

VILNIUS UNIVERSITY
KLAIPĖDA UNIVERSITY
NATURE RESEARCH CENTRE

Justinas
KILPYS

Determination of snow cover characteristics in flat land areas using remote sensing methods

SUMMARY OF DOCTORAL DISSERTATION

Natural Sciences,
Physical Geography (N 006)

VILNIUS 2021

This dissertation was written between 2016 and 2021 at Vilnius University.

Academic supervisor:

Prof. dr. Egidijus, Rimkus (Vilnius University, Natural Sciences, Physical Geography, N 006)

This doctoral dissertation will be defended in a public meeting of the Dissertation Defence Panel:

Chairman – Assoc. Prof. Dr. Donatas, Pupienis (Vilnius University, Natural Sciences, Physical Geography, N 006).

Members:

Prof. Habil. Dr. Algimantas, Česnulevičius (Vilnius University, Natural Sciences, Physical Geography, N 006).

Prof. Habil. Dr. Marek, Kejna (Nicolaus Copernicus University, Natural Sciences, Physical Geography, N 006).

Assoc. Prof. Dr. Jūratė, Sužiedelytė Visockienė (Vilnius Tech, Technological Sciences, Measurement Engineering, T 010).

Dr. Julius, Taminskas (Nature Research Centre, Natural Sciences, Physical Geography, N 006).

The dissertation shall be defended at a public meeting of the Dissertation Defence Panel at 14.00 on 24th September 2021 in meeting room 313 of the Institute of Geosciences (Vilnius University).

Address: M. K. Čiurlionio g. 21/27, LT-03101 Vilnius, Lithuania
Tel. 370 5 239 8202; e-mail: info@chgf.vu.lt.

The text of this dissertation can be accessed at the libraries of Vilnius University, Nature Research Centre, Klaipėda University, as well as on the website of Vilnius University: www.vu.lt/lt/naujienos/ivykiu-kalendorius

VILNIAUS UNIVERSITETAS
KLAIPĖDOS UNIVERSITETAS
GAMTOS TYRIMŲ CENTRAS

Justinas
KILPYS

Sniego dangos rodiklių tyrimas nuotoliniais metodais lyguminėse teritorijose

DAKTARO DISERTACIJOS SANTRAUKA

Gamtos mokslai,
Fizinė geografija (N 006)

VILNIUS 2021

Disertacija rengta 2016–2021 metais Vilniaus universitete.

Mokslinis vadovas:

prof. dr. Egidijus Rimkus (Vilniaus universitetas, gamtos mokslai, fizinė geografija, N 006).

Gynimo taryba:

Pirmininkas – **doc. dr. Donatas Pupienis** (Vilniaus universitetas, gamtos mokslai, fizinė geografija, N 006).

Nariai:

prof. dr. Algimantas Česnulevičius (Vilniaus universitetas, gamtos mokslai, fizinė geografija, N 006).

prof. dr. Marek Kejna (Nicolaus Copernicus universitetas, gamtos mokslai, fizinė geografija, N 006).

doc. dr. Jūratė Sužiedelytė Visockienė (Vilnius Tech, technologijos mokslai, matavimų inžinerija, T 010).

dr. Julius Taminskas (Gamtos tyrimų centras, gamtos mokslai, fizinė geografija, N 006).

Disertacija ginama viešame Gynimo tarybos posėdyje 2021 m. rugsėjo mėn. 24 d. 14 val. Vilniaus universiteto Chemijos ir geomokslų fakulteto Geomokslų instituto 313 auditorijoje.

Adresas: M. K. Čiurlionio g. 21/27, LT-03101 Vilnius, Lietuva, tel. +370 5 239 8202; el. paštas info@chgf.vu.lt.

Disertaciją galima peržiūrėti Vilniaus universiteto, Gamtos tyrimų centro, Klaipėdos universiteto bibliotekose ir VU interneto svetainėje adresu: <https://www.vu.lt/naujienos/ivykiu-kalendorius>

INTRODUCTION

Traditionally, snow cover measurements are done at meteorological stations (MS), but in recent decades, there have been a growing number of applications for satellite remote sensing. Satellite sensors record reflected solar radiation or electromagnetic radiation emitted by the surface itself. When snow cover forms, the intensity of the reflected and emitted radiation changes, enabling the identification of snow. Compared to ground-based measurements, satellite data provide detailed spatial information, even in the remote areas.

However, satellite sensors have their limitations. Application of the optical sensors is limited by the cloud cover, microwave sensors have low resolution and synthetic aperture radar (SAR) has a narrow swath. Researchers are looking for different ways to overcome these constraints and use the advantages of different satellite sensors to extract detailed information about snow cover. This thesis aimed to develop methodology to derive snow cover parameters in lowland areas from satellite sensors.

Motivation

Due to the automation of the measurements in Lithuania from 2000 to 2021, the number of MS reporting the presence of snow decreased by 86.2% (from 65 to 9), the number of snow cover depth measurements decreased by 58.5% (from 65 to 27) and the number of snow surveys decreased more than two-fold (from 65 to 31). Similar trends were also reported in the other Baltic states.

The emerging gap in snow cover monitoring can be filled using satellite observations. The number of satellite observations is increasing every year, but there are large differences in resolution, frequency of observations and the physical characteristics that are determined. There is a constant need to develop data processing algorithms that could derive useful information about snow cover from different satellite sensors.

Research subject

Snow cover detection methodology from satellite remote sensing.

Aim of the study

To analyse the characteristics of snow cover and their trends in the Baltic states and to evaluate the possibilities and accuracy of satellite data applications for snow monitoring.

Study objectives:

1. To analyse the spatial distribution and trends of snow cover characteristics in the Baltic states during the period 1961–2015.
2. To develop a snowfall detection algorithm using Advanced Technology Microwave Sounder (ATMS) data.
3. To derive the annual number of snow cover days (SCD) using Moderate Resolution Imaging Spectroradiometer (MODIS) data and validate it using ground measurements.
4. To improve the SAR snow cover detection algorithm in lowlands (< 500 m) with various land cover types and thin snow cover.
5. To downscale the EUMETSAT Satellite Application Facility on Support to Operational Hydrology and Water Management (HSAF) snow water equivalent (SWE) product using auxiliary environmental data.

Defended statements

1. Data from passive satellite microwave sensors can be used to determine snowfall over land.
2. Filling cloud gaps in snow cover products from optical satellite sensors allows to accurately determine the monthly and annual number of SCD.
3. The accuracy of snow cover determination using SAR data can be increased by using land cover type-specific backscattering thresholds.

Scientific novelty

1. A methodology for the classification of snowfall using ATMS sensor data, and a random forest (RF) machine learning algorithm has been developed.
2. A methodology for filling cloud gaps in the MODIS snow cover product using an additional minimum air temperature filter is proposed.
3. The SAR snow cover detection algorithm was improved by using land cover type-specific backscattering ratio thresholds.
4. A methodology has been developed to downscale HSAF SWE product from ~25 km to ~5 km using auxiliary environmental information.

Relevance and applicability of the study

The methodology proposed in this study can be easily applied for snow monitoring at national meteorological and hydrological services. The application of machine learning algorithms allows us to process large data sets and determine links between parameters, even if the physical connection is not obvious. Results of this research contribute to the practical application of remote sensing in determining various snow cover characteristics.

1. DATA

The satellite and in situ data used in the thesis are summarized in Table 1. The spatial and temporal variation of the data was mainly imposed by the satellite sensors, their resolution and the launch date. The choice of study regions was also influenced by the availability of MS data.

Table 1. Summary of in situ and satellite data used in the thesis.

Data	Resolution	Area	Period	Measured or derived parameters
MS	Point measurements	Global	2015–2018	Precipitation type and amount, air temperature, relative humidity
		Baltic states	1961–2015	Snow depth, snow presence, air temperature, precipitation amount
			2012–2017	SWE
		Lithuania	2002–2018	Snow depth, snow presence
		Šventoji River Basin	2014–2019	Snow depth, air temperature
ATMS	15 km	Global	2015–2018	Precipitation type
MODIS	500 m	Lithuania	2002–2018	Snow cover extent, days with snow over
Sentinel-1 SAR	20 m	Šventoji River Basin	2014–2019	Snow cover extent
Sentinel-2	20 m	Šventoji River Basin	2014–2019	Snow cover extent
HSAF	0.25°	Baltic states	2012–2017	SWE

The main focus area of the study was snow cover characteristics in the Baltic states and Lithuania. Only snowfall classification using ATMS sensors was performed on the global scale. Most of the in situ data from the Baltic states were from 1961 to 2015. Snow cover characteristics (snow presence, snow depth and SWE) were collected from 57 MS: 21 in Estonia, 19 in Latvia and 17 in Lithuania (Figure 1). SWE at meteorological stations is not observed daily, but it is measured every 5 or 10 days during the snow surveys. SWE observations were available from 52 MS in the Baltic states (Figure 1) and field measurements (2016–2019) performed during the study period in the Vilnius district, Neris Regional Park.

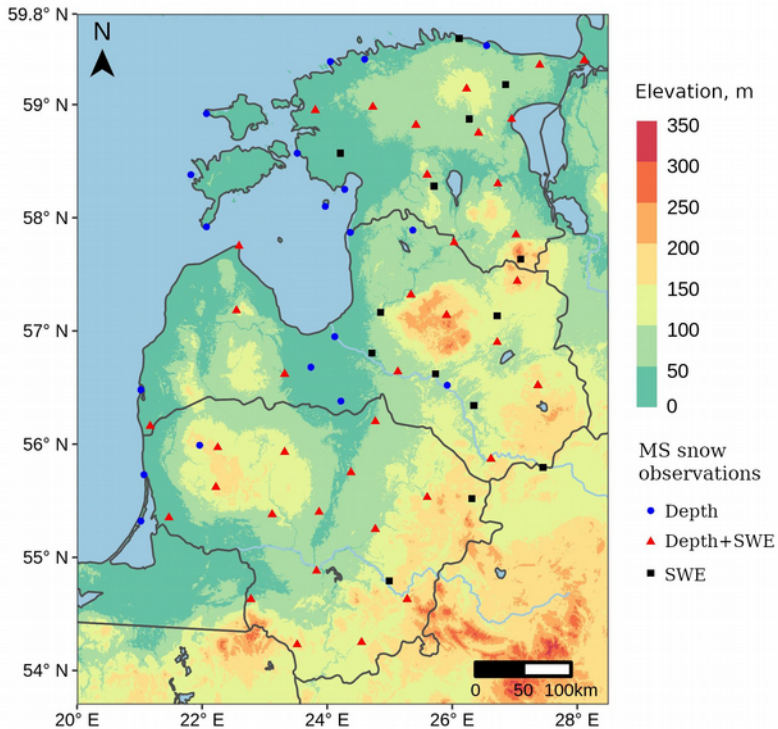


Figure 1. Location of snow observation in the Baltic states.

Gridded meteorological data ($0.5^\circ \times 0.5^\circ$) from the CRU TS4.01 database (Harris, Jones, 2017) was used to analyse the relationship between snow cover parameters and average monthly air temperature and precipitation. Area elevation data was obtained from the Shuttle Radar Topographic Mission Digital Elevation Model (SRTM DEM) of the National Aeronautics and Space Administration (NASA) (Jarvis et al., 2008).

Snowfall detection algorithm was developed using global ATMS data obtained from Earth System Science Interdisciplinary Center (ESSIC) for the period 2012–2018. The ATMS is a microwave sounder that scans the Earth's surface and atmosphere in 22 spectral bands at a frequency of 23.8–185 GHz. ATMS snow classification algorithm training was done using global precipitation observations from the ground station (Figure 2) and NOAA Global Forecast System (GFS) analysis. GFS analysis data was provided in $0.5^\circ \times 0.5^\circ$ resolution with 3 hours time step and 20 different vertical atmospheric levels (200–1000 hPa). A brief description of the main indicators of the GFS model is provided in Table 2.

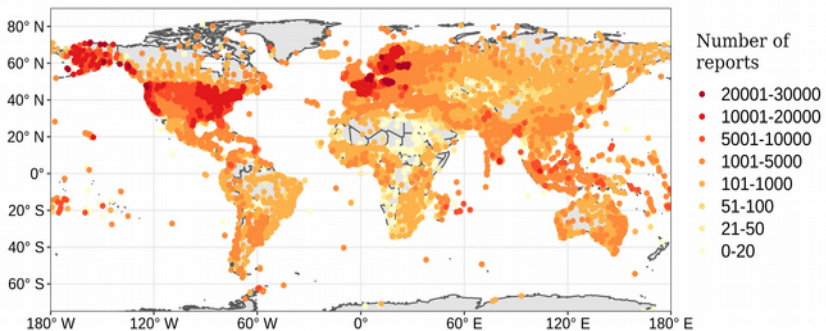


Figure 2. In situ stations with precipitation type measurements. The number of reports is from the period 2012–2018.

Table 2. The main GFS model analysis output parameters.

Abbreviation	Units	Explanation
HGT	gpm	Geopotential height
TMP	K	Temperature
RH	%	Relative humidity
VVEL	Pa/s	Vertical velocity
CLWMR	kg/kg	Cloud water mixing ratio
PRES	Pa	Pressure
CWAT	kg/m ²	Cloud water
PWAT	kg/m ²	Precipitable water
UGRD_10m	m/s	u-component of wind at 10 m
VGRD_10m	m/s	v-component of wind at 10 m
WEASD	kg/m ²	Water equivalent of accumulated snow depth

MODIS 2002–2018 cold period (October–April) data were obtained from the Earthdata database of NASA. In the MOD10A1 (“Terra” satellite) and MYD10A1 (“Aqua” satellite) snow cover products, snow on the surface is determined using the Normalized Difference Snow Index (NDSI) and additional brightness temperature and surface reflection filters (Hall, Riggs, 2016; Hall et al., 1995). NDSI was calculated using MODIS 4 and 6 spectral bands (their central wavelengths are 0.55 and 1.64 μm , respectively) (Hall et al., 2002):

$$NDSI = \frac{VIS_{0.55} - SWIR_{1.64}}{VIS_{0.55} + SWIR_{1.64}} \quad (1)$$

The extent of snow cover also was evaluated using Sentinel-1 SAR data provided by the European Space Agency (ESA). The Sentinel-1 SAR is a C-band radar with a centre frequency of 5.4 GHz (Bourbigot et al., 2016). In this study, ground range detected (GRD) double polarization (VV/VH) SAR data in the interferometric wide (IW) swath mode was used. Using two satellites (Sentinel-1A and

Sentinel-1B), the region of interest (Šventoji River Basin) could be monitored every 6 days. Ascending orbit data for Sentinel-1 was obtained for the cold periods in 2014–2019.

To determine the influence of land cover types on the SAR backscattering coefficient, we used the 2018 CORINE land cover (CLC 2018) database (Büttner et al., 2017). CLC 2018 data was obtained from Copernicus Land Monitoring Service. There were 22 CLC land cover types in the Šventoji River Basin, but they were grouped into seven main categories: 1) coniferous forests; 2) broadleaf/mixed forests; 3) arable land; 4) natural meadows; 5) urbanized territories; 6) wetlands; and 7) water bodies. Validation of the Sentinel-1 SAR-derived snow cover extent in the Šventoji River Basin was done using snow cover maps based on Sentinel-2 NDSI (Drusch et al., 2012). Information on snow thickness and air temperature was obtained from the Ukmergė and Utena MS. Snow depth from these stations was used to classify SAR images with thin (1–5 cm) and thick (> 5 cm) snow cover.

The HSAF SWE product was used to evaluate the spatial distribution of SWE in the Baltic states in 2012–2017. The HSAF SWE product is based on the intensity of terrestrial microwave radiation measured by different passive microwave sensors (AMSR-E, SSM/I and SSMIS) (HSAF, 2018; Trigo et al., 2011). The HSAF SWE detection algorithm is an assimilation of satellite and ground-based snow depth observations. One of the aims of this study was to increase the spatial resolution of the HSAF SWE dataset by using high resolution auxiliary environmental information: 1) 90 m resolution SRTM DEM (Jarvis et al., 2008); 2) 0.25° x 0.25° resolution E-OBS daily minimum temperature (Tmin) data from the E-OBS database (v17.0) (Cornes et al., 2018); 3) 0.08° x 0.08° resolution MODIS Global Land Cover Facility (GLCF) land cover data (Friedl et al., 2009).

2. METHODOLOGY

The homogeneity of in situ snow cover measurements in the Baltic states was tested using the Standard Normal Homogeneity Test (SNHT) (Alexandersson, 1986). Statistically significant changes of the snow cover characteristics were detected in the late 1980s. These changes were unrelated to the relocation of the stations or changes in the measurement methodology, and this period in the Baltic states can be considered as a change in the snow cover regime due to climate change.

The classification of winter was done by using the monthly number of SCD and a hierarchical clustering model (Euclidean distance measure, full connection). The SCD and maximum snow cover depth trends were estimated using the nonparametric Sen slope method (Helsel et al., 2020)). The rate of change was calculated for both monthly and annual data. Statistical significance of the trends was evaluated using the nonparametric Mann-Kendall test ($p < 0.05$).

RF classification was used to identify snowfall cases from ATMS data. Precipitation type observations from MS were used to train RF. Only precipitation cases within the 90-min interval after the ATMS overpass time and within a 50 km buffer zone around the central ATMS beam coordinates were used in the study. This time filter was applied, because it was reported that there was a 30–60 min precipitation intensity delay between ATMS and MS measurements (You et al., 2019). ATMS and MS data also were combined with NOAA GFS analysis data. The compiled dataset described meteorological conditions in the troposphere and at the surface during each ATMS overpass. ATMS brightness temperature and GFS variables were used as predictors in the RF snowfall classification model.

Twenty different iterations of the RF model were performed to select the most important predictors. This allowed one to reduce the number of predictors from 142 to 8. These selected variables were also used to construct a logistic regression model. The accuracy of

both models, RF and logistic regression, was assessed by precipitation observations from MS, which were not used in the RF training.

To derive SCD from MODIS data, cloud gaps have to be filled. Cloud gap filling was done using spatial (Steps I and II), temporal filtering (Step III) and temperature control (Step IV) (Figure 3).

Temporal filtering (Step III) was performed in two ways: backward and forward. In the forward case, the cloudy pixel was filled by its value after the end of the cloudy period. In the backward case, the pixel was filled by its value before the beginning of cloudy conditions. Depending on the meteorological situation, both methods can lead to underestimation or overestimation of the snow cover. However, the mean of these methods allows one to accurately determine SCD, as errors of these two filtering methods are opposite (Foppa, Seiz, 2012).

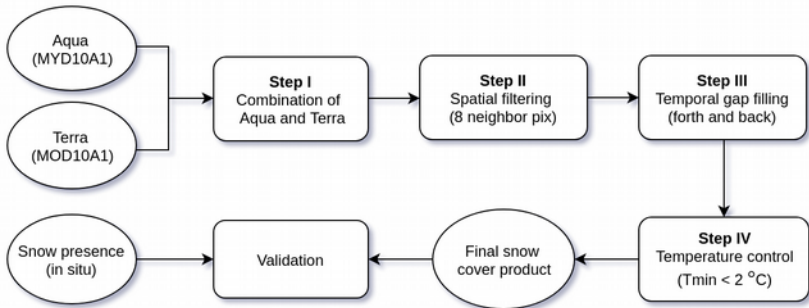


Figure 3. Schematic of MODIS snow cover data processing.

Temperature control (Step IV) was used to reduce the number of false pixel identifications of snow (Figure 3). To determine the optimal critical temperature threshold, minimum daily temperatures from 0.5 to 3.0 °C (with 0.5 °C increments) were tested.

Generated cloud-free MODIS snow cover data was validated using snow cover observations from MS in Lithuania. Validation was based on the confusion matrix and indices derived from it. The

accuracy of monthly and annual SCD derived from MODIS was also assessed using the Climatological Skill Score (SS_{clim}) (Wilks, 2011):

$$SS_{clim} = 1 - \frac{MSE}{MSE_{clim}} \quad (2)$$

where MSE is the root mean square error between MODIS and MS SCD, MSE_{clim} is the root mean square error between the climatological average of SCD (2002–2018) and the monthly (or annual) SCD derived from MS data on a given year. A negative SS_{clim} value would indicate that the MODIS snow product was less accurate than the climatological mean.

The extent of snow cover was also determined using Sentinel-1 SAR observations. The SAR data processing algorithm is presented in Figure 4. After the pre-processing of the SAR data, the SAR backscattering coefficient (σ_0) ratio between the images with snow and the standard image was calculated. Ratios were calculated both for the VV and VH polarizations. The standard SAR image is a mean of seven snow-free cases during the October months in period 2012–2019. The VV and VH polarization σ_0 ratios were fused using a methodology proposed by Nagler et al. (2016) and a combined ratio (R_C) was obtained.

The critical R_C thresholds distinguishing snow and no snow cases were derived using the intersection point of the SAR snow and snow-free R_C value distributions. Two different snow cover thickness categories were considered when determining R_C thresholds: thin snow cover (1–5 cm) and thick snow cover (> 5 cm). The snow depth values were from the Utena and Ukmergė MS in the Šventoji River Basin. The air temperature observations from these MS were used to determine wet and dry snow conditions. Snow cover was considered wet when the minimum daily air temperature (T_{min}) was positive. Snow depth and air temperature filters were applied to Sentinel-1 pixels within a 20 km radius around the MS. The R_C thresholds were determined for the seven most common CLC land cover types in the Šventoji River Basin.

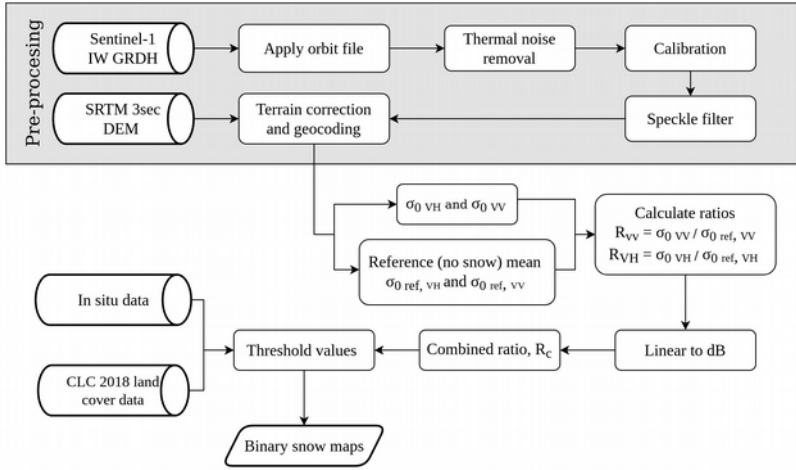


Figure 4. Sentinel-1 IW GRDH data processing scheme. The grey area indicates the SAR pre-processing steps and the white area describes how to generate binary snow cover maps.

The accuracy of the SAR R_C binary snow cover maps was evaluated using NDSI calculated from Sentinel-2 data. NDSI was derived using $0.56 \mu\text{m}$ VIS (Band 3) and $1.61 \mu\text{m}$ SWIR (Band 11) data (see Formula 1).

The downscaling of the HSAF SWE dataset was performed under the assumption that the relationship between snow cover parameters and other environmental factors (elevation, land cover type and minimum air temperature) remains constant at a large and a small scale. The downscaling was performed using two models: multiple linear regression (MLR) and artificial neural network (ANN). SWE measurements from 52 MS in the Baltic states (see Figure 1) were used to train the ANN and to build MLR models and assess their accuracy. The HSAF SWE and auxiliary environmental data used in this study covered winter seasons in the period 2012–2018. Some randomly selected data was left for validation (13%) and other data points were used to build the MLR model and train ANN.

Land cover types were divided into different binary variables. Elevation, T_{\min} and HSAF SWE data were normalized using their minimum and maximum values. The binarization of land cover types and the normalization of other parameters was done because ANN performs better when all variables range from 0 to 1. A four-layer feed forward neural network with the sigmoid activation function was created for this task.

3. RESULTS

3.1. Changes in snow cover characteristics in the Baltic states in 1961–2015

According to the climate normal (1981–2010), the average number of SCD in the Baltic states was 93 (Table 3). The average number of SCD ranged from 56–70 days at the Baltic Sea coast to > 130 days in the northeastern part of the region. The average maximum snow depth varied from 15–20 cm in the coastal areas and southwestern Lithuania to > 35 cm in the hills in northeastern Latvia.

Table 3. Trends of the SCD and maximum snow depth in the Baltic states. The mean values are based on climate normals (1981–2010). Climate trends (Sen slope) were determined using data from the period 1961–2015. The statistical significance of the trend (p) is in parentheses.

Area	Average number of SCD	Trend, days / 10 years	Avg. max. snow depth, cm	Trend, cm / 10 years
Lithuania	87.7	-4.0 ($p = 0.09$)	23.3	-0.38 ($p = 0.58$)
Latvia	93.6	-4.2 ($p = 0.09$)	25.7	-0.61 ($p = 0.53$)
Estonia	97.9	-2.7 ($p = 0.19$)	26.7	-0.32 ($p = 0.68$)
Baltic States	93.4	-3.3 ($p = 0.17$)	25.0	-0.55 ($p = 0.48$)

Using the 1961–2015 MS observations, winters were grouped into five types according to their snow cover regime (Figure 5). The first type includes cold winters, with long snow cover duration and thick snow. The second type includes winters when the snow cover formed late (mostly in early January) but persisted until the end of March. The third type of winter was rare, and only the winters of 1989 and 1990 were attributed to this type. In these winters, snow cover formed very early but later on snow melted and only a short-lived snow cover was observed from January to March. The fourth type includes winters with an average number of SCD but with relatively thick snow in December and January. The fifth type of winter is characterized by very thin and short-lived snow cover and high average air temperature.

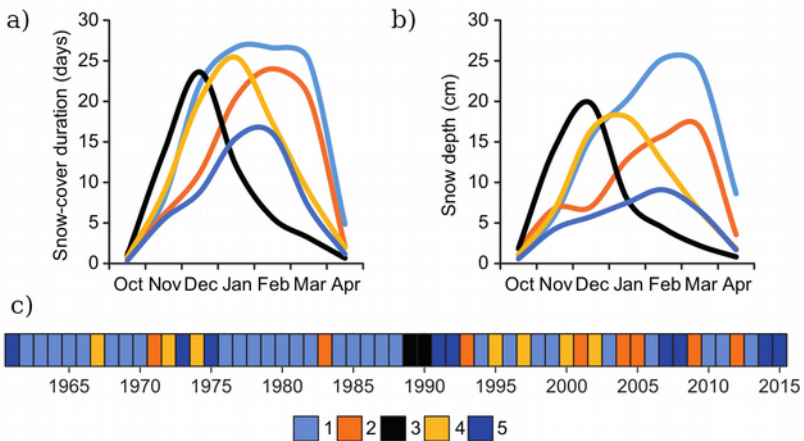


Figure 5. Snow cover regime types in the Baltic states 1961–2015: (a) average number of SCD; (b) maximum snow cover depth and (c) timeseries of snow cover regime types. Winter types (1–5) are in colour scale.

During the study period, the number of SCD in the Baltic states, on average, decreased by 3.3 days per decade ($p = 0.17$) (Figure 6). The largest changes were recorded in Latvia and Lithuania, where

the duration of snow cover decreased by 4.2 ($p = 0.09$) and 4.0 ($p = 0.09$) days per decade, respectively. A decrease in the number of SCD was statistically significant in 35% of the MS in the Baltic states (Figure 7).

During the study period, the average maximum snow depth in the Baltic states remained almost stable (Figure 7). However, in the second half of the study period, larger interannual differences were observed. There was no clear spatial distribution of maximum snow depth trends (Figure 7). A statistically significant ($p < 0.05$) maximum snow depth decreasing trend (-4.8 cm per decade) was determined only at the Väike-Maarja MS in northern Estonia.

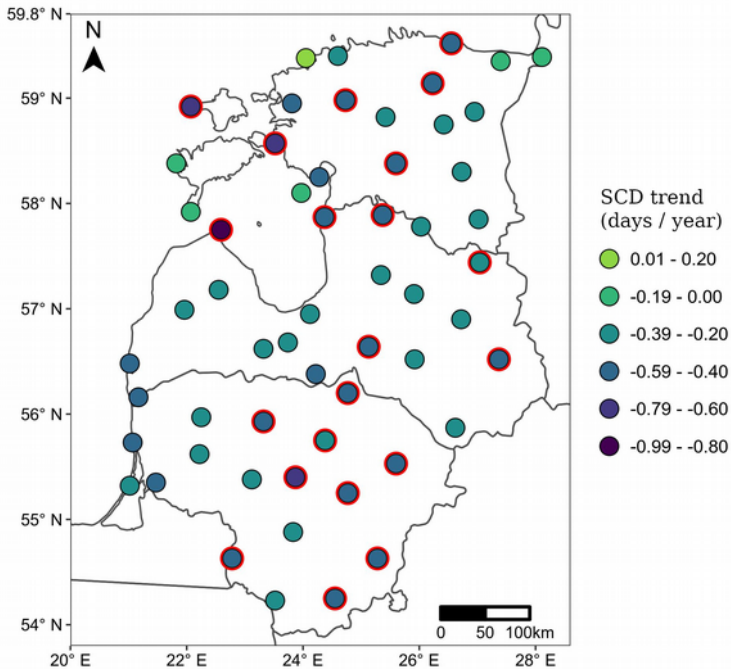


Figure 6. Changes in the number of SCD in the Baltic states during 1961–2015. Stations with statistically significant changes ($p < 0.05$) are circumscribed in red.

The observation that the maximum snow cover depth in the Baltic states did not change significantly, even as the average winter temperature rose rapidly (Jaagus et al., 2014), can be explained by the fact that the maximum snow depth in the Baltic states is not necessarily the result of snow cover accumulation. It may be a short-term phenomenon associated with intense precipitation, rather than with the average conditions of the winter season.

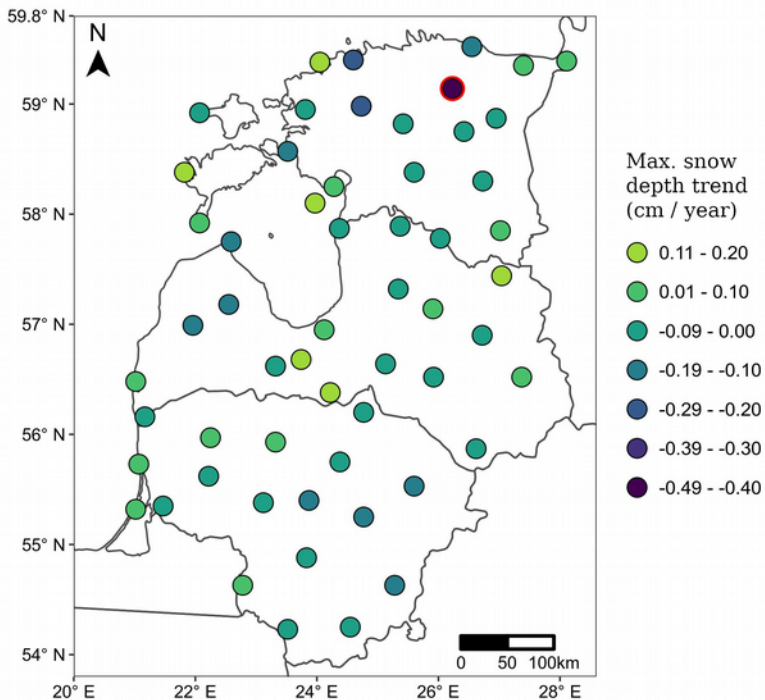


Figure 7. Changes in the average maximum snow cover depth in the Baltic states in 1961–2015. A station with statistically significant changes ($p < 0.05$) is circumscribed in red.

Analysis of the relationship of snow cover parameters with other meteorological variables showed that the SCD is mostly related to the average air temperature in November–March ($r = -0.94$; $p <$

0.00001). The maximum snow depth was related to the air temperature during the snow accumulation period (December–February) ($r = -0.74$; $p < 0.00001$) (Figure 8). Monthly total precipitation did not have a significant effect on the number of SCD. A statistically significant negative correlation ($r = -0.32$; $p = 0.017$) was found only in February. The relationship between amount of monthly precipitation and maximum snow cover depth was insignificant.

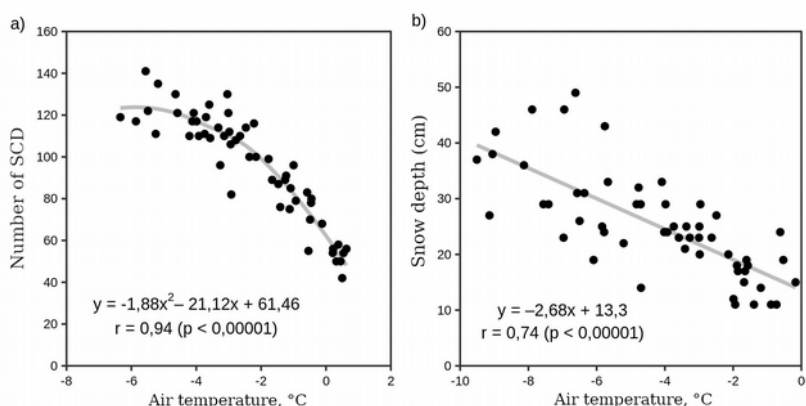


Figure 8. Relationship between snow cover variables and air temperature in the Baltic states 1961–2015: (a) SCD in November–March and (b) maximum snow depth in December–February.

The spatial distribution of snow cover variables in the Baltic states is highly influenced by the distance to the Baltic Sea. The central part of the Baltic Sea is oriented in the south-north direction, so the longitude of the MS is a good approximation of the physical distance to the main part of the Baltic Sea. The correlation coefficient between the longitude of MS and the number of SCD was 0.79 ($p < 0.00001$). This relationship was strongest in December–February (Table 4) when the thermal contrast between the sea and the land was the highest. In these months, there was a clear meridian distribution of SCD. Correlation to the distance to the physical

coastline was not strong (Table 4), as the eastern part of the Gulf of Finland and the Gulf of Riga often freeze in winter, which reduces the thermal contrast between the sea and the mainland.

A statistically significant relationship between snow cover characteristics and latitude was established in October–November and in March–April. During these periods, there was a large south–north air temperature gradient in the region. Another important geographical factor influencing the duration of the snow cover is elevation ($r = 0.52$; $p = 0.00005$) (Table 4). The elevation differences of the analysed MS were not large (the highest measurement point was in Alūksne, Latvia, 197 m). In areas with higher elevation, the air temperature was lower and seasonal snow cover tended to form earlier and melt later.

Table 4. Correlation between the SCD and the maximum snow depth with geographical parameters in the Baltic states in 1961–2015. The first number is SCD and the second number is maximum snow depth. Statistically significant ($p < 0.05$) correlation coefficients are in bold.

Month	Latitude	Longitude	Absolute height	Distance to the coastline
October	0.48/0.48	0.67/0.55	0.35/0.19	0.04/-0.16
November	0.25/ 0.61	0.60/0.59	0.54/0.17	0.23/-0.16
December	0.14/ 0.38	0.79/0.54	0.48/0.31	0.47/0.00
January	0.16/ 0.33	0.80/0.50	0.63/0.40	0.50/0.06
February	0.39/0.40	0.81/0.66	0.52/0.47	0.33/0.15
March	0.64/0.55	0.73/0.71	0.35/0.39	0.07/0.09
April	0.63/0.65	0.51/0.58	0.24/ 0.27	-0.11/-0.07
Annual	0.40/0.35	0.79/0.54	0.52/0.41	0.32/0.07

3.2. Snowfall detection using ATMS data and validation with precipitation observations at ground stations

To determine which ATMS spectral bands were most suitable for snow detection, correlation matrices were generated between ATMS brightness temperature and the GFS global weather model data. At the atmospheric window bands (53.6 GHz and 88.2 GHz) ATMS data correlated with atmospheric temperature (at levels 300–100 hPa) and Earth surface parameters (air temperature and dew point at 2 m height, as well as SWE). However, in cold weather ($T \leq -10$ °C), the correlations between ATMS 165–183 GHz brightness temperature and CLWMR, CWAT and VVEL variables became insignificant. These results are consistent with other studies, which suggested that, in cold weather, water vapour content in the atmosphere is low, and it does not emit enough microwave radiation for ATMS to detect (Kongoli et al., 2015; Meng et al., 2017).

For snowfall detection, ATMS spectral bands with high correlation to tropospheric temperature and Earth surface conditions (50.3–53.6 GHz or Bands 3–6), as well as atmospheric moisture content and vertical velocity (165.6–183 GHz or Bands 17–20), were selected. However, the brightness temperature distributions of these bands during the snowfall, rain and no precipitation events were overlapping (Figure 9). In the case of rain and snowfall, 44–85% of the brightness temperature data were overlapping, and in the case of snowfall and no precipitation events, overlap was 31–67%.

Based on the selected eight most important ATMS and GFS variables (see Methodology), the RF and logistic regression snowfall classification models were developed. The derived logistic regression:

$$\ln\left(\frac{P}{1-P}\right) = 28,17 + 0,0678 * BT_{ch5} + 0,0004 * BT_{ch17} - 0,1399 * BT_{ch18} - 0,0399 * BT_{ch19} + 0,0421 * RH_{800} + 1871,4 * CLWMR_{750} + 1,556 * CWAT - 0,2815 * PWAT \quad (3)$$

where P is snowfall probability; BT_{ch5} is 52.8 GHz brightness temperature (K); BT_{ch17} is 165.5 GHz brightness temperature (K);

BT_{ch18} is 183.3 ± 7.0 GHz brightness temperature (K); BT_{ch19} is 183.3 ± 4.5 GHz brightness temperature (K); RH800 is relative humidity at 800 hPa (%); CLWMR750 is cloud water mixing ratio at 750 hPa (kg/kg); CWAT is cloud water content (kg/m^2); and PWAT is precipitable water (kg/m^2).

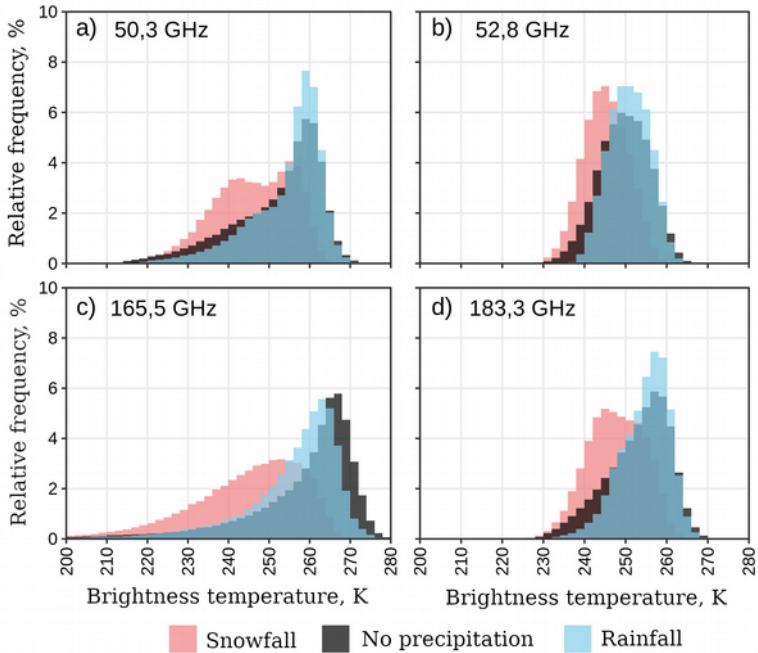


Figure 9. Distribution of ATMS bright temperature values during rainfall, snowfall and no precipitation events at different frequencies: (a) 50.3 GHz; (b) 52.8 GHz; (c) 165.5 GHz and (d) 183.3 GHz.

The snowfall classification results by RF and logistic regression models were validated against MS precipitation type observations and the ATMS snow detection algorithm currently used by the NOAA (Kongoli et al., 2018) (Table 5). The probabilities of snowfall detection for RF and logistic regression were higher (POD = 0.80–0.83) than in the study by Kongoli et al. (2018) (POD = 0.72).

However, the probability of false detection was higher (POFD = 0.23–0.24), and it indicates that models developed in this study tend to overestimate the probability of snowfall. It was also found that, in cold weather conditions ($T < -10\text{ }^{\circ}\text{C}$), ATMS data were not able to detect snowfall accurately. In cold weather, POD decreased to 0.61–0.68 and the POFD increased to 0.26–0.32 (Table 5).

Table 5. Validation scores of snowfall detection using ATMS and GFS data and applying random forest (RF) and logistic regression models. Results in italics show results in cold weather conditions ($T < -10\text{ }^{\circ}\text{C}$). Validation results from the study by Kongoli et al. (2018) are presented for reference.

Method	Predictors	ACC	POD	POFD	HSS
RF	<u>ATMS</u> : BT _{ch5} , BT _{ch17} , BT _{ch18} , BT _{ch19} <u>GFS</u> : RH800, CLWMR750, CWAT, PWAT	0.83 <i>0.68</i>	0.88 <i>0.68</i>	0.23 <i>0.32</i>	0.65 <i>0.35</i>
Logistic regresion	<u>ATMS</u> : BT _{ch5} , BT _{ch17} , BT _{ch18} , BT _{ch19} <u>GFS</u> : RH800, CLWMR750, CWAT, PWAT	0.80 <i>0.68</i>	0.84 <i>0.61</i>	0.24 <i>0.26</i>	0.60 <i>0.35</i>
Hybrid (Kongoli et al., 2018)	<u>ATMS</u> : BT _{ch16} , BT _{ch17} , BT _{ch18} , BT _{ch19} , BT _{ch20} , BT _{ch21} , BT _{ch22} <u>GFS</u> : RH, TEMP, VVEL (1, 2, 3 km), Cloud base and Cloud top height	0.79	0.72	0.17	0.55

The RF snowfall detection model was tested using several intensive snowfall cases in the US in 2014–2018 (Figure 10). Case study examples show that the RF model correctly identified the areas where intense precipitation was recorded (80–100% probability of snowfall was determined). However, not all the observed precipitation was snow. During all analysed cases, the RF model

determined a 50–80% probability of snowfall in some regions where no precipitation was observed. This coincides with validation statistics (POFD = 0.23), which shows that the RF model tends to overestimate snowfall. Spatial errors of the snowfall detection model can be related to the fact that selected ATMS and GFS model variables describe the mean conditions of the troposphere column and are not very suitable for defining cloud boundaries and thickness.

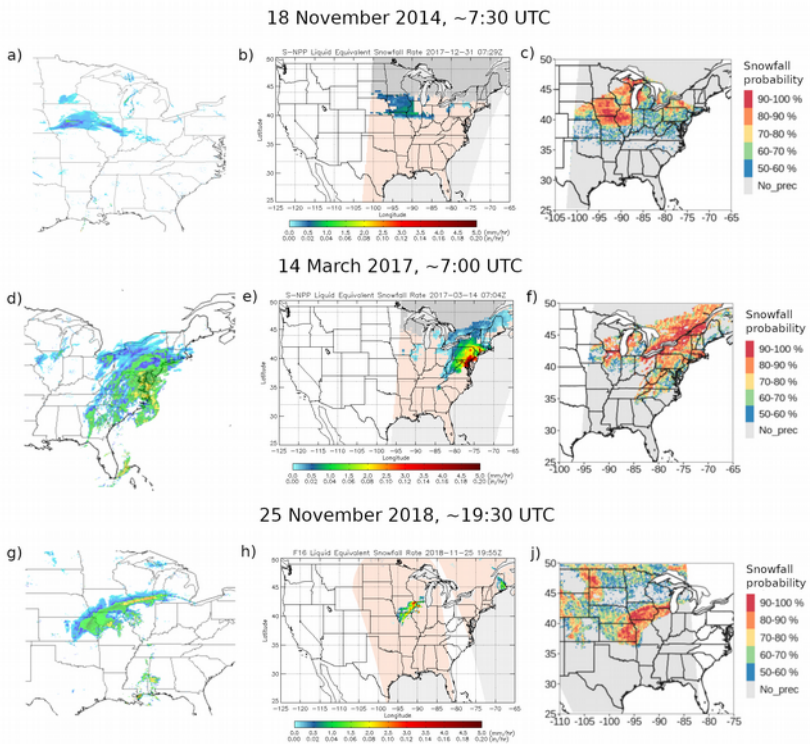


Figure 10. Case studies of random forest (RF) snowfall classification model application. The first column (a, d, g) is a ground radar image (colours show precipitation intensity); second column (b, e, h) – snowfall intensity determined by the hybrid model currently used by the NOAA (mm/h); and third column (c, f, j) – snowfall probability (%) by RF snowfall classification model.

3.3. Number of snow cover days in Lithuania derived from MODIS data

In 2012–2018, according to MOD10A1 (“Terra” satellite) and MYD10A1 (“Aqua” satellite) snow products, 76–78% of the time the Earth’s surface in Lithuania was covered with clouds (Table 6). A combination of Aqua and Terra MODIS data (Step I) and spatial filtering by eight neighbouring pixels (Step II) allowed one to reduce the number of cloud-covered pixels by 11–13% (Table 6). The remaining data gaps were filled using temporal filtering (Step III). A daily minimum air temperature filter was used to reduce the number of false snow cover identification cases (Step IV). It was determined that the 2.0 °C temperature threshold was most effective in reducing the number of snow classification errors.

Table 6. The number of cloud-covered pixels in the MODIS snow cover product at different data processing steps. The date in brackets indicates winter, when the minimum or maximum value of cloud cover was recorded.

Stage	Mean cloud cover, %	Cloud cover minimum, %	Cloud cover maximum, %	Longest continuous period with clouds, days
“Aqua”	78	74 (2006/2007)	84 (2016/2017)	89 (2010/2011)
“Terra”	76	70 (2010/2011)	83 (2016/2017)	58 (2017/2018)
I step	71	64 (2010/2011)	78 (2016/2017)	45 (2009/2010)
II step	65	58 (2010/2011)	72 (2016/2017)	35 (2003/2004)

The overall accuracy of the final cloud-free MODIS snow cover product (ACC = 0.89) was lower than the original Aqua and Terra products (ACC = 0.92–0.95), but the false alarm ratio (FAR) was better (Table 7). Validation results show that cloud removal from the

MODIS snow cover product did not affect the accuracy of the data and increased the availability of surface observations by more than four times (Table 7).

Table 7. Validation results at different stages of the MODIS snow cover product post-processing. N – total number of cloud-free pixels; ACC – accuracy; CSI – critical success index; FAR – false alarm ratio; FBI – frequency bias; POD – probability of detection.

Stage	N	ACC	CSI	FAR	FBI	POD
“Aqua”	12442	0.92	0.77	0.19	1.15	0.93
“Terra”	13656	0.95	0.85	0.12	1.09	0.96
Step I	16917	0.92	0.78	0.18	1.14	0.94
Step II	19983	0.90	0.74	0.22	1.21	0.94
Step III	57732	0.82	0.65	0.29	1.24	0.89
Step IV	57732	0.89	0.74	0.16	1.03	0.86

The generated daily cloud-free MODIS snow cover product allowed one to derive an annual and monthly number of SCD in Lithuania (SCD_{MODIS}) (Figure 11a). Ground-based SCD maps (SCD_{MS}) were generated using snow cover observations from 17 MS and applying bicubic spline interpolation. Both SCD_{MS} and SCD_{MODIS} maps showed that the duration of snow cover in Lithuania increases from west to east and SCD is closely related to the distance from the sea and elevation of the terrain (Figure 11c).

The largest negative differences between SCD_{MODIS} and SCD_{MS} were observed on the furthest eastern corner of Lithuania and on the northwestern border with Latvia. In the highlands and large forested areas, more SCD were recorded using MODIS data compared to the interpolated MS data (Figure 11b). The snow observation network in Lithuania is sparse; therefore, interpolation of MS data causes large

(> 10 days) differences in the annual SCD number between satellite and in situ data on the border regions and in forested areas. In general, MODIS tended to overestimate the number of SCD, and in 74.0% of the territory, the mean annual SCD_{MODIS} was higher than SCD_{MS} by more than 4 days (Figure 11d). Only in 2.7% of the Lithuanian territory (mainly at the eastern and northwestern borders), SCD_{MODIS} was more than 4 days lower than SCD_{MS} .

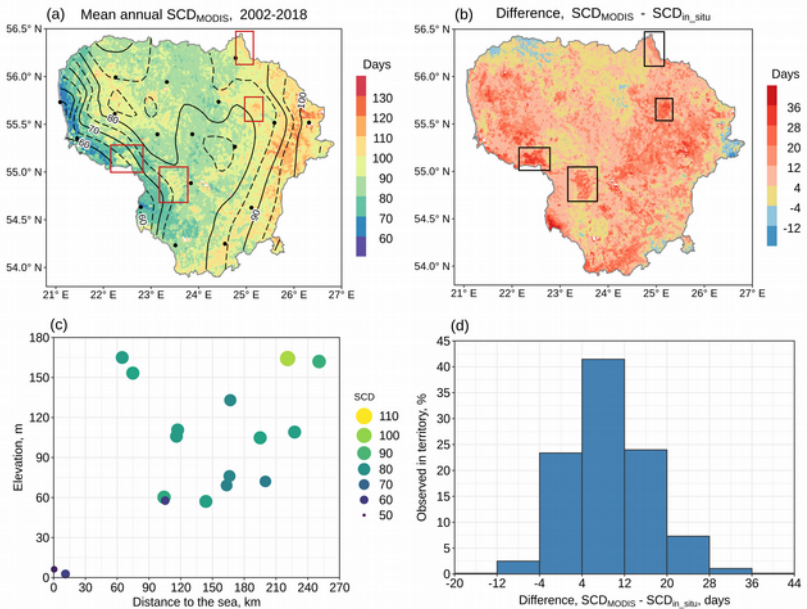


Figure 11. Comparison of average SCD_{MODIS} and SCD_{MS} in Lithuania in 2002–2018: (a) SCD_{MODIS} is in colour scale, SCD_{MS} in isolines, points indicate MS, and some large forest areas are marked with rectangles; (b) map of differences ($SCD_{MODIS} - SCD_{MS}$); (c) relationship of SCD_{MS} with elevation and distance to the Baltic Sea; (d) histogram of differences between SCD_{MODIS} and SCD_{MS} .

The average correlation coefficient between annual SCD_{MODIS} and SCD_{MS} was 0.93, indicating that MODIS can accurately capture the interannual variability of snow cover duration. The average absolute

difference between annual SCD_{MODIS} and SCD_{MS} was 8.5 days, but differences were larger in the coastal areas and smaller in the eastern and northern parts of Lithuania. This can be attributed to the maritime climate of the coastal areas and snow cover being short-lived.

The largest average absolute differences between the monthly SCD_{MODIS} and SCD_{MS} were determined in December. At the beginning of winter, thaws are frequent and cloudy conditions prevail (82% of the time the study area was covered with clouds). Because of these conditions, large biases can be expected when using MODIS data. This is most relevant on the Baltic Sea coast. According to the generated cloud-free MODIS data, the probability of snow cover in the Klaipėda MS in December was twice as high as based on MS observations. In January and February, the differences between SCD_{MODIS} and SCD_{MS} were smaller, as the average air temperature became negative and snow cover remained stable (except for the coast). In March, when the spring snowmelt started, the differences between the monthly SCD_{MODIS} and SCD_{MS} increased again. In April, SCD_{MODIS} and SCD_{MS} were more similar because there were more clear sky surface observations (the study area was covered with clouds 57–60% of the time).

The largest differences in monthly SCD_{MODIS} and SCD_{MS} (12–24 days) were found in 0.4% of pixels. All these pixels were close to the lakes or in the valleys of big rivers. These errors could be caused by the inaccuracies of the land/water mask used in the Terra and Aqua MODIS algorithms (Carroll et al., 2017). Another possible reason is that ice on the river banks and lake shores are classified as snow in the MODIS algorithm and cause large SCD_{MODIS} overestimation errors in spring.

3.4. Snow cover detection in Šventoji River Basin using SAR data

Analysis of Sentinel-1 SAR backscattering ratios (R_C) showed that, in all land cover types, the values were lower when snow cover was thick (> 5 cm) compared to the no snow cases (Figure 12). The lowest critical R_C value separating the snow and no snow cases was determined in urban areas (-0.52 dB) and the highest was in the wetlands (-1.05 dB). In other land cover types, the critical R_C values ranged from -0.78 dB to -0.96 dB (Figure 12). When all land cover types were analysed together, the critical R_C value was -0.78 dB. The highest overlap of thick snow cover and no snow R_C value distributions was determined over water bodies (50.2%), and the lowest was in broadleaf/mixed forests (20.7%).

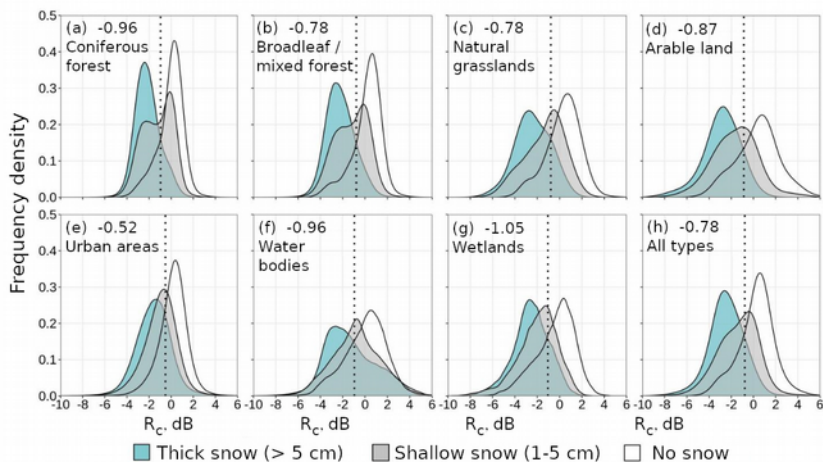


Figure 12. Distributions of R_C values in different land cover types: blue – cases with thick snow cover (> 5 cm), grey – cases with thin snow cover (1–5 cm) and white – no snow cases. The dotted line marks the intersection of thick snow cover and no snow R_C distributions, and it is considered a snow classification threshold in a particular land cover type.

Accuracy of the Sentinel-1 R_C snow classification was assessed using Sentinel-2 NDSI snow cover maps (Sentinel-2_{NDSI}). In the analysed period of 2014–2019, there were only two cases when both satellites passed over the Šventoji River Basin on the same day with low cloudiness: 26 February 2018 and 28 March 2018 (Figure 13). Using these cases for validation showed that the overall accuracy (ACC) of the Sentinel-1 R_C snow mapping was 0.81, the POD was 0.95 and the FAR was 0.29 (Table 8). These statistical scores indicated that the accuracy of snow cover mapping using R_C thresholds was average. A high FAR value and a frequency bias larger than 1 (FBI = 1.34) indicated that the R_C method tended to overestimate snow cover compared to the Sentinel-2_{NDSI}. R_C snow maps were more accurate in the open areas (grasslands and arable land). The worst R_C validation rates were determined in the forested areas. In the coniferous forests, the accuracy of the R_C snow cover map (ACC) was 0.51, POD = 0.96, FAR = 0.86 and HSS = 0.12. Very high FBI and FAR values suggest that Sentinel-1 R_C snow maps may vastly overestimate the presence of snow in the coniferous forests (Table 8).

Table 8. Validation results of R_C snow cover classification using binary Sentinel-2_{NDSI} snow maps as a reference.

Land cover type	ACC	POD	FAR	FBI	HSS
Broadleaf/mixed forests	0.78	0.98	0.39	1.61	0.58
Coniferous forests	0.51	0.96	0.86	6.72	0.12
Urban territories	0.80	0.89	0.23	1.16	0.61
Grasslands	0.92	0.99	0.13	1.13	0.83
Water bodies	0.58	0.54	0.10	0.59	0.18
Wetlands	0.80	0.84	0.26	1.14	0.60
Arable land	0.85	0.99	0.22	1.26	0.71
<i>All types</i>	<i>0.81</i>	<i>0.95</i>	<i>0.29</i>	<i>1.34</i>	<i>0.63</i>

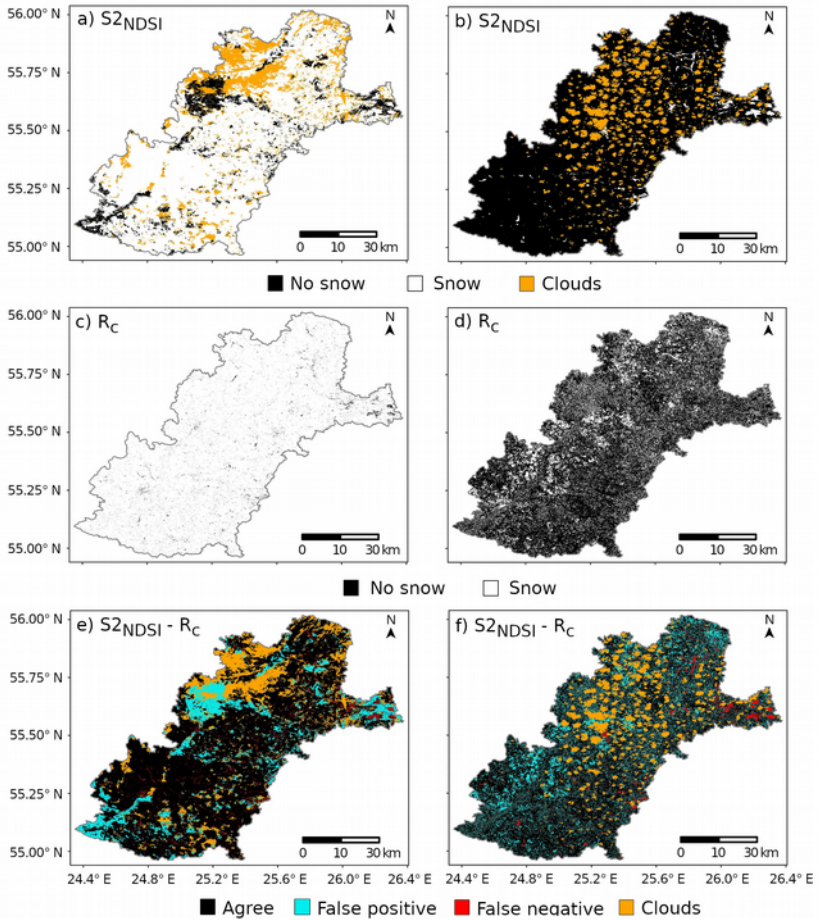


Figure 13. Snow cover maps in the Šventoji River Basin: first column (a, c, e) is 26 February 2018 case and second column (b, d, f) is 28 March 2018. (a) and (b) are Sentinel-2_{NDSI} snow maps; (c) and (d) are R_C snow maps; (e) and (f) are difference maps (Sentinel-2_{NDSI} - R_C).

When using one R_C threshold (-0.78 dB) for all land cover types, the validation results were similar to those discussed above: ACC = 0.80, POD = 0.95 and FAR = 0.32. Heidke's score (HSS) was 0.61

(Table 8). It can be argued that using a single universal R_C threshold provides sufficiently accurate results and can simplify the snow cover mapping algorithm.

Errors in SAR R_C snow cover classification can be caused by several factors, including variations of soil conditions, variations of snow grain size and metamorphism of different snow cover layers (Baghdadi et al., 1999, Dedieu et al., 2014; Paloscia et al., 2017). SAR signal noise can also reduce the accuracy of snow cover mapping (Tsai et al., 2019). Snow cover classification errors in the analysed Šventoji River Basin can also be caused by thin snow and frequent thaws. In very thin snow cover, the main contribution to the SAR signal backscattering comes from signal interaction with soil, rather than with snow cover.

3.5. Downscaling of the satellite SWE data

Downscaling of the HSAF SWE product was done using MLR and ANN models, SWE data from 52 MS and auxiliary spatial DEM, land cover and minimum daily air temperature data in the Baltic states.

Using the variable importance criteria of the ANN model, the six most significant SWE predictors were selected. The ANN model was trained using 500 epochs to avoid overfitting of the model. Using the chosen environmental predictors, the MLR model was also derived:

$$\begin{aligned} VAS_{LRM} = & -0,0027 + 0,7169 * SWE_{HSAF} + 0,0089 * elev - \\ & -0,0019 * conif_f - 0,0061 * mixed_broad_f - \\ & -0,0013 * arable_l - 0,0025 * urban_l \end{aligned} \quad (4)$$

where SWE_{HSAF} – HSAF SWE value (mm); *elev* – height based on SRTM digit elevation model (m); *conif_f* – coniferous forests (0/1); *mixed_broad_f* – mixed and broadleaf forests (0/1); *arable_l* – arable land (0/1); *urban_l* – urbanized areas (0/1). The coefficient of determination (R^2) of the derived MLR was 0.89.

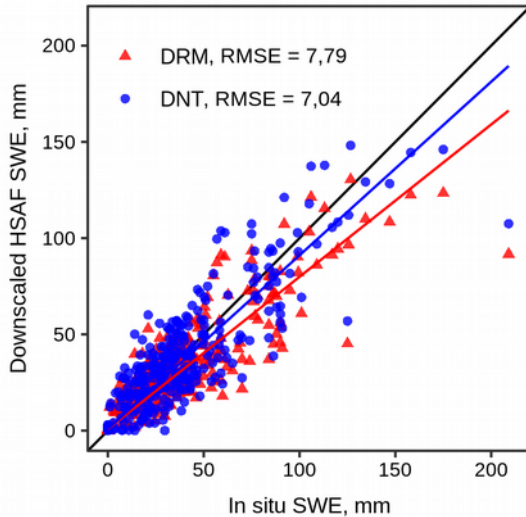


Figure 14. Comparison of downscaled HSAF SWE data using ANN and MLR models and SWE measurements during the snow surveys.

Using ANN and MLR models, downscaled HSAF SWE data was compared to SWE observations from snow surveys in the Baltic states in 2017–2018. The validation results showed that the SWE values generated from the ANN and MLR models were similar (Figure 14). Downscaled SWE data obtained using ANN was slightly more accurate (RMSE = 7.04; ME = 0.03) when compared to MLR output (RMSE = 7.79; ME = -1.08). In both cases, the correlation coefficient between the downscaled HSAF SWE data and in situ observations was 0.94.

Although the validation statistics indicate that the downscaling of SWE using the ANN model is more accurate, analysis of spatial distribution revealed that, in some cases, ANN create snow (SWE > 0 mm); it was not determined in the original HSAF product (Figure 15c). The main reason for these errors is that land cover types in the ANN model had higher relative importance than HSAF SWE data. Therefore, in some cases, even if HSAF SWE was 0 mm, values of

other parameters (elevation, land cover type and T_{\min}) determined that the DNT model output was higher than 0 mm. To avoid these undesirable results of the ANN model, more weight should be given to the HSAF SWE. Another option could be using other ANN activation functions (e.g., ReLU – Rectified Linear Unit or TLU – Threshold Logic Unit).

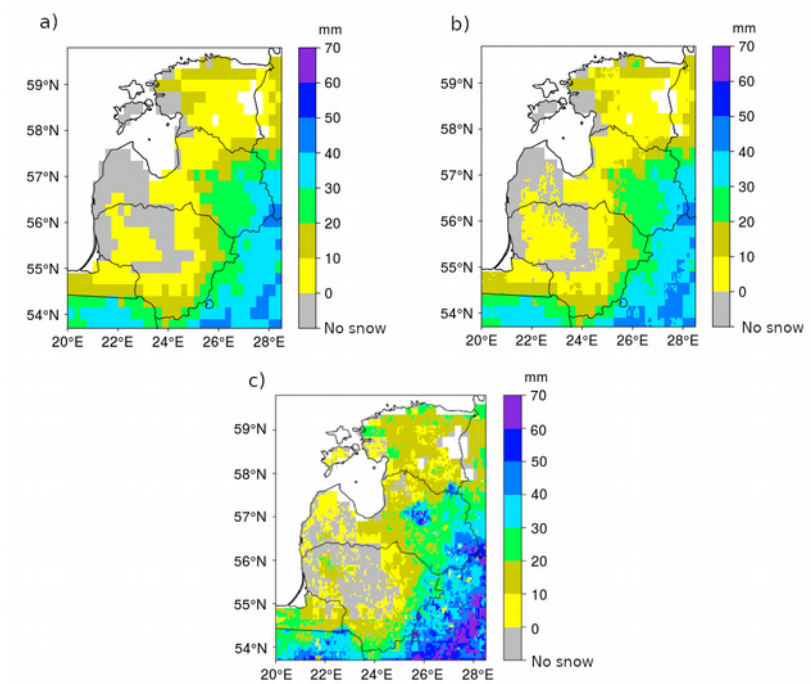


Figure 15. SWE in the Baltic states on 10 February 2017: (a) original $0.25^\circ \times 0.25^\circ$ HSAF SWE product; (b) downscaled SWE using MLR model; (c) downscaled SWE using ANN model ($0.05^\circ \times 0.05^\circ$).

CONCLUSIONS

1. The spatial distribution of snow cover characteristics in the Baltic states is mainly determined by the elevation and longitude, which is an approximation of the distance to the central part of the Baltic Sea. During the period 1961–2015, the number of snow cover days in the Baltic states on average decreased by 3.3 days per decade. The average maximum snow depth during the same period did not change significantly.

2. The ATMS brightness temperature recorded during the snowfall at 165.5–183.3 GHz frequency is highly correlated with the moisture content and vertical air velocity in the troposphere. ATMS and GFS data were used to develop logistic regression and RF models for snowfall detection. The output from both models had a similar probability of snow detection and probability of false detection (POD = 0.80–0.83; POFD = 0.23–0.24). The RF algorithm was superior in the situations where snowfall was rare. In cold weather conditions ($T < -10$ °C), the accuracy of both snow detection models was lower.

3. Cloud gaps in the MODIS snow cover product can be effectively removed using spatial and temporal filtering and minimum daily air temperature control ($T_{\min} < 2.0$ °C). The average correlation coefficient between the annual number of SCD in Lithuania, determined using cloud-free MODIS data and MS observations, was 0.93. MODIS tended to overestimate the snow cover duration, and the average absolute difference between annual satellite and MS observations was 8.5 days. Annual and monthly MODIS snow cover data was more detailed and accurate over the forested areas and at the border regions compared to the interpolated MS data. The largest MODIS snow cover data biases were determined in the coastal areas where snow cover is often short-lived.

4. It was determined that, in Lithuania Sentinel-1 SAR backscattering changes when snow cover is thicker than 5 cm. Depending on the land cover type, the critical SAR ratio (R_C) values for snow (> 5 cm) detection ranged from -0.52 dB to -1.05 dB. The universal R_C threshold for all land cover types was -0.78 dB. Comparison of SAR R_C snow cover maps with Sentinel-2 NDSI images showed that R_C snow detection was accurate in the grasslands and arable land (POD = 0.99; FAR = 0.13–0.22). However, over the forests, especially conifer, SAR R_C snow cover detection methodology performed poorly (FAR = 0.86).

5. Downscaling of the HSAF SWE product was done using detailed auxiliary information (elevation, land cover types and minimum air temperature) and ANN and MLR models. Comparison with in situ measurements showed that ANN output SWE was more accurate (RMSE = 7.04) than using the MLR model (RMSE = 7.79). Spatial analysis of the downscaled SWE showed that, in some cases, ANN produced snow cover (SWE > 0 mm), where it was not present in the original HSAF data. The main reason for these errors was that, in the ANN model, a higher relative variable importance was assigned to land cover types, rather than the original HSAF SWE data.

LIST OF PUBLICATIONS

Publications in journals with impact factor included into *Clarative Analytics Web of Science* database:

- **Kilpys J.**, Pipiraitė-Januškienė S., Rimkus E. (2020). Snow climatology in Lithuania based on the cloud-free moderate resolution imaging spectroradiometer snow cover product. *International Journal of Climatology*, 1–17.
- Rimkus E., Briede A., Jaagus J., Stonevicius E., **Kilpys J.**, Viru B. (2018). Snow-cover regime in Lithuania, Latvia and Estonia and its relationship to climatic and geographical factors in 1961–2015. *Boreal Environment Research*, 23, 193-208.

Reviewed publications:

- Pipiraitė-Januškienė S., **Kilpys J.**, Rimkus E. (2018). Palydovinės informacijos apie sniego dangą trūkių užpildymo metodų vertinimas. *Geologija. Geografija*, 4(3), 103-112.

CURRICULUM VITAE

Name, Surname Justinas Kilpys
Birth date and place 1985-12-09, Šiauliai, Lithuania
E-mail kilpysj@gmail.com

Education

2016–2021 PhD studies in physical geography, Vilnius University
2008–2011 Masters’s degree in geography with specialization in hydrometeorology, Vilnius University
2004–2008 Bachelor’s degree in geography with specialization in hydrology and meteorology, Vilnius University

Work experience

2018–2021 Junior researcher, Institute of Geosciences, Vilnius University
2011–2017 Senior specialist, Climatology Division, Lithuanian Hydrometeorological Service
2009–2010 Environmental expert assistant, Baltic Environmental Forum, Lithuania

Other professional activities

2019–2020 Project manager of European Space Agency PECS project “University Course and Public Lectures on Earth Observations (UniEO)”.
2019 Jun-Nov Visiting researcher at Cooperative Institute for Satellite Earth System Studies (CISESS), University of Maryland, USA
2012 Jun-Dec World Meteorological Organization (WMO) fellowship at Federal Office of Meteorology and Climatology (MeteoSwiss), Switzerland

NOTES

NOTES

NOTES

Vilnius University Press
9 Saulėtekio Ave., Building III, LT-10222 Vilnius
Email: info@leidykla.vu.lt, www.leidykla.vu.lt
Print run 20