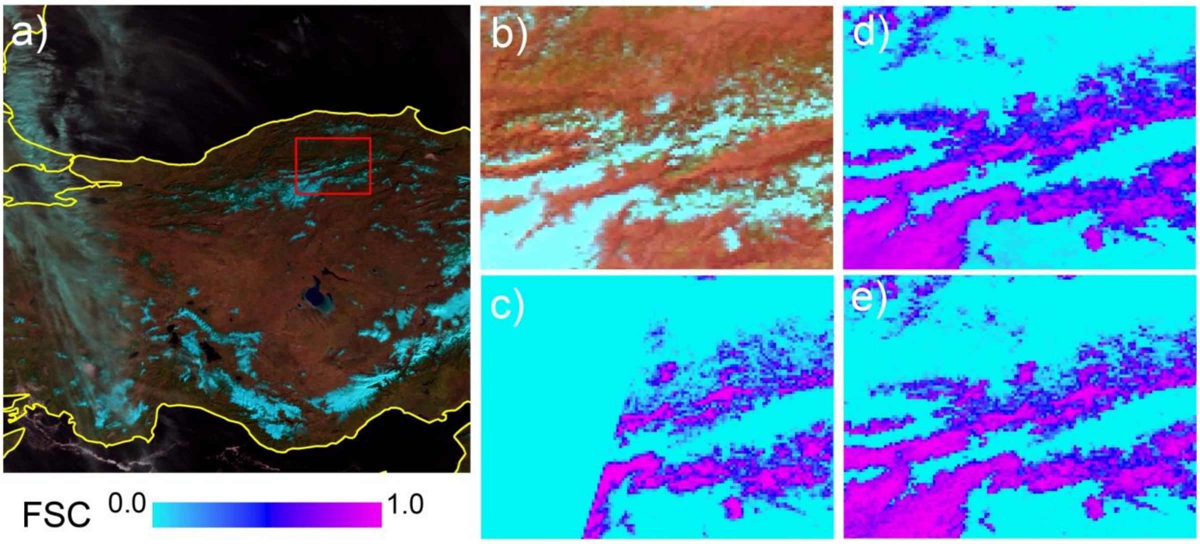


Figure 14. Turkey Test Image 12: a) Metop RGB image (R: Band 3a, G: Band 2, B: Band 1), b) Metop RGB image of the test area, c) Sentinel 2-derived reference FSC map (please note that the Sentinel 2 tile has no data in the left side of the scene), d) MARS FSC image, and e) RF FSC image.

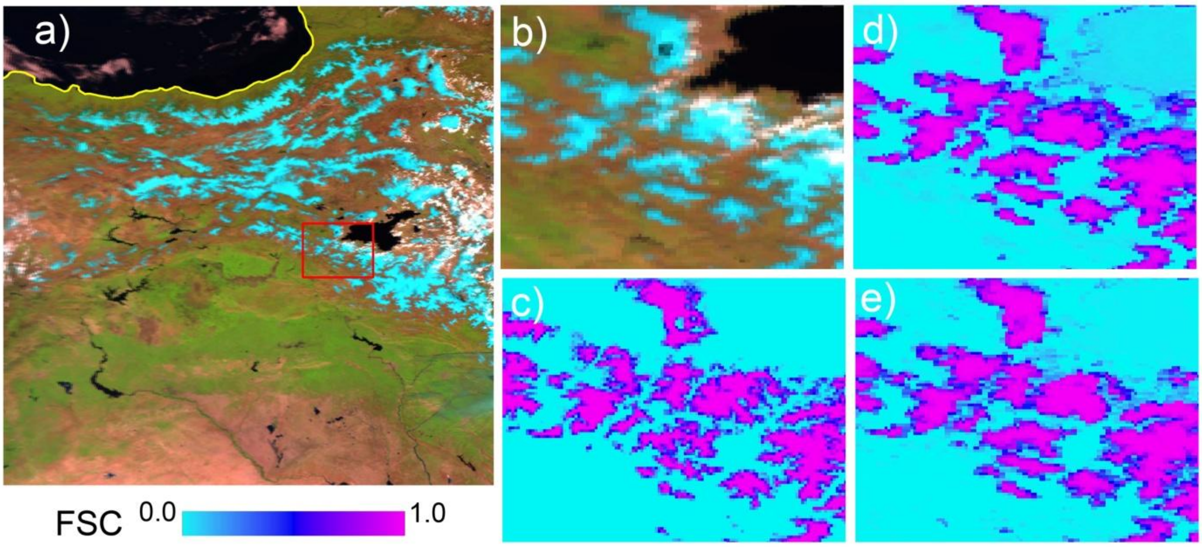


Figure 15. Turkey Test Image 16: a) Metop RGB image (R: Band 3a, G: Band 2, B: Band 1), b) Metop RGB image of the test area, c) Sentinel 2-derived reference FSC map, d) MARS FSC image, and e) RF FSC image.

3.2 Final MARS Model on the Independent Test Dataset

The performance of the final MARS model (max_BFs = 10 and max_INT = 1) is evaluated on 25 test scenes and the results are given in Table 14.

Table 14. Performance metrics of the final MARS model together with H35 on the independent test dataset (Please note that Test 5 and Test 20 on H35 gives NaN for *R* since all available H35 pixels are equal to 0. Additionally, Test 18 and Test 24 give NaN for Both RMSE and *R* since there is no available H35 pixel).

Test Set		**MARS** **max_INT = 1** **max_BF = 10**		**H35**	
		RMSE	*R*	RMSE	*R*
Alps	**Test 1**	0.3420	0.4107	0.3363	0.5866
	Test 2	0.2278	0.8144	0.3567	0.1489
	Test 3	0.2480	0.8488	0.2887	0.7190
	Test 4	0.2913	0.6539	0.4755	0.6538
	Test 5	0.3441	0.3249	0.4311	NaN
	Test 6	0.2149	0.8679	0.4154	0.7685
Tatra	**Test 7**	0.3189	0.6320	0.0827	0.1813
	Test 8	0.1887	0.8282	0.2495	0.7814
	Test 9	0.2624	0.7656	0.5823	0.3577
	Test 10	0.1922	0.5900	0.1532	0.8549
Turkey	**Test 11**	0.1351	0.9408	0.4759	0.9088
	Test 12	0.1981	0.7927	0.3080	0.8704
	Test 13	0.1329	0.8875	0.3471	0.5413
	Test 14	0.2854	0.6663	0.2940	0.4501
	Test 15	0.1737	0.9136	0.2787	0.6289
	Test 16	0.2558	0.8051	0.3577	0.7500
Greenland	**Test 17**	0.1571	0.8082	0.7131	0.3206
	Test 18	0.3395	0.5575	NaN	NaN
	Test 19	0.3163	0.3141	0.5390	0.5008
	Test 20	0.3029	0.4238	0.5081	NaN
	Test 21	0.1634	0.6161	0.7667	0.0772
Norway	**Test 22**	0.2021	0.5595	0.5975	0.1466
	Test 23	0.1733	0.8089	0.5580	0.3957
	Test 24	0.2951	0.5267	NaN	NaN
	Test 25	0.1641	0.6518	0.7903	0.3416

The corresponding scatter plots of the final MARS model for Alps, Tatra, Turkey, Greenland and Norway are presented in Figures 16-20, respectively. In these scatter plot diagrams, the mean value of the FSC estimated by the corresponding model within each incremental 10% FSC range is plotted against the mean of Sentinel 2-based reference FSC within that 10% FSC range.

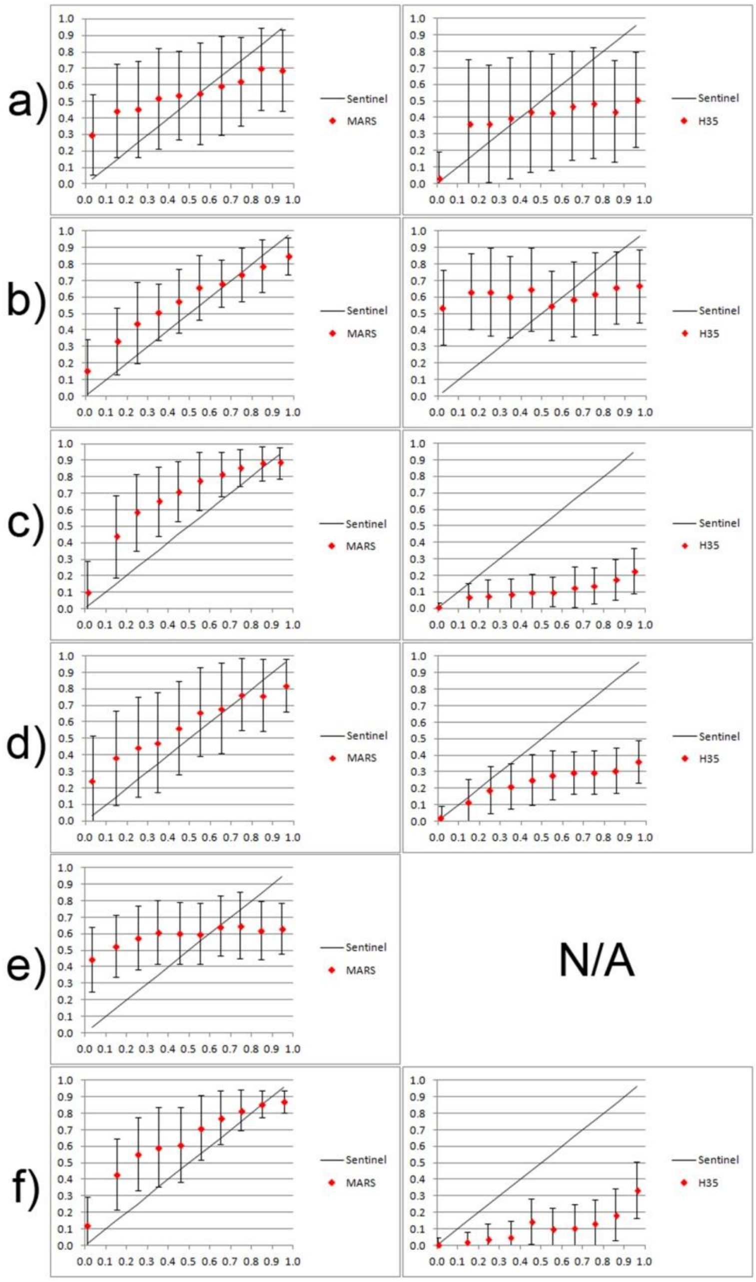


Figure 16. Scatter plots on the independent test images for Alps: a) Test 1, b) Test 2, c) Test 3, d) Test 4, e) Test 5 and f) Test 6 (Please note that the red dots indicate the mean estimated FSC by the model within each 10% FSC range and the vertical error bars indicate the corresponding standard deviations with respect to Sentinel 2-derived reference FSC values).

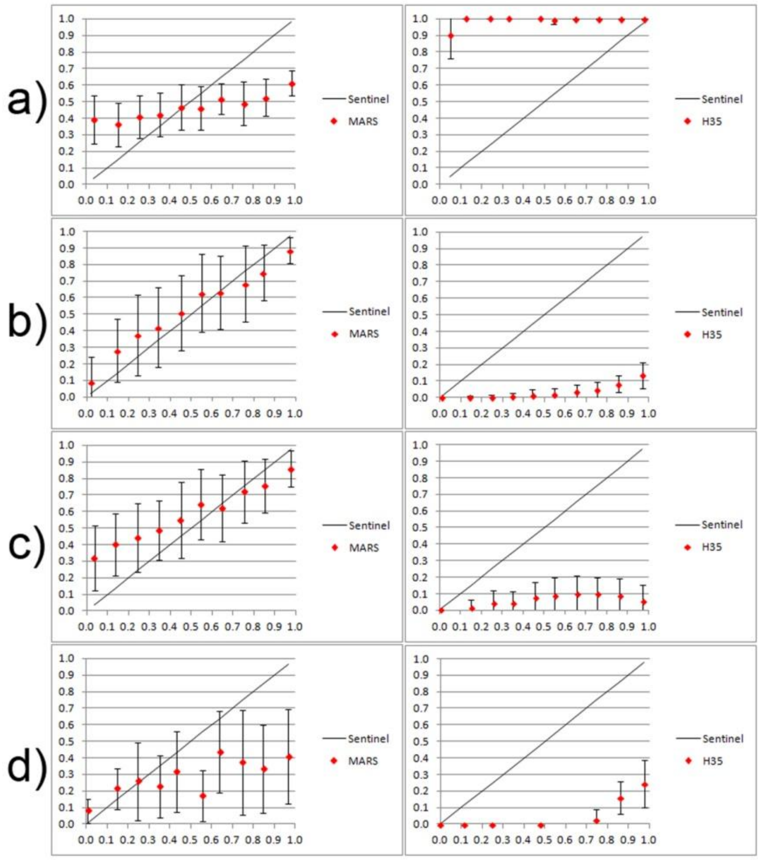


Figure 17. Scatter plots on the independent test images for Tatra: a) Test 7, b) Test 8, c) Test 9 and d) Test 10 (Please note that the red dots indicate the mean estimated FSC by the model within each 10% FSC range and the vertical error bars indicate the corresponding standard deviations with respect to Sentinel 2-derived reference FSC values).

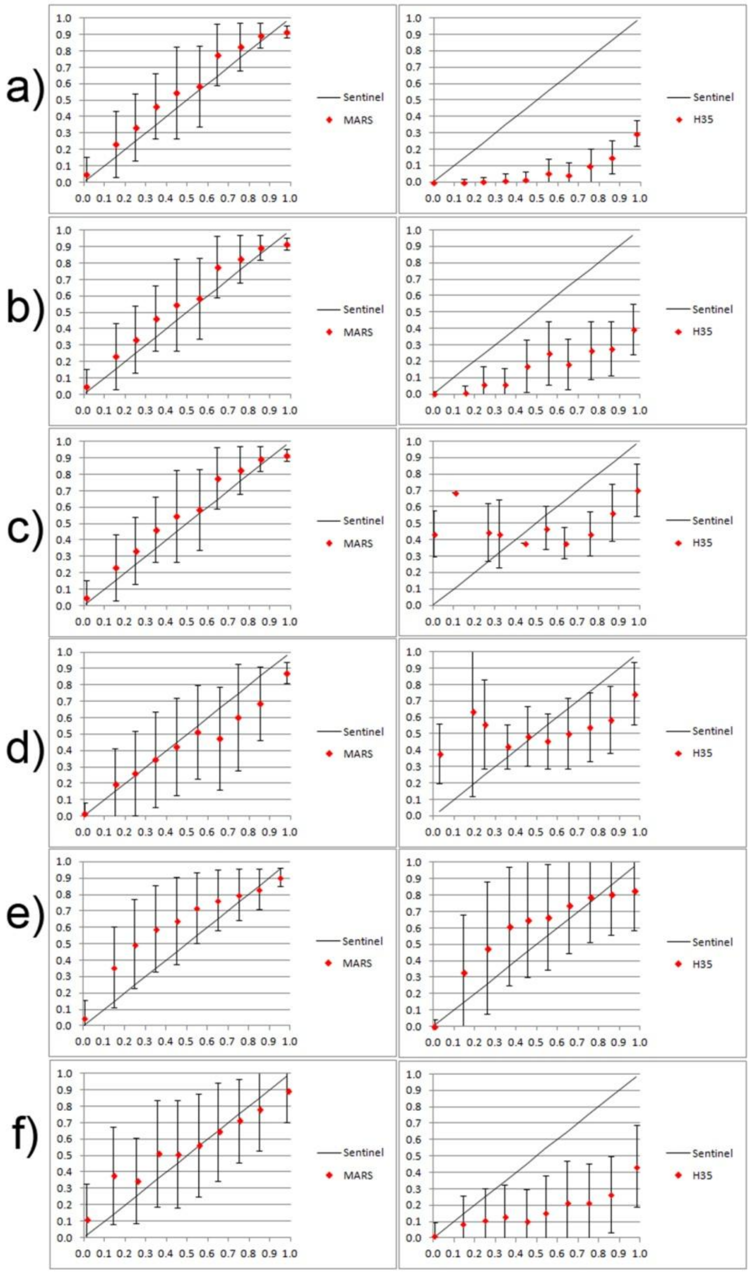


Figure 18. Scatter plots on the independent test images for Turkey: a) Test 11, b) Test 12, c) Test 13, d) Test 14, e) Test 15 and f) Test 16 (Please note that the red dots indicate the mean estimated FSC by the model within each 10% FSC range and the vertical error bars indicate the corresponding standard deviations with respect to Sentinel 2-derived reference FSC values).

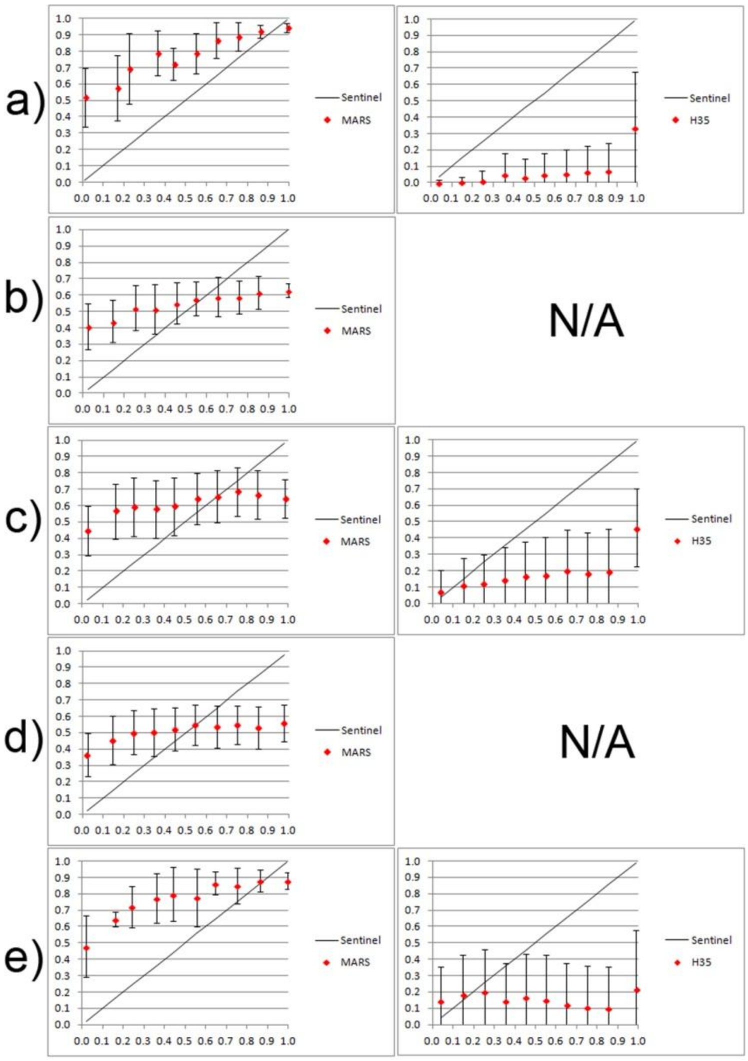


Figure 19. Scatter plots on the independent test images for Greenland: a) Test 17, b) Test 18, c) Test 19, d) Test 20 and e) Test 21 (Please note that the red dots indicate the mean estimated FSC by the model within each 10% FSC range and the vertical error bars indicate the corresponding standard deviations with respect to Sentinel 2-derived reference FSC values).

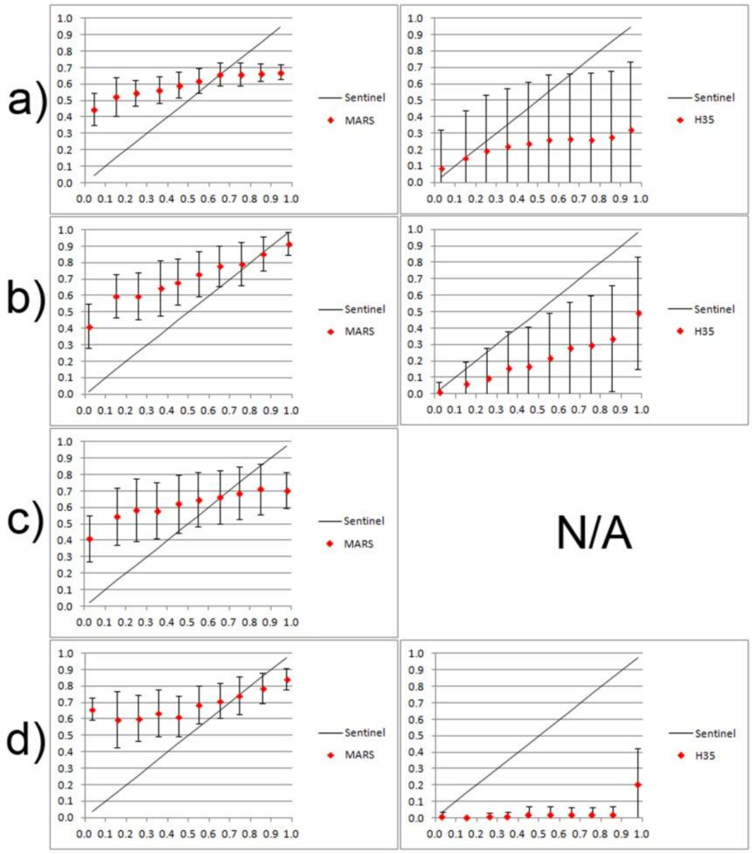


Figure 20. Scatter plots on the independent test images for Norway: a) Test 22, b) Test 23, c) Test 24 and d) Test 25 (Please note that the red dots indicate the mean estimated FSC by the model within each 10% FSC range and the vertical error bars indicate the corresponding standard deviations with respect to Sentinel 2-derived reference FSC values).

The final MARS model's BFs with the corresponding coefficients are given in Table 15.

Table 15. BFs and their associated coefficients in the final additive MARS FSC model. The max_BFs in the forward pass is set as 10 and after the backward pass the remaining number of terms in the final model is 8 including the intercept ("h" denotes MARS hinge function, cf. Equation 2).

BFs	Coefficient
Intercept	0.1069
h(19.32 - Band_1)	0.0296
h(Band_1 - 19.32)	0.0047
h(17.4 - Band_3a)	0.0098
h(Band_3a - 17.4)	0.0201
h(NDSI + 0.229783)	0.8436
h(0.617874 - NDSI)	0.2590
h(NDSI - 0.617874)	-1.4265

It is of value to mention a problem encountered regarding the Sen2Cor scene processing software during the analysis. On some of test scenes, it is observed that Sen2Cor underestimates the snow cover by misclassifying snow on the shady parts of the slopes as water which leads to lower *R* values on the test image dataset. To give a specific example, close up spatial subsets on Test scenes 1, 5, 19 and 24 are given in Figures 21-24.

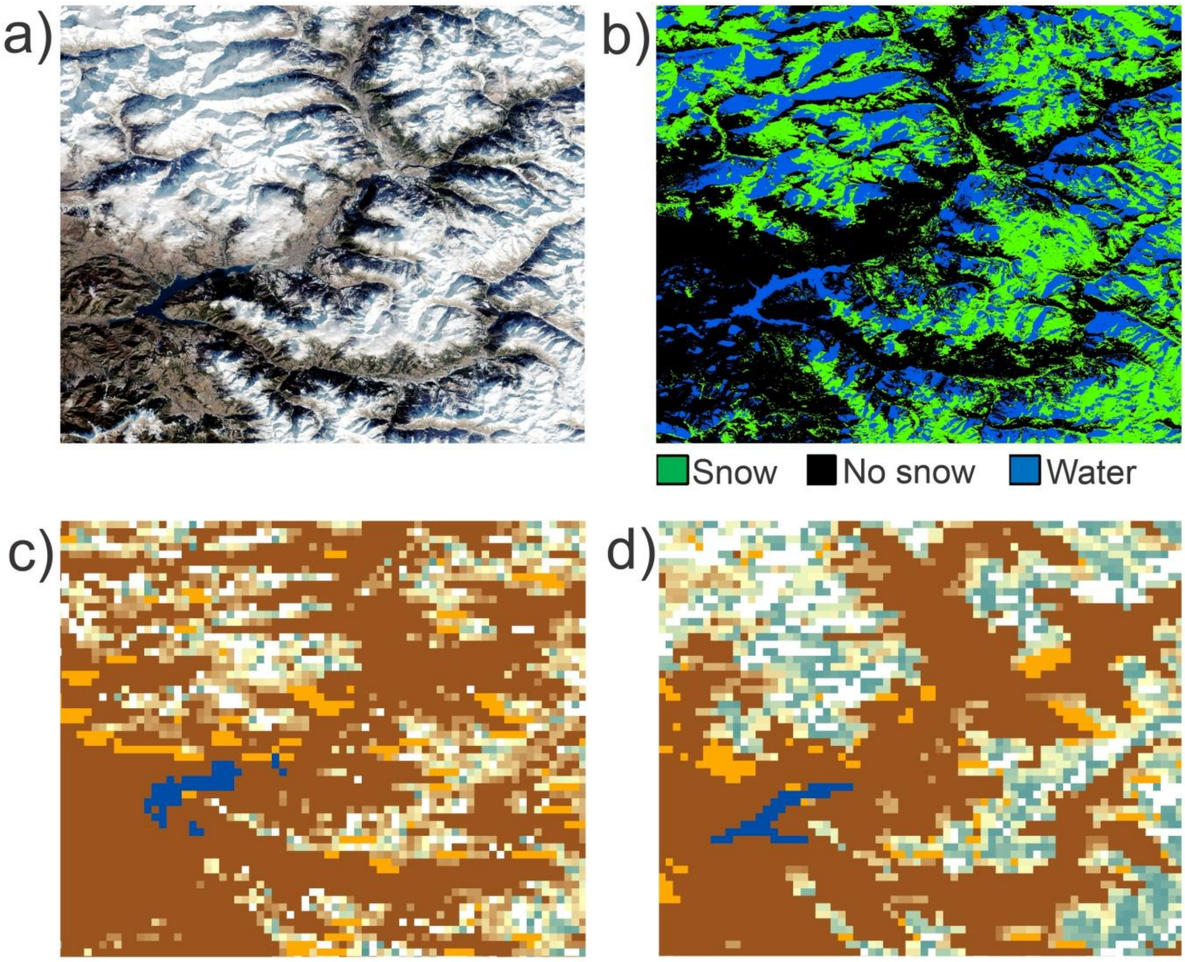


Figure 21. 20 Jan 2019 Test 1 scene close-up spatial subsections (Alps) a) Sentinel 2 RGB image, b) Sen2Cor classified image, c) original H35 image and d) MARS-based H35 image.

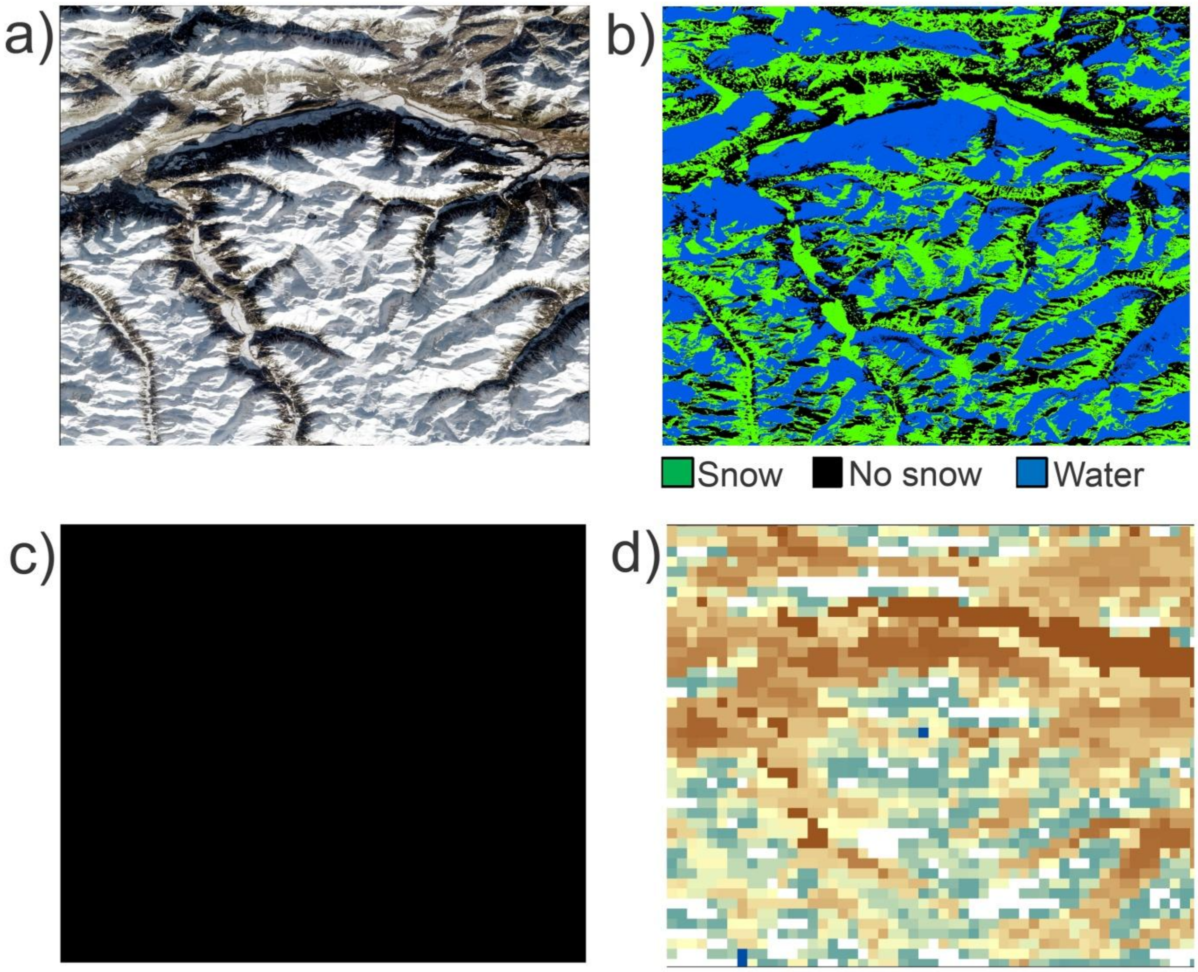


Figure 22. 18 Dec 2018 Test 5 scene close-up spatial subsections (Alps) a) Sentinel 2 RGB image, b) Sen2Cor classified image, c) original H35 image (In this spatial subset no data come from H35) and d) MARS-based H35 image.

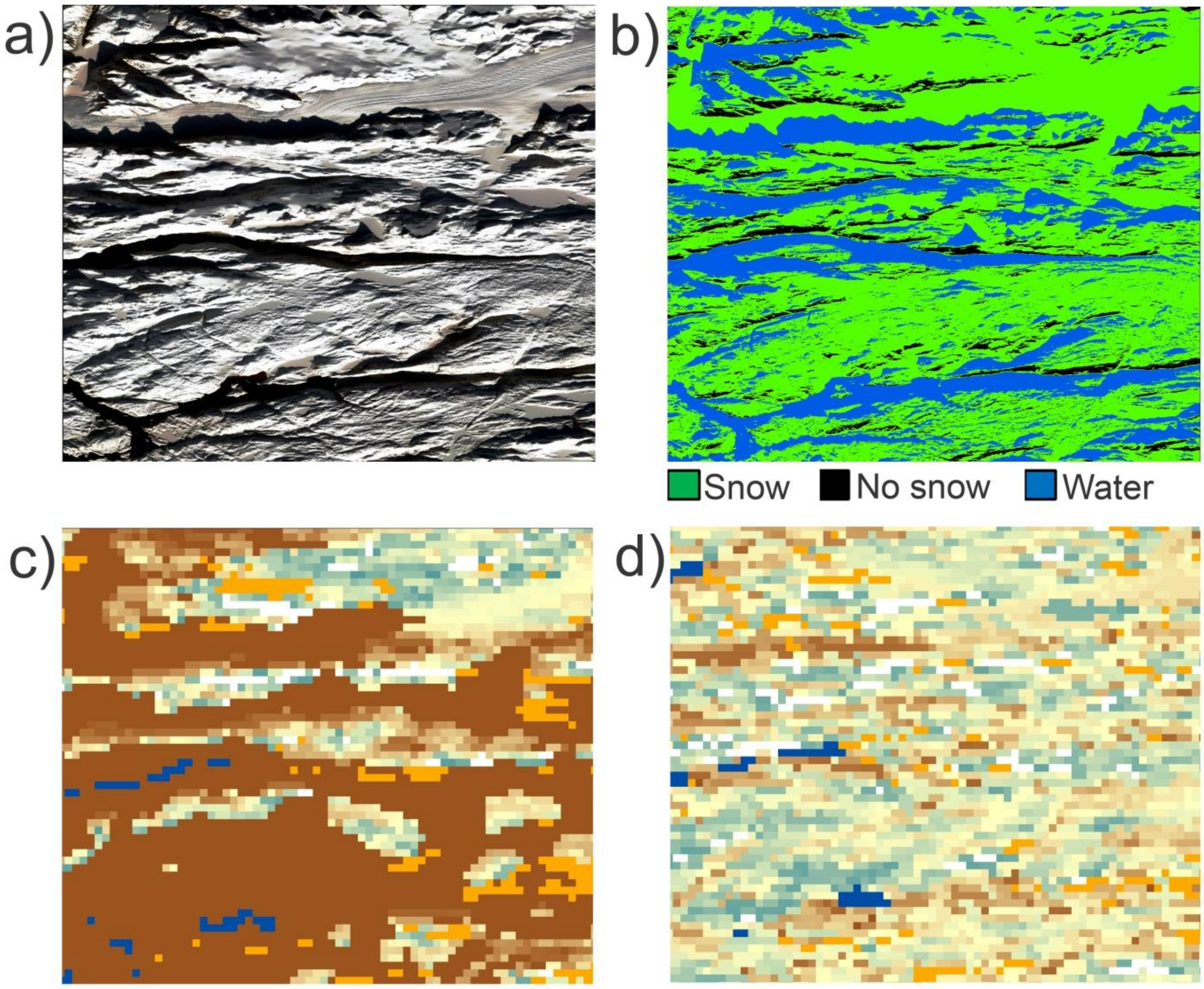


Figure 23. 15 Feb 2019 Test 19 scene close-up spatial subsections (Greenland) a) Sentinel 2 RGB image, b) Sen2Cor classified image, c) original H35 image (In this spatial subset no data come from H35) and d) MARS-based H35 image.

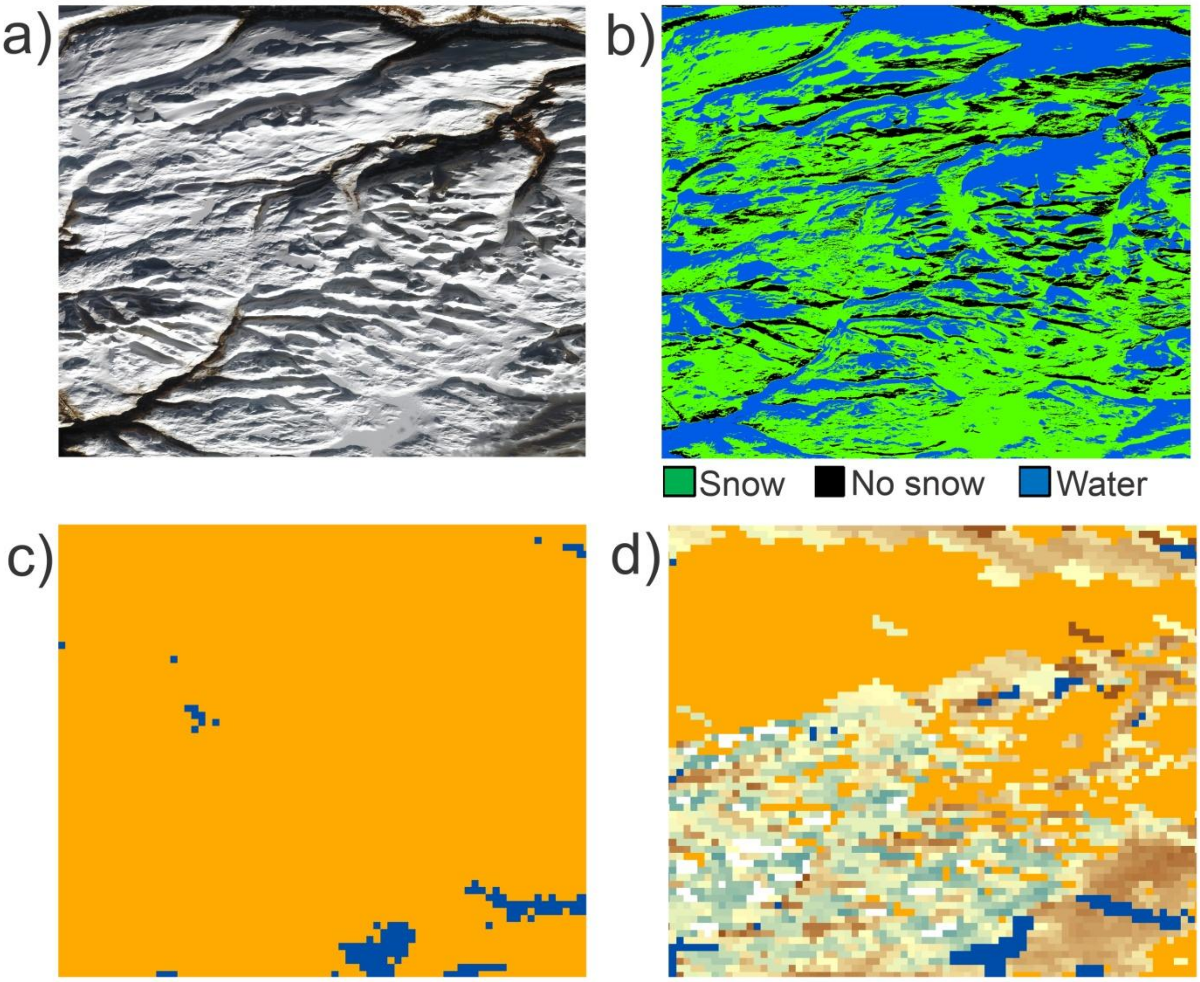


Figure 24. 5 Feb 2019 Test 24 scene close-up spatial subsections (Norway) a) Sentinel 2 RGB image, b) Sen2Cor classified image, c) original H35 image (In this spatial subset no data come from H35) and d) MARS-based H35 image.

3.3 Flowchart of MARS-based H35 Product for the Mountainous Areas

The flowchart of the proposed MARS FSC modelling approach for the mountainous parts is shown in Figure 25.

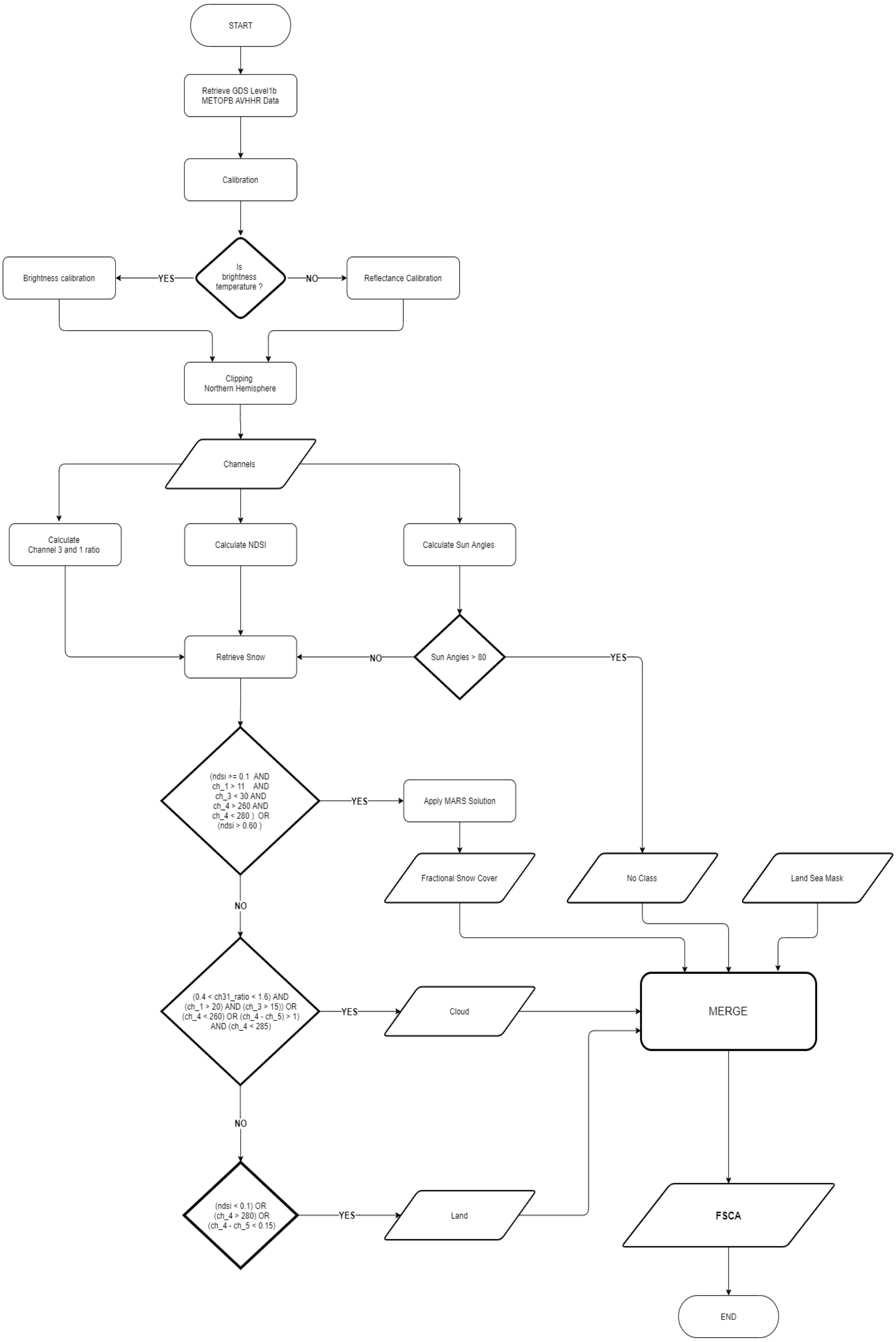


Figure 25. MARS-based H35 processing flowchart.

According to the process flowchart, the AVHRR bands are read after the necessary brightness and reflectance calibration process. Sun angle information is additionally derived in order to attain no-class value to pixels for which the sensor viewing conditions are too challenging, i.e., sun angle > 80 degrees. If it is decided that a pixel contains snow cover, then the final MARS model is employed to estimate the FSC value of that pixel. Once the cloud and the land pixels are decided all layers, including the land-sea mask, are merged to obtain the final FSC product. The whole process chain to generate a MARS-based H35 product takes 40-45 minutes on average, which meets with the requirements specified on the associated ATBD.

3.4 Visual Comparison: MARS-Based H35 versus MODIS MOD09GA

In order to analyze the spatial consistency of the FSC prediction by the final MARS model, H35 and MODIS MOD09GA (Vermote et al. 2015) false color RGB composite image pairs are compared for several dates as shown in Figures 26-32. In MOD09GA false color images, the following band combination is used: R = band 6, G = band 2 and B = band 4. In this band combination, ice and snow appear as bright blue; whereas, water bodies are near black. Saturated soil can be seen also in blue, and clouds are still white.

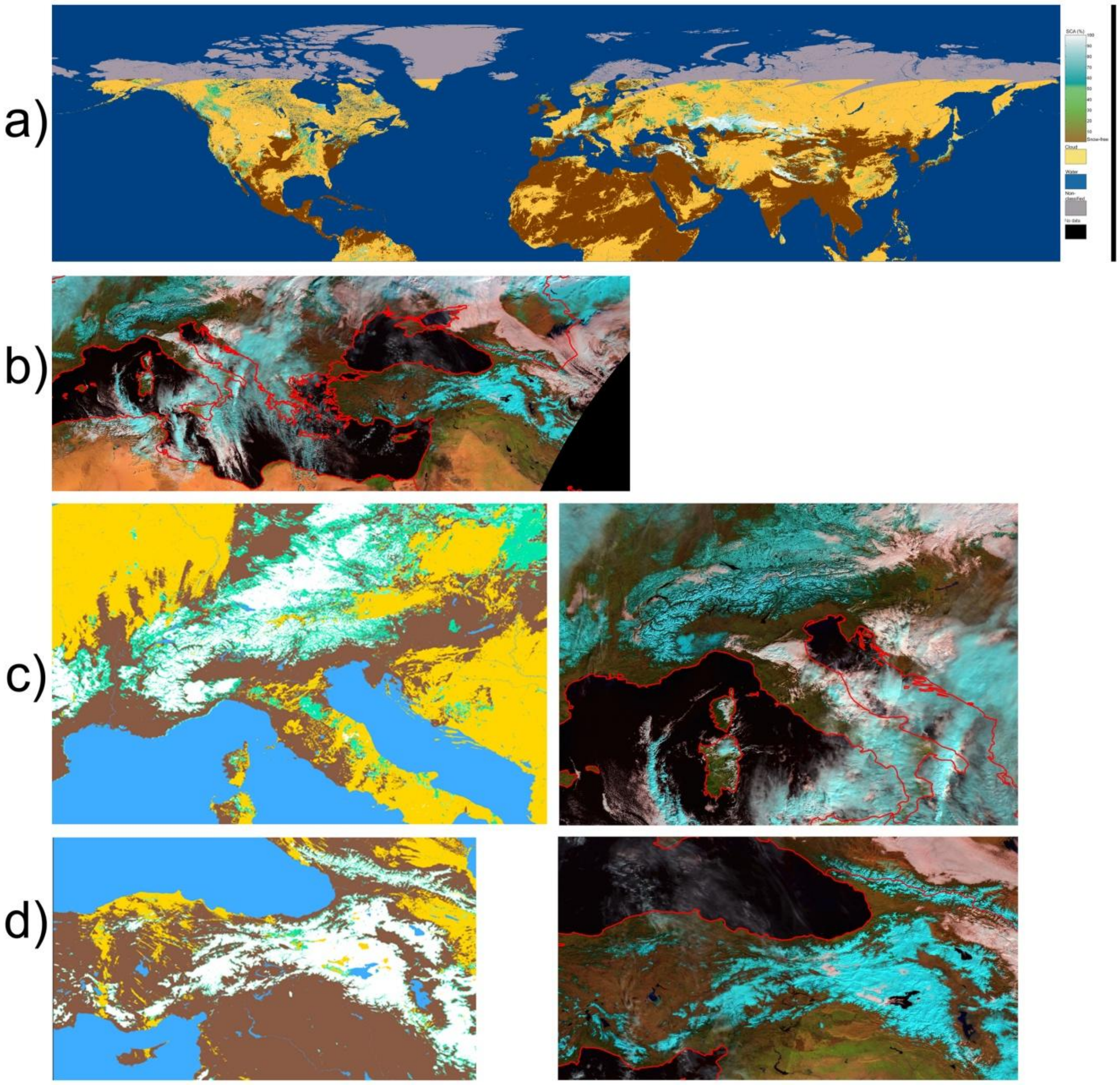


Figure 26. 4 Feb 2019: a) MARS-based H35, b) MOD09GA False RGB, c) H35 vs MOD09GA over Alps and d) H35 vs. MOD09GA over Turkey.

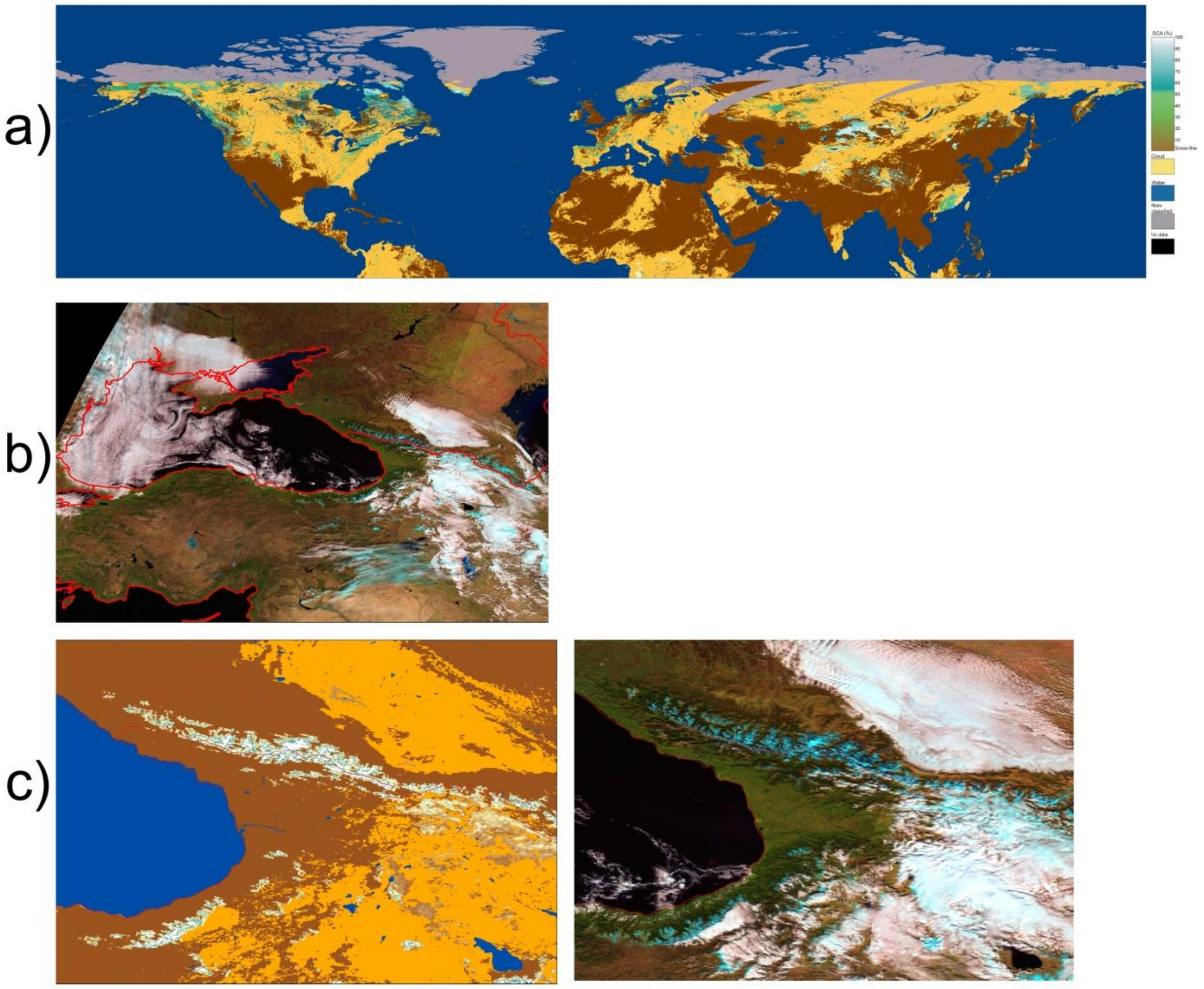


Figure 27. 2 Nov 2018: a) MARS-based H35, b) MOD09GA False RGB, c) H35 vs. MOD09GA over Turkey.

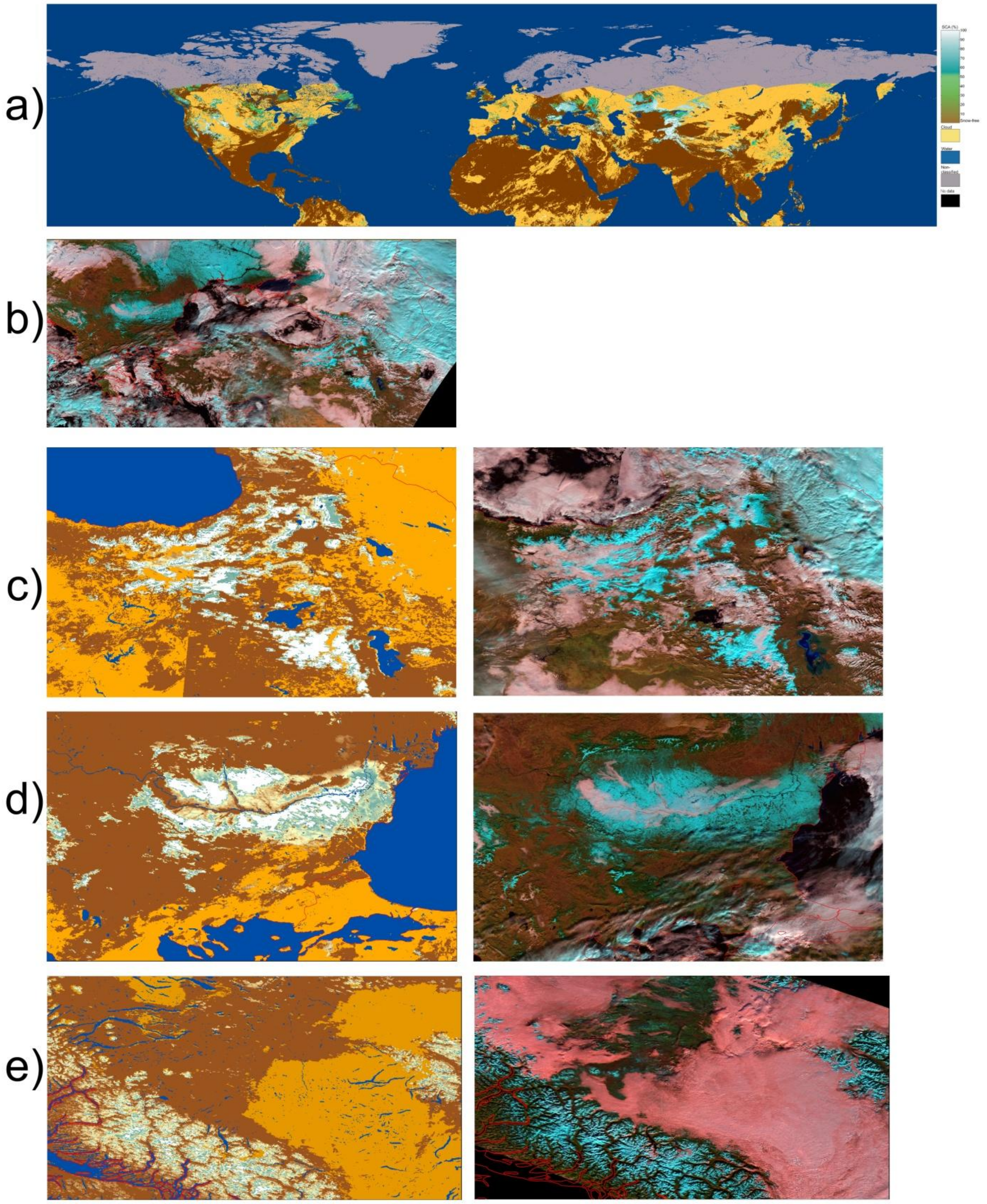


Figure 28. 2 Dec 2018: a) MARS-based H35, b) MOD09GA False RGB, c) H35 vs. MOD09GA over Turkey, d) H35 vs. MOD09GA over Bulgaria, e) H35 vs. MOD09GA over Mount Waddington Canada.

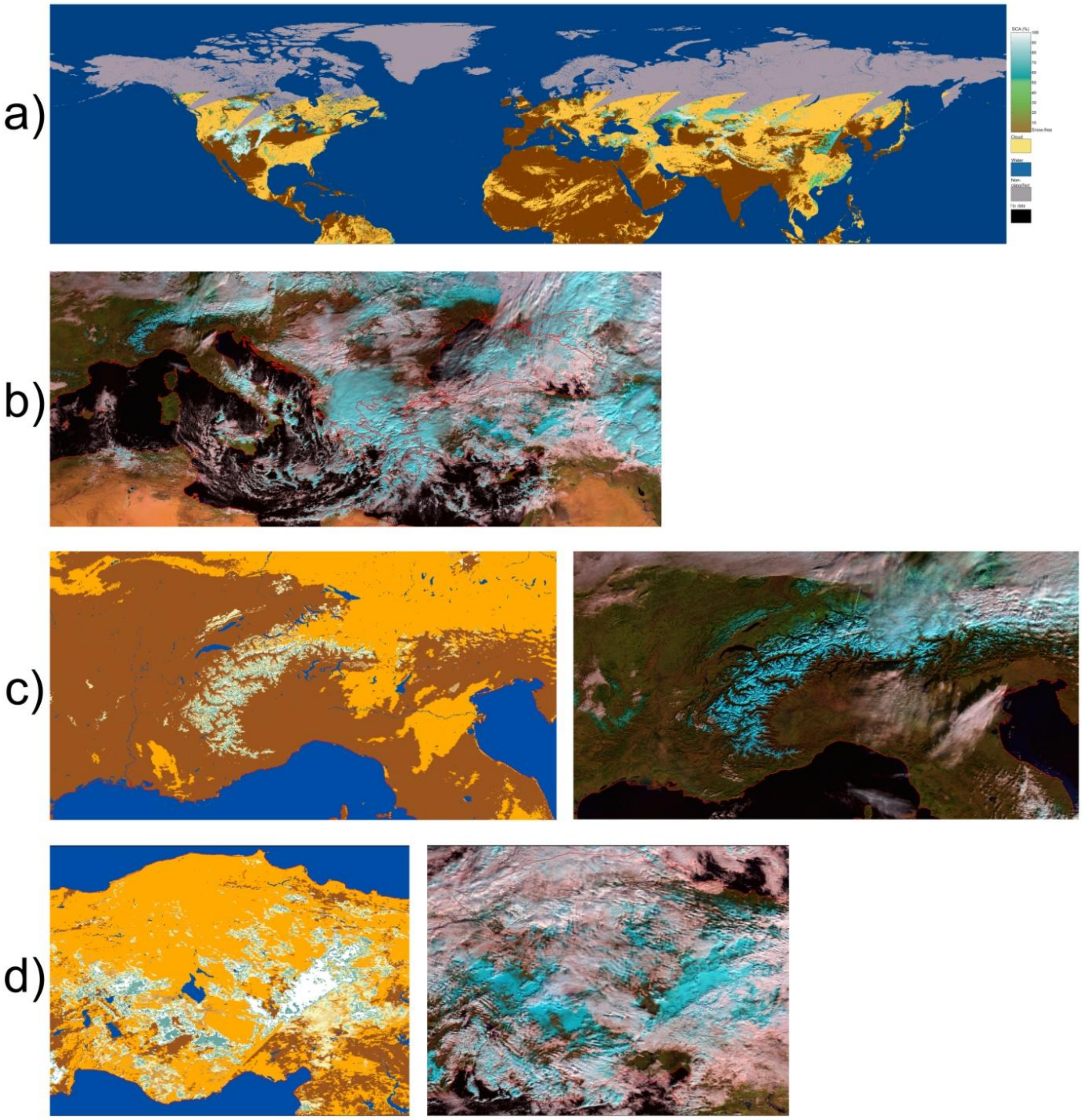


Figure 29. 3 Jan 2019: a) MARS-based H35, b) MOD09GA False RGB, c) H35 vs. MOD09GA over Alps, d) H35 vs. MOD09GA over Turkey.

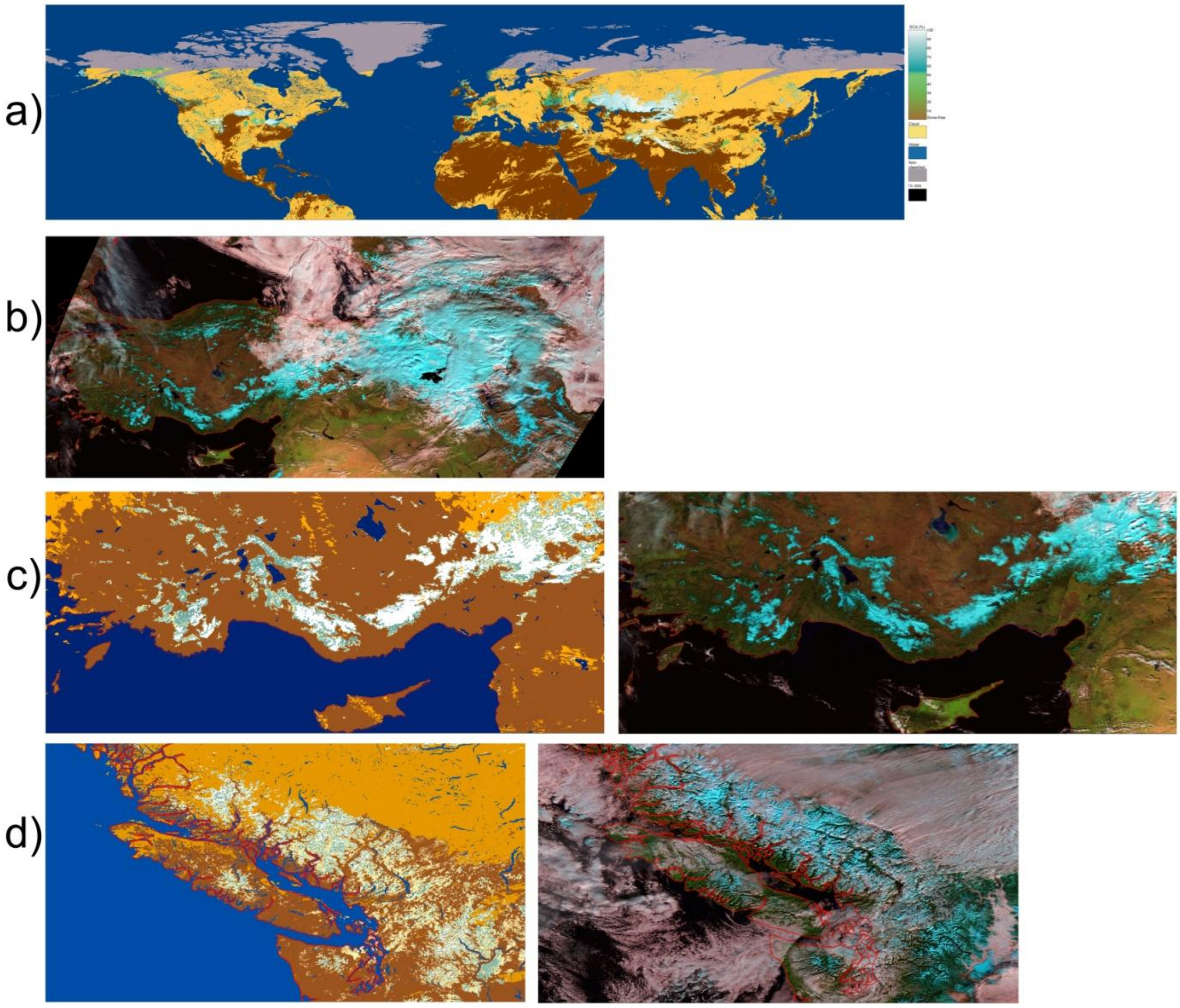


Figure 30. 2 Feb 2019: a) MARS-based H35, b) MOD09GA False RGB, c) H35 vs. MOD09GA over Turkey, d) H35 vs. MOD09GA over Mount Waddington Canada.

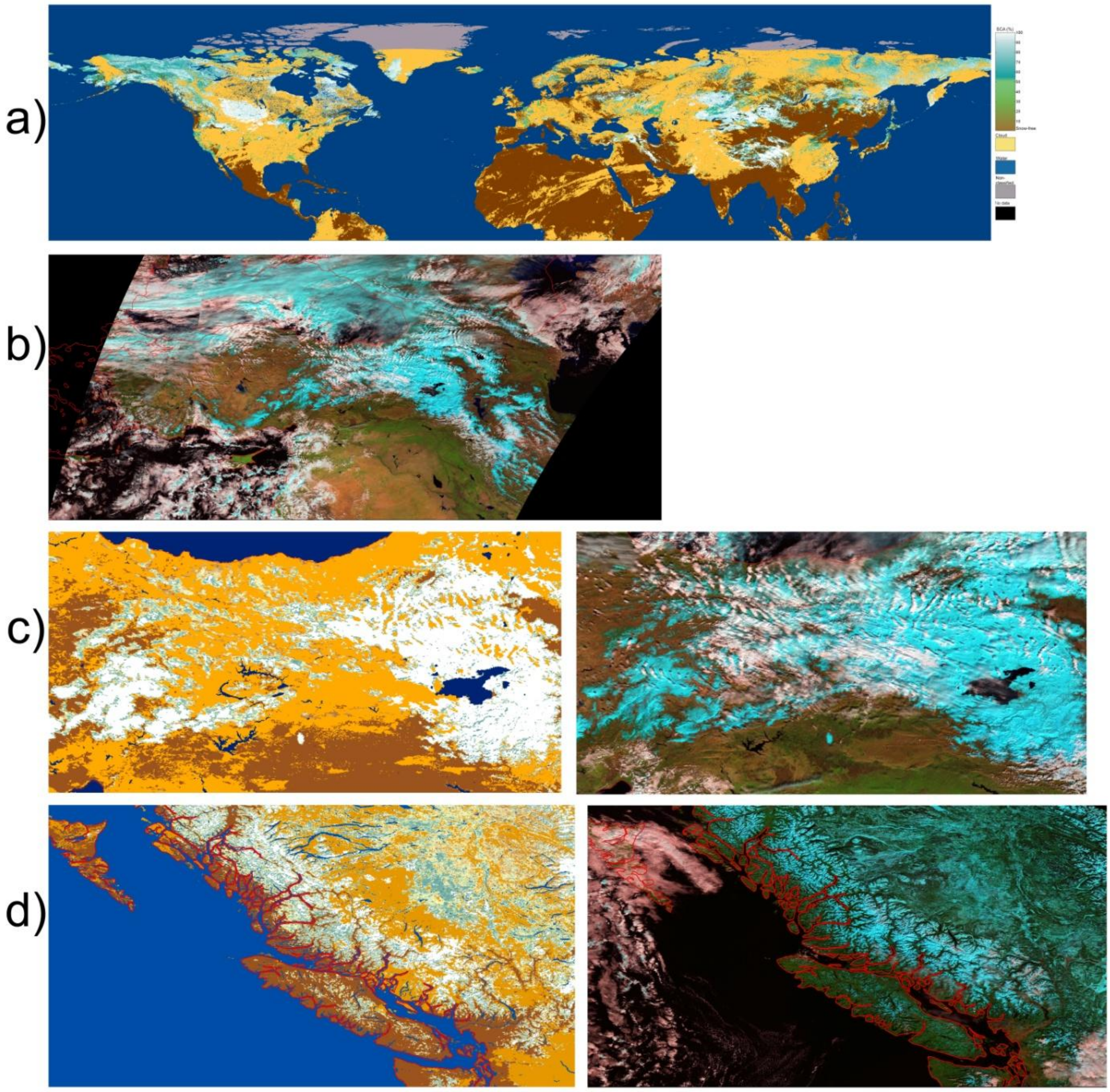


Figure 31. 2 Mar 2019: a) MARS-based H35, b) MOD09GA False RGB, c) H35 vs. MOD09GA over Turkey, d) H35 vs. MOD09GA over Mount Waddington Canada.

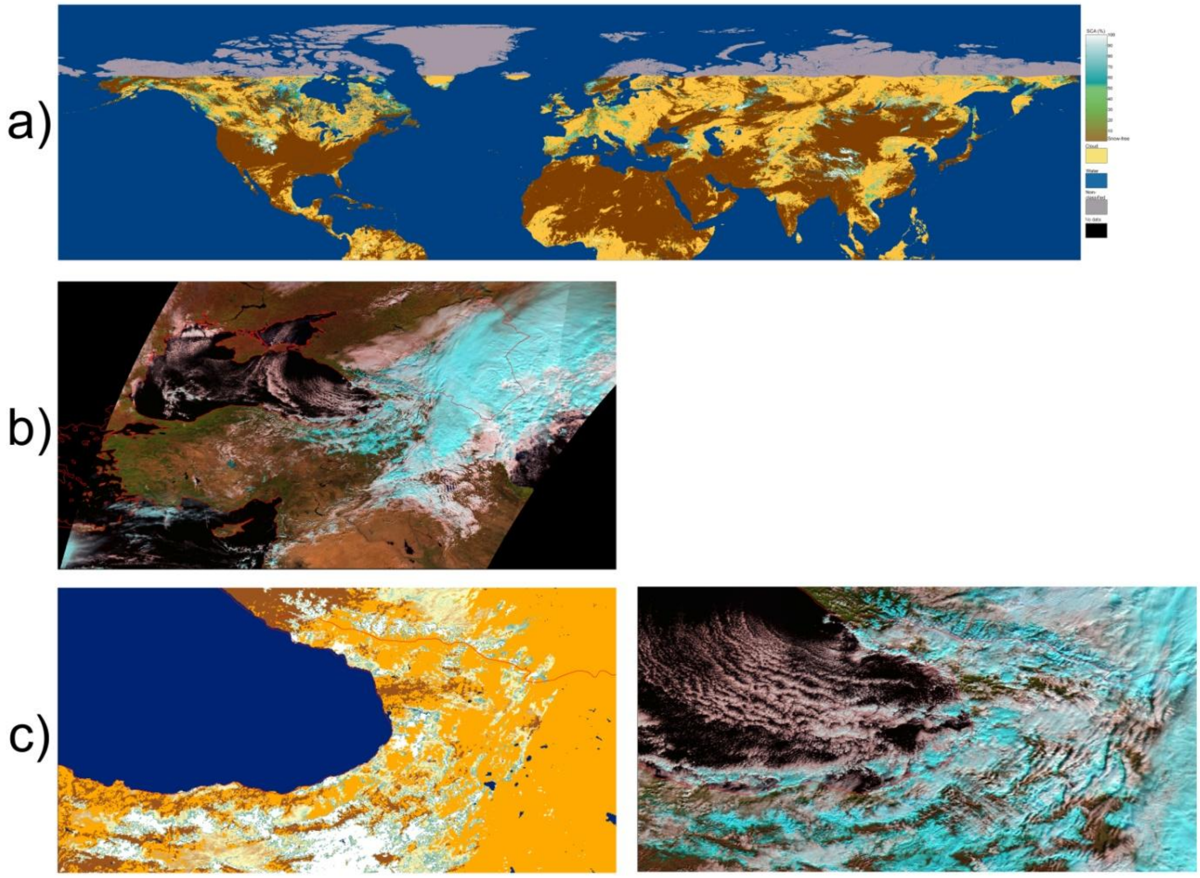


Figure 32. 2 Nov 2019: a) MARS-based H35, b) MOD09GA False RGB, c) H35 vs. MOD09GA over Turkey.

3.5 Variable Importances of MARS and RF Models

In MARS, the relative importance of a certain predictor variable can be obtained by determining the rate of change in the sum of squared errors without all the BFs containing that variable. The most important variable with the highest score is the variable that leads to the greatest decrease in the predictive ability of the model after it is completely discarded from the model. Contrarily, less significant variables receive lower scores corresponding to the ratio of degradation in the fit generated by these variables to that of the top scoring variable and the results are given on a scale of 0 to 100% over *generalized cross validation* (GCV) or *residual sum of squares* (RSS). The predictor variable with the most significant contribution has always relative importance of 100%.

In RFs, "permutation importance" is used to derive the variable importances. The method is based on the idea that if a certain predictor variable is not important, then randomly reordering its values, while keeping the rest in the same order, should not degrade the prediction accuracy (Biau and Scornet 2016; Cutler et al. 2012). The difference between the

OOB error rate from the original dataset and the OOB error rate of the model trained with the permuted predictor in terms of MSE for regression is used as the measure of variable importance. In case of an important predictor, permuting of its values leads to an increase in the OOB error rate.

The most important two predictors for the initial MARS model are NDSI and Band 2 with variable importance scores 100% and 24.8%, respectively, as shown in Table 16. NDSI and Band 2 are again the top ranked predictors for the initial RF model with relative importance scores with 2.93 and 2.20, whereas Band 1 (1.79) and NDVI (1.77) variables are relatively less important for the model's predictive performance.

Table 15. Variable importances of MARS and RF models. MARS variable importances are given over RSS within 0-100%, whereas RF variable importances are expressed in terms of increase in OOB error rates.

Predictor Name	Initial MARS Model	Initial RF Model	Final MARS Model
Band 1	18.4	1.7903	29.3
Band 2	24.8	2.2036	0.0 (unused)
Band 3a	20.2	1.9112	11.6
Band 4	14.8	1.9200	0.0 (unused)
Band 5	15.6	2.0843	0.0 (unused)
NDSI	100.0	2.9298	100.0
NDVI	18.8	1.7693	0.0 (unused)

As seen in Table 15, the initial MARS model uses the complete set of predictor variables, whereas only three predictors are employed in the final MARS model, which are NDSI, Band 1 and Band 3a with the corresponding variable importance scores 100%, 29.3% and 11.6%, respectively.

4. DISCUSSIONS

Recent technological improvements in all kinds of measurement devices (which is AVHRR in our case) create a gigantic and continuously growing supply of information (i.e., big data) to analyze. This situation forces us, earth scientists, to position the machine learning issues within the areas of *Data Mining* and *Remote Sensing*.

In this AS study, our main task is to develop an alternative methodology for the mountainous part of the H35 product by employing MARS algorithm instead of using traditional formula-based approaches who have their scientific origins in the natural sciences, especially, in physics.

According to our previous experiences on FSC mapping by utilizing machine learning approaches (Çiftçi et al. 2017; Kuter 2020; Kuter et al. 2017; Kuter et al. 2018; Kuter et al. 2016), MARS and RF are more advantageous as compared to artificial neural networks (ANN) and support vector machines (SVM) methods since:

- ANNs and SVMs are notorious for their black-box nature, i.e., it is impossible to exactly explain in a comprehensible form the process through which a given decision (output) is made by the ANN.
- Training process is expensive and time consuming in ANNs due to the large number model tuning parameters.
- SVMs are not favorable for large-sized datasets due to the large demand of the quadratic programming.
- Finding the suitable kernel function for a specific application constitutes a significant challenge.

On the other hand MARS and RF have the following superiorities:

- MARS and RF can handle both categorical and continuous data.
- They are suitable for high dimensional and complex datasets.
- Model building process is fast in MARS and RF since both algorithms depends on few model tuning parameters (i.e., 2 parameters for MARS and 1 parameter for RF)
- They both provide variable importance scores.

The high *R* and low RMSE values on the independent test images indicate that MARS-based H35 are in good agreement with reference FSC values. Additionally, visual comparisons with MODIS images reveal that the proposed methodology can successfully capture the spatial distribution of snow cover.

One challenging issue to be kept in mind is the underestimation of the snow cover in Sen2Cor scene classification software as shown in Figures 21-24. It should be emphasized that for all machine learning algorithms, the training data must be highly representative in order to effectively reveal the underlying relationship between the predictor variables and its response and to have good generalization ability on unseen data sets. Thus, utmost attention has been paid while selecting the Sentinel 2 scenes used in the analysis.

The results obtained during this AVS study will be represented as a scientific article which will also include rigorous validation of the MARS-based H35 product for the whole 2018-2019 snow season.

5. REFERENCES

Akyürek, Z., Hall, D.K., Riggs, G.A. & Sensoy, A. (2010). Evaluating the Utility of the Ansa Blended Snow Cover Product in the Mountains of Eastern Turkey. *International Journal of Remote Sensing, 31,* 3727-3744.

Belgiu, M. & Drăguţ, L. (2016). Random Forest in Remote Sensing: A Review of Applications and Future Directions. *ISPRS Journal of Photogrammetry and Remote Sensing, 114,* 24-31.

Beniston, M. (2003). Climatic Change in Mountain Regions: A Review of Possible Impacts. In H.F. Diaz (Ed.), *Climate Variability and Change in High Elevation Regions: Past, Present & Future* (pp. 5-31). Dordrecht, Netherlands: Springer.

Biau, G. & Scornet, E. (2016). A Random Forest Guided Tour. *Test, 25,* 197-227.

Boulesteix, A.-L., Janitza, S., Kruppa, J. & König, I.R. (2012). Overview of Random Forest Methodology and Practical Guidance with Emphasis on Computational Biology and Bioinformatics. *WIREs Data Mining and Knowledge Discovery, 2,* 493-507.

Breiman, L. (1996). Bagging Predictors. *Machine Learning, 24,* 123-140.

Breiman, L. (2001). Random Forests. *Machine Learning, 45,* 5-32.

Brown, R.D., Brasnett, B. & Robinson, D. (2003). Gridded North American Monthly Snow Depth and Snow Water Equivalent for Gcm Evaluation. *Atmosphere-Ocean, 41,* 1-14.

Brown, R.D. & Goodison, B.E. (1996). Interannual Variability in Reconstructed Canadian Snow Cover, 1915-1992. *Journal of Climate, 9,* 1299-1318.

Çiftçi, B.B., Kuter, S., Akyürek, Z. & Weber, G.W. (2017). Fractional Snow Cover Mapping by Artificial Neural Networks and Support Vector Machines. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, IV-4/W4,* 179-187.

Clark, M.P., Slater, A.G., Barrett, A.P., Hay, L.E., McCabe, G.J., Rajagopalan, B. & Leavesley, G.H. (2006). Assimilation of Snow Covered Area Information into Hydrologic and Land-Surface Models. *Advances in Water Resources, 29,* 1209-1221.

Cline, D.W., Bales, R.C. & Dozier, J. (1998). Estimating the Spatial Distribution of Snow in Mountain Basins Using Remote Sensing and Energy Balance Modeling. *Water Resources Research, 34,* 1275-1285.

Colditz, R.R. (2015). An Evaluation of Different Training Sample Allocation Schemes for Discrete and Continuous Land Cover Classification Using Decision Tree-Based Algorithms. *Remote Sensing, 7*, 9655-9681.

Cutler, A., Cutler, D.R. & Stevens, J.R. (2012). Random Forests. In C. Zhang, & Y. Ma (Eds.), *Ensemble Machine Learning: Methods and Applications* (pp. 157-175). Boston, MA: Springer US.

Díaz-Uriarte, R. & De Andres, S.A. (2006). Gene Selection and Classification of Microarray Data Using Random Forest. *BMC Bioinformatics, 7,* 3.

Dozier, J. (1992). Opportunities to Improve Hydrologic Data. *Reviews of Geophysics, 30,* 315-331.

Foster, J., Liston, G., Koster, R., Essery, R., Behr, H., Dumenil, L., Verseghy, D., Thompson, S., Pollard, D. & Cohen, J. (1996). Snow Cover and Snow Mass Intercomparisons of General Circulation Models and Remotely Sensed Datasets. *Journal of Climate, 9,* 409-426.

Friedman, J.H. (1991). Multivariate Adaptive Regression Splines. *The Annals of Statistics, 19,* 1-67.

Gafurov, A. & Bárdossy, A. (2009). Cloud Removal Methodology from Modis Snow Cover Product. *Hydrology and Earth Sytem Sciences, 13,* 1361-1373.

Guan, H., Li, J., Chapman, M., Deng, F., Ji, Z. & Yang, X. (2013). Integration of Orthoimagery and Lidar Data for Object-Based Urban Thematic Mapping Using Random Forests. *International Journal of Remote Sensing, 34,* 5166-5186.

Hall, D.K. & Martinec, J. (1985). *Remote Sensing of Ice and Snow*. USA: Chapman and Hall.

Hartman, M.D., Baron, J.S., Lammers, R.B., Cline, D.W., Band, L.E., Liston, G.E. & Tague, C. (1999). Simulations of Snow Distribution and Hydrology in a Mountain Basin. *Water Resources Research, 35,* 1587-1603.

Hastie, T., Tibshirani, R. & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. (2nd ed.). NY, USA: Springer.

Hosaka, M., Nohara, D. & Kitoh, A. (2005). Changes in Snow Cover and Snow Water Equivalent Due to Global Warming Simulated by a 20km-Mesh Global Atmospheric Model. *Sola, 1,* 93-96.

Houborg, R. & McCabe, M.F. (2018). A Hybrid Training Approach for Leaf Area Index Estimation Via Cubist and Random Forests Machine-Learning. *ISPRS Journal of Photogrammetry and Remote Sensing, 135,* 173-188.

Hüsler, F., Jonas, T., Wunderle, S. & Albrecht, S. (2012). Validation of a Modified Snow Cover Retrieval Algorithm from Historical 1-Km Avhrr Data over the European Alps. *Remote Sensing of Environment, 121,* 497-515.

Kuter, S. (2020). Completing the Machine Learning Saga in Fractional Snow Cover Estimation from Modis Terra Reflectance Data: Random Forests Versus Support Vector Regression (under Review). *Remote Sensing of Environment*

Kuter, S., Akyürek, Z., Kuter, N. & Weber, G.-W. (2017). An Alternative Method for Snow Cover Mapping on Satellite Images by Modern Applied Mathematics. In A.A. Pinto, & D. Zilberman (Eds.), *Modeling, Dynamics, Optimization and Bioeconomics Ii* (pp. 267-292). Switzerland: Springer International Publishing AG.

Kuter, S., Akyurek, Z. & Weber, G.W. (2018). Retrieval of Fractional Snow Covered Area from Modis Data by Multivariate Adaptive Regression Splines. *Remote Sensing of Environment, 205,* 236-252.

Kuter, S., Akyürek, Z. & Weber, G.W. (2016). Estimation of Subpixel Snow-Covered Area by Nonparametric Regression Splines. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W1*, 31-36.

Laternser, M. & Schneebeli, M. (2003). Long-Term Snow Climate Trends of the Swiss Alps (1931–99). *International Journal of Climatology, 23*, 733-750.

Lawrence, R.L., Wood, S.D. & Sheley, R.L. (2006). Mapping Invasive Plants Using Hyperspectral Imagery and Breiman Cutler Classifications (Randomforest). *Remote Sensing of Environment, 100*, 356-362.

Lehning, M., Völksch, I., Gustafsson, D., Nguyen, T.A., Stähli, M. & Zappa, M. (2006). Alpine3d: A Detailed Model of Mountain Surface Processes and Its Application to Snow Hydrology. *Hydrological Processes, 20*, 2111-2128.

Marty, C. (2008). Regime Shift of Snow Days in Switzerland. *Geophysical Research Letters, 35*, L12501.

Mueller-Wilm, U. (2019). Sen2cor Configuration and User Manual - V2.8. Available at: http://step.esa.int/thirdparties/sen2cor/2.8.0/docs/S2-PDGS-MPC-L2A-SUM-V2.8.pdf.

Nitze, I., Barrett, B. & Cawkwell, F. (2015). Temporal Optimisation of Image Acquisition for Land Cover Classification with Random Forest and Modis Time-Series. *International Journal of Applied Earth Observation and Geoinformation, 34,* 136-146.

Pelletier, C., Valero, S., Inglada, J., Champion, N. & Dedieu, G. (2016). Assessing the Robustness of Random Forests to Map Land Cover with High Resolution Satellite Image Time Series over Large Areas. *Remote Sensing of Environment, 187,* 156-168.

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

Romanov, P., Tarpley, D., Gutman, G. & Carroll, T. (2003). Mapping and Monitoring of the Snow Cover Fraction over North America. *Journal of Geophysical Research: Atmospheres, 108*

Timm, B.C. & McGarigal, K. (2012). Fine-Scale Remotely-Sensed Cover Mapping of Coastal Dune and Salt Marsh Ecosystems at Cape Cod National Seashore Using Random Forests. *Remote Sensing of Environment, 127,* 106-117.

Tramontana, G., Ichii, K., Camps-Valls, G., Tomelleri, E. & Papale, D. (2015). Uncertainty Analysis of Gross Primary Production Upscaling Using Random Forests, Remote Sensing and Eddy Covariance Data. *Remote Sensing of Environment, 168*, 360-373.

Vermote, E.F., Roger, J.C. & Ray, J.P. (2015). Modis Surface Reflectance User's Guide (Collection 6) - Version 1.4. *MODIS Land Surface Reflectance Science Computing Facility*

Yu, X., Hyyppä, J., Vastaranta, M., Holopainen, M. & Viitala, R. (2011). Predicting Individual Tree Attributes from Airborne Laser Point Clouds Based on the Random Forests Technique. *ISPRS Journal of Photogrammetry and Remote Sensing, 66*, 28-37.