


RESEARCH ARTICLE

Open Access



Creation and environmental applications of 15-year daily inundation and vegetation maps for Siberia by integrating satellite and meteorological datasets

Hiroki Mizuochi^{1,2,3*} , Taiga Sasagawa⁴, Akihiko Ito⁵, Yoshihiro Iijima⁶, Hotaek Park^{2,7}, Hirohiko Nagano⁸, Kazuhito Ichii⁹ and Tetsuya Hiyama²

Abstract

As a result of climate change, the pan-Arctic region has seen greater temperature increases than other geographical regions on the Earth's surface. This has led to substantial changes in terrestrial ecosystems and the hydrological cycle, which have affected the distribution of vegetation and the patterns of water flow and accumulation. Various remote sensing techniques, including optical and microwave satellite observations, are useful for monitoring these terrestrial water and vegetation dynamics. In the present study, satellite and reanalysis datasets were used to produce water and vegetation maps with a high temporal resolution (daily) and moderate spatial resolution (500 m) at a continental scale over Siberia in the period 2003–2017. The multiple data sources were integrated by pixel-based machine learning (random forest), which generated a normalized difference water index (NDWI), normalized difference vegetation index (NDVI), and water fraction without any gaps, even for areas where optical data were missing (e.g., cloud cover). For the convenience of users handling the data, an aggregated product is provided, formatted using a 0.1° grid in latitude/longitude projection. When validated using the original optical images, the NDWI and NDVI images showed small systematic biases, with a root mean squared error of approximately 0.1 over the study area. The product was used for both time-series trend analysis of the indices from 2003 to 2017 and phenological feature extraction based on seasonal NDVI patterns. The former analysis was used to identify areas where the NDVI is decreasing and the NDWI is increasing, and hotspots where the NDWI at lakesides and coastal regions is decreasing. The latter analysis, which employed double-sigmoid fitting to assess changes in five phenological parameters (i.e., start and end of spring and fall, and peak NDVI values) at two larch forest sites, highlighted a tendency for recent lengthening of the growing period. Further applications, including model integration and contribution to land cover mapping, will be developed in the future.

Keywords Continental-scale water and vegetation maps, Satellite data, Pan-Arctic region, Data fusion, Machine learning, Trend analysis, Phenological feature extraction

*Correspondence:

Hiroki Mizuochi

mizuochi.hiroki@aist.go.jp

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

1 Introduction

The pan-Arctic region, which includes Siberia, is known for its extremely cold climate. However, it has recently been experiencing a higher rate of temperature increase than other regions of the Earth's surface, mainly as a result of global climate change (Hantemirov et al. 2022, Arctic Monitoring and Assessment Programme (AMAP) 2021). The higher temperatures increase the risk of climatic extremes and disasters, such as heatwave-induced wildfires (Witze 2020), thawing permafrost that promotes the release of greenhouse emissions from these areas (Yokohata et al. 2020), and river ice jams and consequent flooding (Madaeni et al. 2020; Sakai et al. 2015).

In order to evaluate the impact of climate change on the pan-Arctic region, detailed monitoring of specific environmental parameters is essential. This has been accomplished using remote sensing techniques. Water-related parameters (Suzuki and Matsuo 2019) such as total water storage, snow water equivalent (SWE), soil moisture, and surface water coverage (Velicogna et al. 2012; Suzuki et al. 2020; Yang et al. 2007; Bartsch et al. 2009; Watts et al. 2012; Mizuochi et al. 2021a) have been used to investigate local and global hydrological cycles. Another focus is vegetation parameters (Nagai et al. 2019) such as plant functional type (i.e., land cover) map, aboveground biomass, leaf area index, and growing season duration (Leroy et al. 2006; Myneni et al. 2001; Kushida et al. 2007; Buitenwerf et al. 2015), which are important for ecological monitoring and carbon cycle research. Water and vegetation parameters are closely interlinked via energy budgets and meteorological and ecological schemes in relation to evapotranspiration, albedo, vegetation growth, canopy interception of precipitation, and surface aerodynamic properties (e.g., Mizuochi et al. 2021b).

Optical remote sensing has been traditionally used in such environmental studies to monitor changes in surface water and vegetation. Band indices, such as the (modified) normalized difference water index (NDWI: Xu 2006) and the normalized difference vegetation index (NDVI: Hatfield et al. 1984; Perry and Lautenschlager 1984), are widely used to characterize spectral reflectance features and enhance the appearance of water and vegetation bodies in satellite imagery, respectively.

Another promising approach is microwave remote sensing using either active radar or passive radiometers, which has the advantage of being less sensitive to cloud cover than optical sensors. Backscatter signals observed by synthetic aperture radar (SAR) can be used to extract open water and forest areas and to estimate surface soil moisture content (Twele et al. 2016; Zakharov et al. 2020; Reiche et al. 2018). Microwave radiometers also can estimate the surface water fraction, soil moisture content

and vegetation optical depth (Fily et al. 2003; Owe et al. 2008; Moesinger et al. 2020).

When using satellite data to retrieve physical parameters, the divergent characteristics of the data and the technical difficulty in data handling often result in trade-offs between the spatial and temporal resolutions of the resulting product. For example, previous broad-scale water maps (e.g., open water and wetland distribution, water fraction within pixels) have resolution ranging from tens to hundreds of meters (Lehner and Döll 2004; Fluet-Chouinard et al. 2015; Yamazaki et al. 2015; Pekel et al. 2016). While they offer detailed spatial information, they often provide no temporal information, or if such information is provided, the time intervals are long. Conversely, maps with high temporal resolutions (i.e., daily to monthly) typically provide relatively coarse spatial information (e.g., several tens of kilometers) (Papa et al. 2010; Schroeder et al. 2015).

To overcome such trade-offs, researchers have used multiple datasets in a synergistic manner (i.e., data fusion; Belgiu and Stein 2019). There has been particular interest in combining image data obtained by optical and microwave sensors (Suzuki and Matsuo 2019; Beamish et al. 2020). This is because information that is missing from optical sensor observations due to cloud cover or insufficient solar illumination can be compensated for using data obtained by microwave sensors. Many applications of such data fusion approaches have been reported for wetland mapping (Chasmer et al. 2020; Zhang et al. 2021; Mizuochi et al. 2021a), snow and thawing detection (Armstrong and Brodzik 2001; Kim et al. 2012) and vegetation and ecosystem monitoring (Kimball et al. 2009; Mavrovic et al. 2023). These studies involved intercomparison of different data sources and/or complementary inputs for process-based models or machine learning (ML). However, there have been few studies on simultaneous mapping of areas of water and vegetation that are interlinked. In addition, the creation and handling of broad-scale datasets that combine both high temporal resolution (e.g., daily) and moderate spatial resolution (e.g., sub-kilometer scale) are challenging, although there has been one report on the creation of daily maps of soil moisture content with a spatial resolution of 1 km (Zhu et al. 2023).

The goal of the present study was to produce simultaneous continental-scale inundation and vegetation maps for Siberia. By integrating optical and microwave satellite data with other ancillary data, maps were obtained with a daily temporal resolution and a moderate spatial resolution of 500 m. In this paper, the map generation process and the characteristics and utility of the created products are described in detail. It is expected that this data will provide fundamental information that can be used

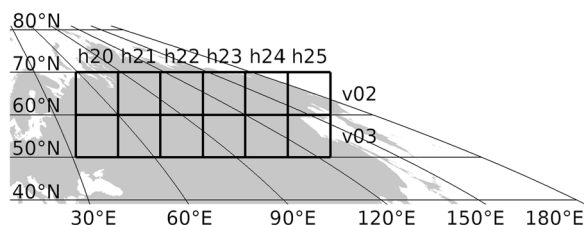


Fig. 1 MODIS tiles used in this study. The tiles (h20–h25, v02–v03) were superimposed on a sinusoidal map projection of the study area

to analyze the impacts of climate change on the risk of disasters, and also to clarify changes in water and carbon cycles in the pan-Arctic region. As example applications, we performed a time-series analysis of the NDWI and NDVI on a continental scale and extracted phenological features at two experimental sites.

2 Methods

2.1 Target area and product specifications

The target is a wide area of Siberia, approximately corresponding to 50–70° N and 31–180° E at maximum (Fig. 1). Considering potential applications (i.e., large-scale monitoring and model integration), computational resources and the spatial coverage of permafrost over Siberia (Shestakova et al. 2021), the specifications of our product are shown in Table 1.

To meet these requirements, we relied mainly on the Moderate Resolution Imaging Spectroradiometer (MODIS) to identify vegetation and water distributions at optimal spatiotemporal resolutions. In this way, the spatiotemporal resolution of our target product is consistent with those of MODIS: the original spatial resolution was 500 m, and the temporal resolution was daily. The product duration (2003–2017) was determined after considering the temporal characteristics of all of the data sources (see Sect. 2.2.1).

In polar regions in particular, commonly used geographic projections (i.e., latitude/longitude) exaggerate distances along the longitude direction, so distorting the shape of objects on the map and increasing pixel-based computational costs. Consequently, in the present study, data fusion processing was conducted using a MODIS sinusoidal projection (Fig. 1), which is originally provided at a resolution of 500 m. For convenience, an aggregated product using a latitude/longitude projection with a 0.1° grid resolution is also provided as a postprocessing output (see Sect. 2.2.5).

2.2 Algorithm and data sources

The basic algorithm used for map generation was previously developed by us for use in a limited study area

(Mizuochi et al. 2021a). In the ML-based algorithm used in this study, a pixel-based random forest (RF; Breiman 2001) approach is used to fill observation gaps in MODIS data (explained variable) by referring to other matching data sources (explanatory variable). Subsequently, bias correction is applied using a conditional generative adversarial network (pix2pix; Mirza and Osindero 2014; Isola et al. 2017). This study expands the previous study area to a pan-Arctic scale and includes an update to the NDVI map, the details of which are summarized in Table 1. To conserve computational resources, no pix2pix bias correction was used, as the procedure only improved the RMSE accuracy by ~1% (Mizuochi et al. 2021a).

The overall procedure involves (1) data download, (2) preprocessing including extraction of feature variables, map translation and coregistration, (3) ML training and prediction, and (4) postprocessing including validation, visualization, creation of the water fraction maps, and product format conversion.

Freely available optical and microwave satellite data, together with reanalysis data, were downloaded from the relevant websites (see the ‘Availability of data and materials’ section). Preprocessing and postprocessing treatments that include geospatial analysis were implemented mainly using GRASS GIS (ver. 7.8.2) and QGIS (3.10.4). ML processing was implemented using Python 3.8.10 with PyCaret 2.3.10 for preliminary model selection and hyperparameter tuning, scikit-learn 1.2.1 was used for ML training and prediction, and ray 2.2.0 was used for multiprocessing. Detailed explanations of each step are provided in the following subsections.

2.2.1 Data preparation

The data sources and physical parameters were similar to those used by Mizuochi et al. (2021a); however, vegetation optical depth (VOD) data (Moesinger et al. 2020) were newly added since they may be useful for NDVI prediction. The data sources and retrieved physical parameters are summarized in Table 2.

Surface reflectance data obtained by the MODIS instrument on the Aqua satellite (MYD09GA, collection 6.1) were used to calculate the NDWI and NDVI as follows:

$$\text{NDWI} = \frac{G - \text{SWIR}}{G + \text{SWIR}} \quad (1)$$

$$\text{NDVI} = \frac{\text{NIR} - R}{\text{NIR} + R} \quad (2)$$

where G, R, NIR, SWIR are the green, red, near infrared, and short-wave infrared surface reflectance, which correspond to MODIS band 4 (545–565 nm), band 1

Table 1 Specifications of developed product

| Items | This study | Previous study |
|---------------------------------|---|---|
| Spatial resolution | 500 m (original sinusoidal output) 0.1° (aggregated lat–long map) | 500 m |
| Observation frequency | Daily | Daily |
| Spatial extent | Latitude: 50.00–70.00°N, Longitude: approx. 31.12–180.00°E (Corresponding to MODIS sinusoidal tile v02–03, h20–25) | Latitude: 62.02–63.08°N, Longitude: 129.32–130.38°E |
| Duration | 2003.1.1–2017.12.31 | 2012.3.7–2018.3.31 |
| Number of explanatory variables | 23 | 4, 6, or 23 |
| Output parameters | NDWI, NDVI, water fraction | NDWI, water fraction |

For user convenience, both original resolution maps on sinusoidal projection (i.e., MODIS tile base) and merged maps on latitude–longitude projection are provided. Differences from our previous product (Mizuochi et al. 2021a, b) are also described

MODIS: Moderate Resolution Imaging Spectroradiometer, NDWI: normalized difference water index, NDVI: normalized difference vegetation index

(620–670 nm), band 2 (841–876 nm), and band 7 (2105–2155 nm), respectively.

Cloud, cloud shadow, snow, and cloud-adjacent pixels were screened out (i.e., treated as missing data to be filled by ML) by referring to accompanying quality assurance data (“state_1km” subset with bitmask that indicates cloud and snow states). Also, pixels recorded during shadow time in the high-latitude winter were also treated as missing data to be filled using ML. The quality controlled NDWI and NDVI were the explained variables to be predicted from the following low-resolution explanatory variables.

Level-3 brightness temperature data, obtained by the Advanced Microwave Scanning Radiometer-EOS (AMSR-E) on the Aqua satellite and AMSR-2 on the Global Change Observation Mission-Water (GCOM-W) satellite, were used to calculate the polarization index (e.g., Moradzadeh and Srivastava 2021; Sawada et al. 2017) and surface water fraction. These indices are sensitive to soil moisture and/or abundance of surface water. The calculation method employed for the indices was the same as that in Mizuochi et al. (2021a). In addition, the SWE product retrieved from the AMSR series was used to determine snow masks. To maintain consistency with the AMSR series (on the Aqua satellite or a satellite with a similar overpass time), the other MODIS product on the Terra satellite platform (MOD09GA) was not used.

Sixteen meteorological parameters provided by ERA5-land (ECMWF 2023) along with the VOD product (Moesinger et al. 2020) estimated from multiple microwave satellite data sources (i.e., SSM/I, TMI, AMSR-E, AMSR2 and WindSat), were also used. These were selected because they were likely to be sensitive to the water and vegetation conditions on the surface. In addition, day of year (DOY) information was used to represent the season. To ensure continuity between the end of

one year and the beginning of the next, DOY was calculated in cosine and sine forms as follows:

$$\text{DOY}_{\cos} = \cos\left(2\pi \times \frac{\text{DOY}}{365.25}\right) \text{ and } \text{DOY}_{\sin} = \sin\left(2\pi \times \frac{\text{DOY}}{365.25}\right) \quad (3)$$

In total, 23 low-resolution maps (derived from AMSR series, ERA5-Land, VOD and DOY) were assumed to be the common explanatory variables to predict the two explained variables (MODIS NDWI and NDVI), since the NDWI (water) and NDVI (vegetation) are likely to be closely interlinked.

2.2.2 Preprocessing

Preprocessing consisted of calculating the abovementioned variables, converting map projections, coregistering maps, and performing climatology calculations. All of the processes were managed on individual MODIS sinusoidal tiles, which encompassed 12 tiles in total (vertical tile numbers 02–03 and horizontal tile numbers 20–25). First, the MODIS HDF files were converted into raw binary images containing 2400 × 2400 pixels using a sinusoidal projection with the HDF-EOS-to-GeoTIFF-conversion tool (HDF-EOS Tools and Information Center 2023). Using GIS software, data from other data sources (i.e., explanatory variables) were processed to align them with the MODIS sinusoidal image. This involved resampling, cropping (i.e., cutting out the region of interest), and converting different data formats and on different projections into raw binary images with matching locations and dimensions (i.e., coregistration). Importantly, the process employed nearest-neighbor resampling, which involves oversampling (i.e., dividing a coarse pixel into multiple pixels while retaining the same pixel value). This step was necessary as the other data sources have lower resolution than MODIS.

Table 2 Data source and physical parameters for data fusion

| Parameter number | Data name | Spatial resolution | Observation frequency | Duration | Source | Physical parameters |
|------------------|---------------|--------------------|-------------------------------------|---|-------------------------|--|
| 1 | MODIS | 500 m | Daily | 2002–present | USGS (2023) | NDWI |
| 2 | MODIS | 500 m | Daily | 2002–present | USGS (2023) | NDVI |
| 3 | AMSR-E, AMSR2 | 10 km | Daily | 2002–2011 (AMSR-E), 2012–present (AMSR2) | JAXA (2023) | Polarization index |
| 4 | AMSR-E, AMSR2 | 10 km | Daily | 2002–2011 (AMSR-E), 2012–present (AMSR2) | JAXA (2023) | Fraction of water surface (estimated at 18.7 and 36.5 GHz bands) |
| 5 | ERA5-Land | 0.1° | Hourly (aggregated into daily mean) | 1981–present | ECMWF (2023) | Dewpoint temperature at 2 m height (K) |
| 6 | ERA5-Land | 0.1° | Hourly (aggregated into daily mean) | 1981–present | ECMWF (2023) | Surface and subsurface runoff (m) |
| 7 | ERA5-Land | 0.1° | Hourly (aggregated into daily mean) | 1981–present | ECMWF (2023) | Skin reservoir water content (meters of water equivalent) |
| 8 | ERA5-Land | 0.1° | Hourly (aggregated into daily mean) | 1981–present | ECMWF (2023) | Land surface temperature (K) |
| 9 | ERA5-Land | 0.1° | Hourly (aggregated into daily mean) | 1981–present | ECMWF (2023) | Snow depth (meters of water equivalent) |
| 10 | ERA5-Land | 0.1° | Hourly (aggregated into daily mean) | 1981–present | ECMWF (2023) | Soil temperature in 0–7 cm layer (K) |
| 11 | ERA5-Land | 0.1° | Hourly (aggregated into daily mean) | 1981–present | ECMWF (2023) | Soil temperature in 7–28 cm layer (K) |
| 12 | ERA5-Land | 0.1° | Hourly (aggregated into daily mean) | 1981–present | ECMWF (2023) | Soil temperature in 28–100 cm layer (K) |
| 13 | ERA5-Land | 0.1° | Hourly (aggregated into daily mean) | 1981–present | ECMWF (2023) | Soil temperature in 100–289 cm layer (K) |
| 14 | ERA5-Land | 0.1° | Hourly (aggregated into daily mean) | 1981–present | ECMWF (2023) | Air temperature at 2 m height (K) |
| 15 | ERA5-Land | 0.1° | Hourly (aggregated into daily mean) | 1981–present | ECMWF (2023) | Total evaporation (meters of water equivalent) |
| 16 | ERA5-Land | 0.1° | Hourly (aggregated into daily mean) | 1981–present | ECMWF (2023) | Total precipitation (m) |
| 17 | ERA5-Land | 0.1° | Hourly (aggregated into daily mean) | 1981–present | ECMWF (2023) | Volume of water in 0–7 cm soil layer (m ³ /m ³) |
| 18 | ERA5-Land | 0.1° | Hourly (aggregated into daily mean) | 1981–present | ECMWF (2023) | Volume of water in 7–28 cm soil layer (m ³ /m ³) |
| 19 | ERA5-Land | 0.1° | Hourly (aggregated into daily mean) | 1981–present | ECMWF (2023) | Volume of water in 28–100 cm soil layer (m ³ /m ³) |
| 20 | ERA5-Land | 0.1° | Hourly (aggregated into daily mean) | 1981–present | ECMWF (2023) | Volume of water in 100–289 cm soil layer (m ³ /m ³) |
| 21 | DOY | – | Daily | – | – | Cosine and sine of DOY |
| 22 | VODCA | 0.25° | Daily | 2002–2018 (C-band), 1997–2018 (X-band) | Moesinger et al. (2020) | Vegetation optical depth |

AMSR: Advanced Microwave Scanning Radiometer, DOY: day of year, ECMWF: European Centre for Medium-Range Weather Forecasts, ERA5: ECMWF Reanalysis v5, JAXA: Japan Aerospace Exploration Agency, MODIS: Moderate Resolution Imaging Spectroradiometer, NDWI: normalized difference water index, NDVI: normalized difference vegetation index, USGS: United States Geological Survey, VODCA: vegetation optical depth climate archive

After coregistration, daily climatology maps were prepared for the explanatory variables by averaging the maps for the same DOY over the period 2003–2017. These daily climatology maps were assumed to represent typical

values of the variables on a given DOY and were used to fill any missing pixels in the explanatory maps for ML prediction (see Sect. 2.2.3).

Table 3 Rate of missing pixels in each explanatory variable (parameter numbers 3–22) and number of available explanatory variables

| MODIS tiles | No. 3, 4 (36.5 GHz) | No. 4 (18.7 GHz) | No. 5–21 | No.22 | Mean number of available variables | Minimum number of available variables |
|-------------|------------------------|------------------|----------|-------|---------------------------------------|---|
| h20v02 | 0.64 | 0.61 | 1.00 | 0.74 | 21 | 18 |
| h21v02 | 0.64 | 0.61 | 1.00 | 0.80 | 21 | 18 |
| h22v02 | 0.64 | 0.61 | 1.00 | 0.86 | 22 | 18 |
| h23v02 | 0.64 | 0.61 | 1.00 | 0.89 | 22 | 18 |
| h24v02 | 0.64 | 0.61 | 1.00 | 0.90 | 22 | 18 |
| h25v02 | 0.64 | 0.61 | 1.00 | 0.89 | 22 | 18 |
| h20v03 | 0.64 | 0.61 | 1.00 | 0.66 | 21 | 18 |
| h21v03 | 0.64 | 0.61 | 1.00 | 0.76 | 21 | 18 |
| h22v03 | 0.64 | 0.61 | 1.00 | 0.78 | 21 | 18 |
| h23v03 | 0.64 | 0.61 | 1.00 | 0.80 | 22 | 18 |
| h24v03 | 0.64 | 0.61 | 1.00 | 0.87 | 22 | 18 |
| h25v03 | 0.64 | 0.61 | 1.00 | 0.90 | 22 | 18 |

Each value was temporally averaged over the study duration, and spatially averaged within each tile and masked for ocean area. The correspondence between the parameter numbers and the explanatory variables is detailed in Table 2 (number 1 and 2 were excluded because they are explained variables)

2.2.3 ML processing

In our data fusion scheme, pixel-based ML was used to train matching pairs of explanatory and explained variables, allowing the prediction of explained maps without any gaps (Mizuochi et al. 2018; 2021a). Before initiating ML processing, a preliminary study was conducted to select the ML model and to tune the hyperparameters using a limited number of sampling pixels (100 pixels) from several tiles. In total, 25 ML algorithms (different types of regression models) were compared based on ten-fold cross validation using PyCaret. RF was chosen as a robust regressor, with the following hyperparameters: maximum tree depth 5, minimum impurity decrease 0.005, minimum samples at leaf node 2, minimum samples for internal node splitting 7 and number of trees 60.

Strictly speaking, the rank of the models fluctuated depending on the tile, the presence of absence of snow, the NDWI or NDVI (detailed below), and the referred accuracy criteria. However, ensemble models of decision trees (RF and extremely randomized trees) usually showed top scores. For simplicity, we decided to use RF for all the tiles, since it is widely used in remote sensing (Zhao et al. 2022) and is also consistent with our previous research (Mizuochi et al. 2021a).

The spectral features of snow are characterized by a continuous decrease in reflectance from the visible to the infrared region (Petty 2006), a pattern that resembles that for water bodies in NDWI and NDVI maps. To consider the sensitivity of the snow pixels to the NDWI and NDVI, RF models with and without snow were created separately by checking if the SWE retrieved by AMSR-E/AMSR2 was positive or zero. Four original images taken in May, August, November, and February of a randomly

picked year during the study period were selected for validation of each season. Each variable was then normalized by subtracting the spatiotemporal mean from the original values and dividing by the spatiotemporal standard deviation for the RF training.

For each pixel, historical match-up pairs were searched during the period 2003–2017, where all the explained and explanatory variables existed altogether. In other words, one matching pair containing a full set of 23 explanatory and 2 explained variables was used as one training sample for each the pixel.

SWE and ERA5-Land included permanently missing pixels for water and ocean masks. To avoid wasting information for the other variables and to make a wall-to-wall product, ML was performed on these pixels by filling zero for SWE and the spatially averaged values of climatology maps for ERA5-Land.

After RF training, RF prediction was performed for each pixel: explanatory maps were again input to the trained RF model to predict the explained maps (i.e., MODIS NDWI and NDVI) without gaps. To this end, it was ensured that the explanatory maps at this stage had no missing values. Temporary gaps (much fewer than in the explained maps but still non-negligible) were filled by daily climatology maps. It should be noted that this treatment was applied only during the prediction process, not during the training process, to avoid overfitting to the average map. Ancillary flag maps indicating the absence/presence of this treatment (0: original explanatory variables, 1: explanatory variables filled by climatology data) and maps showing the presence of snow were also produced.

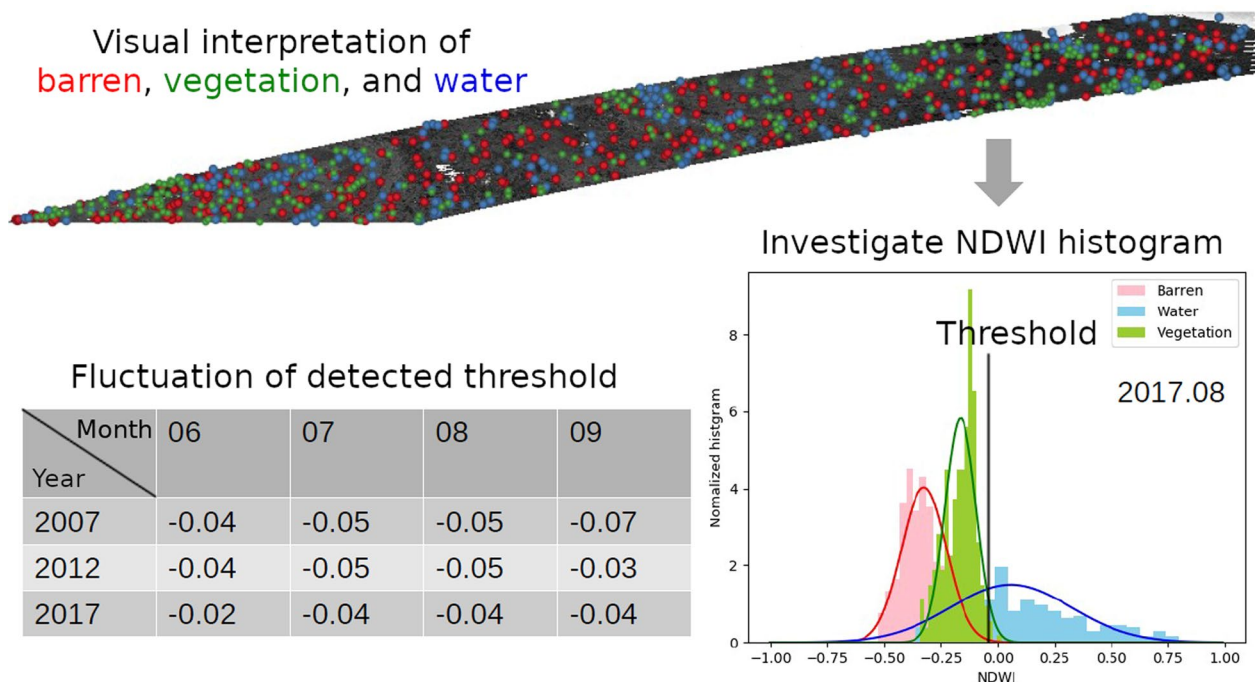


Fig. 2 Scheme of NDWI thresholding. (Top) Visual interpretation of three land cover categories (red: barren ground, green: vegetation, and blue: water); (bottom left) Summary table of extracted thresholds for different seasons; (bottom right) Example of maximum likelihood approach using data obtained in August 2017

Table 3 summarizes the tile-averaged rate of missing pixels filled by climatology data in each explanatory map, and the number of available explanatory variables. AMSR-related parameters (no. 3, 4) were approximately 60% available, and ERA-related parameters (no. 5 to 21) were 100% available, both of which were insensitive to the tile location. VOD (no. 22) was sensitive to the tile location, fluctuating in the range 66–90%. In all cases, 21–22 of a total of 23 parameters were available in the spatiotemporal average for all tiles. 18 parameters were obtained even in the worst case, which is understandable because ERA5-Land and DOY are at least available at any time. Therefore, if some variables were missed and filled by climatology data, the remaining variables are expected to offer realistic input data to some extent.

2.2.4 Postprocessing: creating water maps by NDWI thresholding

Inundation pixels were extracted as physically intuitive information by thresholding the NDWI created by ML. Through visual interpretation of NDWI maps and high-resolution satellite imagery from Google Earth, ground references were established for three typical land cover categories (barren land, vegetation and water). Three hundred samples for each category were then used to produce category-dependent NDWI histograms (Fig. 2).

A Gaussian distribution was fitted to each histogram, and the intersection between the Gaussian curve representing water and those representing the other categories was extracted as the tentative NDWI threshold for the data (i.e., the maximum likelihood approach).

The tentative thresholds between vegetation and water, derived from 12 (snow-free) monthly averaged maps for a particular tile (h24v02), showed a limited fluctuation range (Table 4). Consequently, the average value (NDWI = -0.043) with a standard error of 0.003 (N = 12) was assumed as the fixed threshold for all seasons. Utilizing this threshold, binary maps were generated to show 0 (non-water) and 1 (water).

To assess the uncertainty in the water map, a sensitivity analysis of the threshold setting was conducted. In the average map derived from the 12 monthly NDWI scenes, the threshold was slightly adjusted from the average value (NDWI = -0.043) within one standard deviation (0.012) using Gaussian noise. This process was repeated 12 times to calculate variations in the water fraction across the entire tile (h24v02). The mean water fraction and the standard error (N = 12) were 0.032 ± 0.001, corresponding to 3% uncertainty in water fraction estimation.

This study also considered fluctuating wetlands, i.e., areas where the extent of inundation changes seasonally, rather than permanent open water bodies such as lakes, ponds, rivers and reservoirs, which often attract

Table 4 Summary of estimated NDWI thresholds in different seasons

| Year/Month | Threshold between vegetation and water | Threshold between barren land and water | Sample size of barren land (pixels) | Sample size of vegetation (pixels) | Sample size of water (pixels) |
|------------|--|---|-------------------------------------|------------------------------------|-------------------------------|
| 2007/06 | -0.04 | -0.07 | 286 | 295 | 291 |
| 2007/07 | -0.05 | -0.15 | 299 | 300 | 297 |
| 2007/08 | -0.05 | -0.16 | 299 | 300 | 300 |
| 2007/09 | -0.07 | -0.12 | 144 | 194 | 185 |
| 2012/06 | -0.04 | -0.12 | 159 | 189 | 178 |
| 2012/07 | -0.05 | -0.14 | 284 | 288 | 274 |
| 2012/08 | -0.05 | -0.13 | 297 | 298 | 297 |
| 2012/09 | -0.03 | -0.10 | 116 | 113 | 121 |
| 2017/06 | -0.02 | -0.05 | 229 | 272 | 256 |
| 2017/07 | -0.04 | -0.15 | 299 | 300 | 300 |
| 2017/08 | -0.04 | -0.16 | 300 | 300 | 300 |
| 2017/09 | -0.04 | -0.08 | 130 | 124 | 139 |

NDWI: normalized difference water index

special interest in estimating methane emissions, a key process affecting climate change (Zhang et al. 2021). Thus, the mode value for the summer season was subtracted from the original water maps to create what we refer to as “floating water maps”, which likely represent areas that are intermittently inundated, and which exclude permanent open water.

2.2.5 Postprocessing: creation of analysis-ready data (ARD)

In addition to the raw binary maps based on MODIS tiles, which cover the wall-to-wall NDVI, NDWI, water, and quality flags at the original spatial resolution of 500 m, a merged product was also created using a commonly used latitude/longitude projection. The product was resampled using the nearest neighbor method and stored in GeoTiff format. To reduce the cost of data handling on a continental scale, the original resolution was aggregated into a 0.1° pixel spacing. Within these pixels, the average NDVI and NDWI values were stored, along with the fractional coverage of both permanent and intermittent wet areas.

2.3 Validation

The ML result was validated for each tile over four seasons (Table 5), by comparing the generated MODIS NDVI and NDWI images with the original images that were not used in the ML training process (i.e., by employing a hold-out approach). Spatial patterns and statistical criteria were evaluated, with the latter including the mean error, root mean squared error (RMSE), correlation coefficient[®], slope and offset of the regression line between the original and created images. These values

Table 5 Acquisition date of original MODIS images used for validation in each tile

| MODIS tile | Spring | Summer | Fall | Winter |
|------------|------------|------------|------------|------------|
| h20v02 | 2009/05/16 | 2006/08/27 | 2004/11/01 | 2009/02/19 |
| h21v02 | 2005/05/01 | 2007/08/19 | 2008/11/02 | 2005/02/22 |
| h22v02 | 2015/05/05 | 2016/08/11 | 2017/11/01 | 2004/02/13 |
| h23v02 | 2004/05/13 | 2012/08/09 | 2006/11/02 | 2013/02/11 |
| h24v02 | 2011/05/09 | 2013/08/31 | 2008/11/03 | 2016/02/15 |
| h25v02 | 2011/05/29 | 2014/08/28 | 2003/11/15 | 2015/02/27 |
| h20v03 | 2003/05/06 | 2008/08/09 | 2014/11/04 | 2004/02/03 |
| h21v03 | 2011/05/06 | 2015/08/02 | 2016/11/14 | 2014/02/09 |
| h22v03 | 2017/05/03 | 2017/08/16 | 2004/11/16 | 2005/02/03 |
| h23v03 | 2007/05/03 | 2003/08/01 | 2015/11/21 | 2010/02/26 |
| h24v03 | 2017/05/07 | 2014/08/11 | 2006/11/15 | 2015/02/22 |
| h25v03 | 2012/05/09 | 2010/08/18 | 2004/11/01 | 2006/02/11 |

MODIS: Moderate Resolution Imaging Spectroradiometer

were summarized for each MODIS tile and image acquisition season.

Spatiotemporal patterns were also checked for the 0.1° ARD product. In addition to visualizing the NDVI and NDWI, spatially averaged time series and temporally averaged spatial patterns of the original and floating water maps were compared with a previous wetland map (WAD2M; Zhang et al. 2021). WAD2M primarily draws upon water fraction maps derived from multiple microwave satellite data sources (SWAMPS; Jensen and McDonald 2019), supplemented with ancillary satellite data and static maps for snow and permanent water masking, plus bias correction. The recent product, which emerged from the integration of multiple data sources

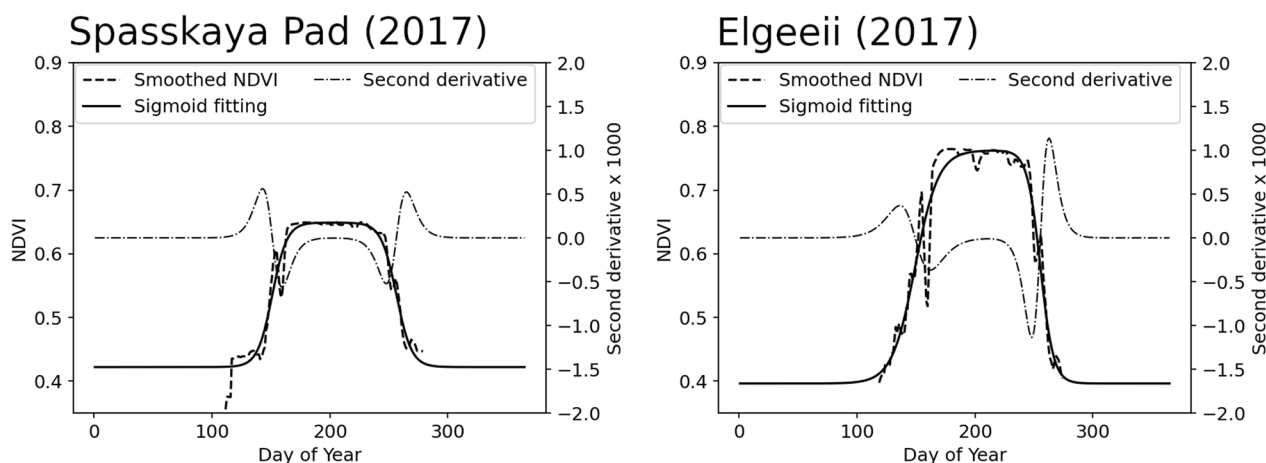


Fig. 3 Examples of double sigmoid fitting for seasonal NDVI. The NDVI with a 5-day moving average, sigmoid fitting and second derivative plots for sites at Spasskaya Pad (left) and Elgeei (right) during 2017 are shown

through different schemes and at varying spatiotemporal resolutions (0.25° , monthly), seems useful for assessing the advantages and potential disadvantages of our product.

The terrestrial water storage (TWS) anomaly is also useful for time-series comparisons with surface water (Suzuki et al. 2018). Gravity Recovery and Climate Experiment (GRACE) monthly water mass datasets were prepared (Swenson 2012; Landerer and Swenson 2012), with contributions from the Center for Space Research (CSR), GeoForschungsZentrum Potsdam (GFZ), and NASA Jet Propulsion Laboratory (JPL), culminating in the calculation of the ensemble mean of the three datasets (Sakumura et al. 2014). In addition, as in Suzuki et al. (2018), reanalysis data from Global Land Data Assimilation System (GLDAS) version 2 (Li et al. 2019) were also compared for the other TWS data source. The TWS includes snowpack, river runoff and groundwater and is not directly comparable to the NDWI and NDVI. However, it is helpful for interpreting time-series NDWI (and derived surface water area) data in relation to the water budget.

2.4 Application: trend analysis and phenological feature extraction

The first application of the product developed in this study involved trend analysis at a continental scale. Time-series of snow-masked NDVI and NDWI maps were used to calculate the 15-year trend (i.e., Theil Sen's regression slope) for each 0.1° pixel. The statistical significance of the trends was calculated using the Mann–Kendall test. The results were compared with previous reports for 15-year trends in vegetation and water in polar regions.

The second application involved the extraction of phenological features at two study sites covered by larch

forest (Spasskaya Pad: $62^\circ 15' 17''$ N, $129^\circ 37' 10''$ E; Elgeei: $60^\circ 01' 01''$ N, $133^\circ 49' 53''$ E), utilizing the high observation frequency of our product (Fig. 3); detailed descriptions of each study site can be found in Nagano et al. (2022). As a useful approach for phenological monitoring in terrestrial ecosystems, a moving average (window size=5 days) of snow-masked NDVI time series data was fitted using a double sigmoid function (e.g., Ide and Oguma 2013; Myers et al. 2019; Yan et al. 2019) as follows:

$$\text{NDVI}(t) = \text{NDVI}_b + \frac{1}{2} \text{NDVI}_a [\tanh(p(t - D_i)) - \tanh(q(t - D_d))], \quad (4)$$

where NDVI_b is the baseline value, NDVI_a is the amplitude, D_i is the day of year when the NDVI increases most rapidly, D_d is the day of year when the NDVI decreases most rapidly, and p and q determine the rate of increase and decrease of the NDVI, respectively. The initial parameter values were set to $\text{NDVI}_b = 0.4$, $\text{NDVI}_a = 0.2$, $D_i = 150$, $D_d = 250$, and $p = q = 0.001$, based on a preliminary investigation of interannual phenology patterns over the sites. These parameters were then precisely fitted to the actual NDVI time-series data during the growing season (assumed to be from $\text{DOY} = 100$ to $\text{DOY} = 280$), employing the *curve_fit* function of the Python SciPy library.

Then, local maximum and minimum values of the second derivative in the fitted function were identified, allowing for the determination of four commonly used phenological features (i.e., D_1 , D_2 , D_3 , and D_4), which correspond to the start of spring (SOS), end of spring (EOS), start of fall (SOF), and end of fall (EOF), respectively. The date of the peak NDVI (D_5) was also extracted, and the 15-year trends in these

Table 6 Definition of five phenological parameters

| Parameter | Meaning | Definition |
|-----------|------------------------|---|
| D_1 | Start of spring (SOS) | DOY of maximum second derivative of NDVI fitting curve during spring season |
| D_2 | End of spring (EOS) | DOY of minimum second derivative of NDVI fitting curve during spring season |
| D_3 | Start of fall (SOF) | DOY of maximum second derivative of NDVI fitting curve during fall season |
| D_4 | End of fall (EOF) | DOY of minimum second derivative of NDVI fitting curve during fall season |
| D_5 | Peak of growing season | DOY of maximum NDVI in fitting curve |

Spring and fall seasons were roughly assumed from DOY = 100 to DOY = 180, and from DOY = 200 to DOY = 280, respectively, during which time the local extremum in the second derivative was searched for

DOY: day of year, NDVI: normalized difference vegetation index

five parameters were investigated. A summary of the individual definitions of these phenological features is given in Table 6.

3 Results and discussion

3.1 Created data and validation

Data fusion using RF produced wall-to-wall MODIS-like NDWI and NDVI maps for the period 2003–2017 with enhanced the spatiotemporal extent compared to the existing data fusion product (Mizuochi et al. 2021a).

Examples of comparisons between the original and newly created maps in a tile are shown in Figs. 4 and 5. The examples show that the cloud gaps present in the original images were completely filled in the images created by RF regression, closely resembling the original spatial patterns and values of the NDWI and NDVI. The scatterplots show that images in the fall season deviated from the 1:1 (i.e., unity) lines, and that spatially discontinuous patterns were observed in the southwest area.

The discrepancy in the scatterplot analysis was quantified by Theil-Sen's regression slope for each tile and season (Table 7). While the regression slope was close to unity for both the NDWI and NDVI in summer, large deviations representing uncertainties in the values were observed in the seasons affected by snow or soil freezing. Since the RF models were trained separately for snow and no-snow pixels, the mixture of these pixels within a scene resulted in discontinuous spatial patterns and poor regression among multiple clusters in the scatterplots (Fig. 6, h22v03). The other reason for the poor regression may have been that the limited variable range of the NDWI and NDVI in the snow-covered winter season creates a featureless spectral pattern over the entire scene (Fig. 6, h23v02). Different tiles likely had different snow or soil freezing event distributions, resulting in varying accuracies among tiles (Table 7).

Table 7 also shows the small mean error for all tiles (-0.01 ± 0.04 for NDWI, and 0.02 ± 0.02 for NDVI), indicating minimal biases in the reproduced images. The

RMSE for all tiles was approximately 0.1 for both the NDWI and NDVI.

Seasonal changes were clearly observed in the snow-masked NDWI and NDVI time-series data, averaged over the entire study region (Fig. 7A). The NDVI oscillated between being high in the growing season and low in the winter season, while the NDWI was high in the winter season and low in the growing season. A high local NDWI peak was observed in the spring season, possibly due to spring soil thawing and/or snow melting which rapidly supplied water to rivers and wetlands. Given the high ratio of null pixels in the winter season (~ 1.0), the low NDVI and the high NDWI observed at this time were attributed to the limited number of pixels affected by snow, even after the application of snow masking, which produced spectral patterns that differed from the original land cover. In addition, the low solar incidence angle in the winter season could reduce the signal-to-noise ratio for the measured radiance as well as the reliability of atmospheric correction for the optical data, potentially resulting in poor quality of the original MODIS products. As a result, studies on vegetation and surface water using our ARD product should focus on spring through fall, with special consideration given to snow masking.

Gap-filling by the climatology maps (i.e., an average value for the same date over 15 years) can potentially lead to bias in the prediction. This is particularly the case when predicting recent climatic extremes away from the 15-year average. In fact, slight differences in patterns around the beginning of 2012 were likely caused by a gap in the microwave data (AMSR-E terminated on 2011/10/03 and AMSR2 started on 2012/07/02), which was filled by the daily climatology maps for ML prediction.

Time-series data for the original and floating water fraction (Fig. 7B) showed consistent seasonal variability, with high values recorded in the summer season (~ 0.07) and the spring local peak, and low values recorded in the winter season. The baseline for the original water

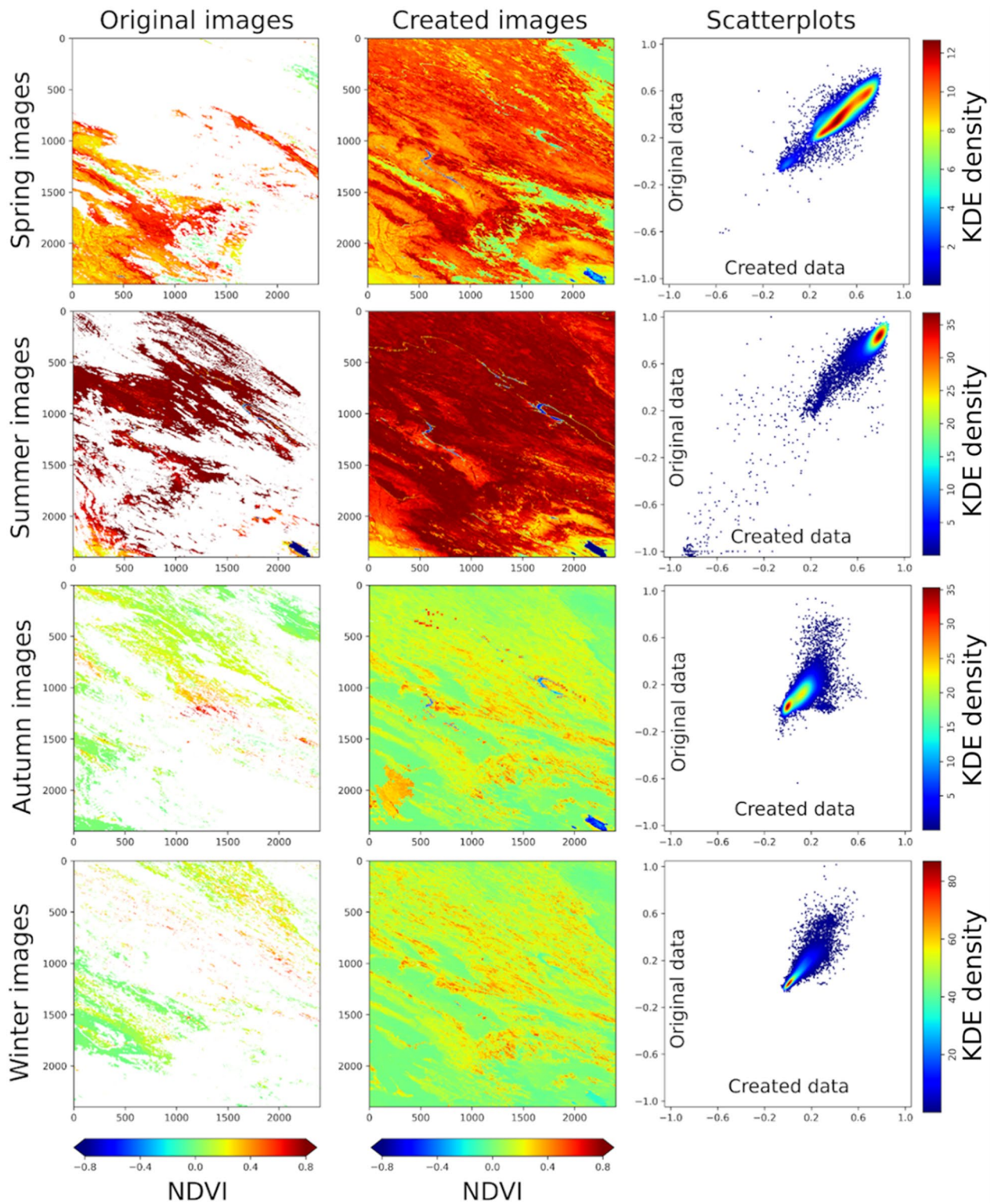


Fig. 4 Comparison between original and created MODIS NDVI maps for tile h23v03. Spring image: 2007/05/03, summer image: 2003/08/01, fall image: 2015/11/21, winter image: 2010/02/26. Color scale of the scatterplot was configured by kernel density estimation (KDE)

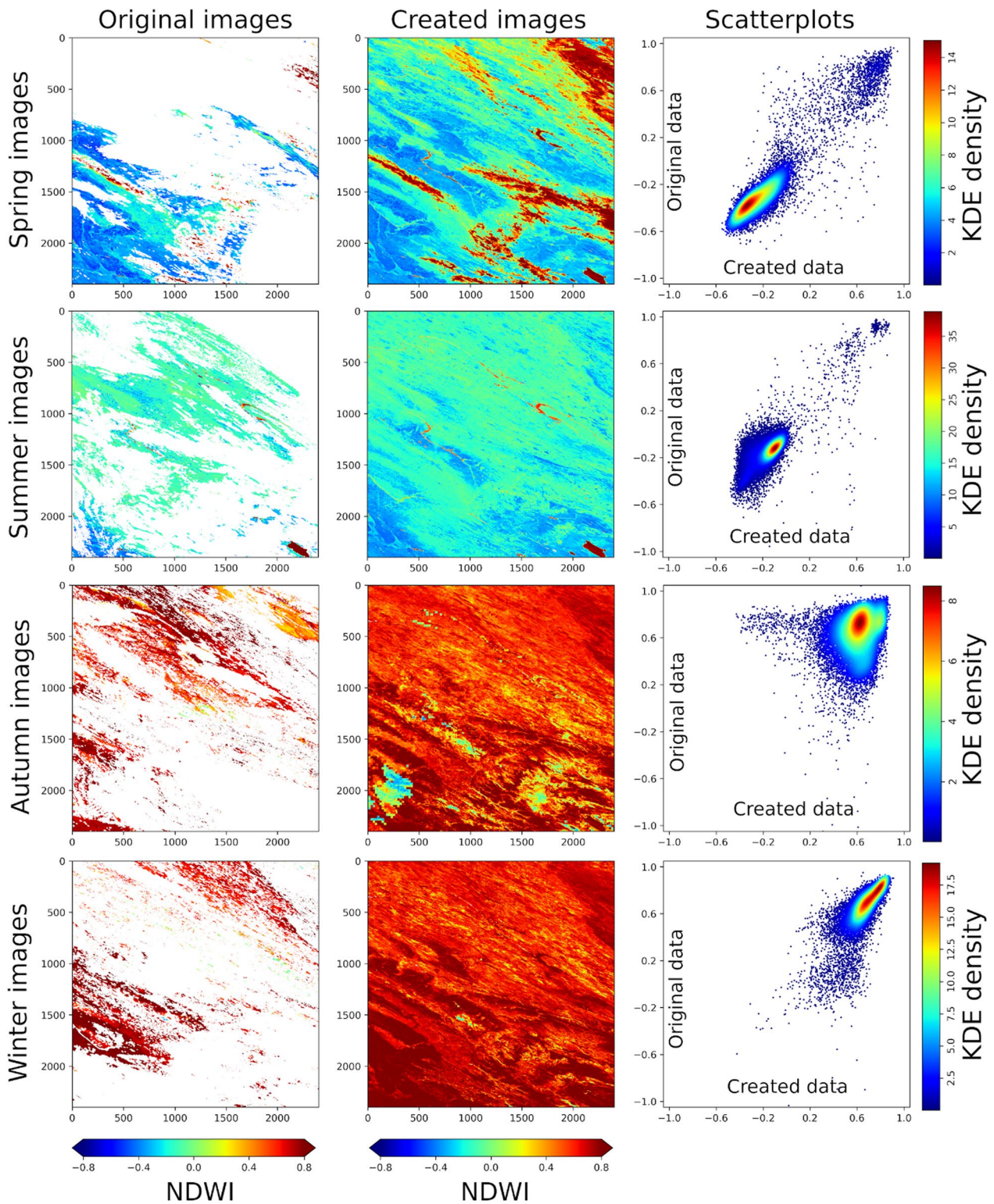


Fig. 5 Same as Fig. 4, but for MODIS NDWI

Table 7 Accuracy of RF prediction averaged for each tile, for each season with standard deviation

| | Slope | Offset | Mean error | RMSE | Correlation coefficient |
|------------------|-----------|------------|------------|-----------|-------------------------|
| h20v02 | 0.72±0.43 | 0.24±0.28 | -0.01±0.03 | 0.05±0.02 | 0.58±0.30 |
| | 1.32±0.33 | -0.02±0.03 | 0.01±0.02 | 0.09±0.17 | 0.80±0.12 |
| h20v03 | 0.86±0.12 | 0.01±0.20 | 0.00±0.11 | 0.12±0.07 | 0.48±0.22 |
| | 0.87±0.10 | 0.02±0.05 | 0.03±0.05 | 0.03±0.02 | 0.60±0.16 |
| h21v02 | 1.19±0.28 | -0.07±0.24 | -0.01±0.02 | 0.41±0.74 | 0.34±0.35 |
| | 1.10±0.15 | -0.02±0.05 | 0.01±0.05 | 0.12±0.22 | 0.54±0.28 |
| h21v03 | 0.81±0.35 | 0.12±0.27 | 0.01±0.09 | 0.09±0.09 | 0.40±0.30 |
| | 0.98±0.16 | -0.01±0.04 | -0.01±0.03 | 0.02±0.01 | 0.67±0.23 |
| h22v02 | 0.79±0.27 | 0.18±0.25 | -0.05±0.07 | 0.03±0.02 | 0.43±0.18 |
| | 0.80±0.30 | 0.00±0.09 | 0.03±0.06 | 0.01±0.00 | 0.72±0.12 |
| h22v03 | 0.67±0.28 | 0.22±0.34 | -0.09±0.19 | 0.15±0.15 | 0.54±0.29 |
| | 0.74±0.41 | 0.01±0.03 | 0.04±0.09 | 0.03±0.02 | 0.58±0.22 |
| h23v02 | 1.03±0.21 | -0.01±0.22 | -0.01±0.08 | 0.03±0.03 | 0.56±0.24 |
| | 0.61±0.29 | -0.01±0.05 | 0.04±0.06 | 0.02±0.02 | 0.50±0.19 |
| h23v03 | 0.75±0.34 | 0.08±0.28 | 0.01±0.04 | 0.24±0.11 | 0.34±0.29 |
| | 0.92±0.12 | 0.05±0.07 | -0.01±0.05 | 0.10±0.16 | 0.61±0.23 |
| h24v02 | 1.00±0.27 | 0.05±0.13 | 0.02±0.05 | 0.02±0.00 | 0.54±0.38 |
| | 0.53±0.35 | -0.00±0.05 | 0.03±0.01 | 0.01±0.02 | 0.51±0.35 |
| h24v03 | 1.03±0.09 | -0.05±0.11 | 0.03±0.07 | 0.15±0.16 | 0.67±0.19 |
| | 1.05±0.25 | 0.01±0.05 | 0.00±0.08 | 0.17±0.13 | 0.49±0.09 |
| h25v02 | 0.82±0.26 | 0.22±0.19 | -0.05±0.07 | 0.05±0.06 | 0.59±0.26 |
| | 0.75±0.37 | -0.05±0.04 | 0.07±0.07 | 0.16±0.19 | 0.65±0.11 |
| h25v03 | 1.10±0.07 | -0.06±0.11 | 0.05±0.14 | 0.25±0.13 | 0.66±0.04 |
| | 1.01±0.19 | -0.01±0.12 | 0.03±0.14 | 0.34±0.37 | 0.61±0.17 |
| Spring | 0.87±0.23 | 0.02±0.25 | 0.02±0.11 | 0.08±0.07 | 0.75±0.11 |
| | 0.83±0.22 | -0.01±0.05 | 0.04±0.06 | 0.05±0.07 | 0.74±0.14 |
| Summer | 1.05±0.14 | 0.03±0.05 | -0.03±0.05 | 0.20±0.43 | 0.62±0.26 |
| | 1.03±0.14 | 0.02±0.10 | 0.00±0.09 | 0.24±0.25 | 0.64±0.20 |
| Fall | 0.83±0.43 | 0.11±0.35 | -0.01±0.13 | 0.16±0.15 | 0.36±0.19 |
| | 0.82±0.39 | -0.01±0.05 | 0.03±0.07 | 0.07±0.13 | 0.45±0.13 |
| Winter | 0.84±0.24 | 0.15±0.17 | -0.01±0.03 | 0.09±0.09 | 0.31±0.16 |
| | 0.89±0.42 | -0.02±0.02 | 0.01±0.02 | 0.01±0.02 | 0.59±0.21 |
| All tile average | 0.90±0.17 | 0.08±0.12 | -0.01±0.04 | 0.13±0.12 | 0.51±0.11 |
| | 0.89±0.22 | -0.00±0.02 | 0.02±0.02 | 0.09±0.10 | 0.61±0.09 |

Regression was performed using the created values as the X-axis and the original values as the Y-axis, with 1% samples randomly extracted from entire pixels. For each cell, the upper row is the validation of the NDWI and the lower row is that of the NDVI

RF: random forest, NDWI: normalized difference water index, NDVI: normalized difference vegetation index, RMSE: root mean squared error

fraction was adjusted to zero in the floating water fraction. Comparison with a previous wetland map (i.e., WAD2M; Zhang et al. 2021) showed similar seasonal changes during the spring to fall seasons, although large discrepancies were observed in the winter season. The water fraction obtained in the present study for the winter season was lower than that for WAD2M, in both the original and floating cases. This difference was likely due to differences in the definitions used for snow masking. Specifically, our product strictly masked pixels where the

microwave SWE was positive. This screened almost all of the pixels in the winter season and resulted in the water fraction being nearly zero. Conversely, the water fraction data from SWAMPS version 3.2 (Jensen and McDonald 2019), mainly used in WAD2M, applied a more relaxed approach to snow masking, which resulted in a substantial water fraction even in the winter season.

In the spring to fall seasons, the original WAD2M values were close to the lower limit of our original water fraction. In contrast, those for the offset-corrected

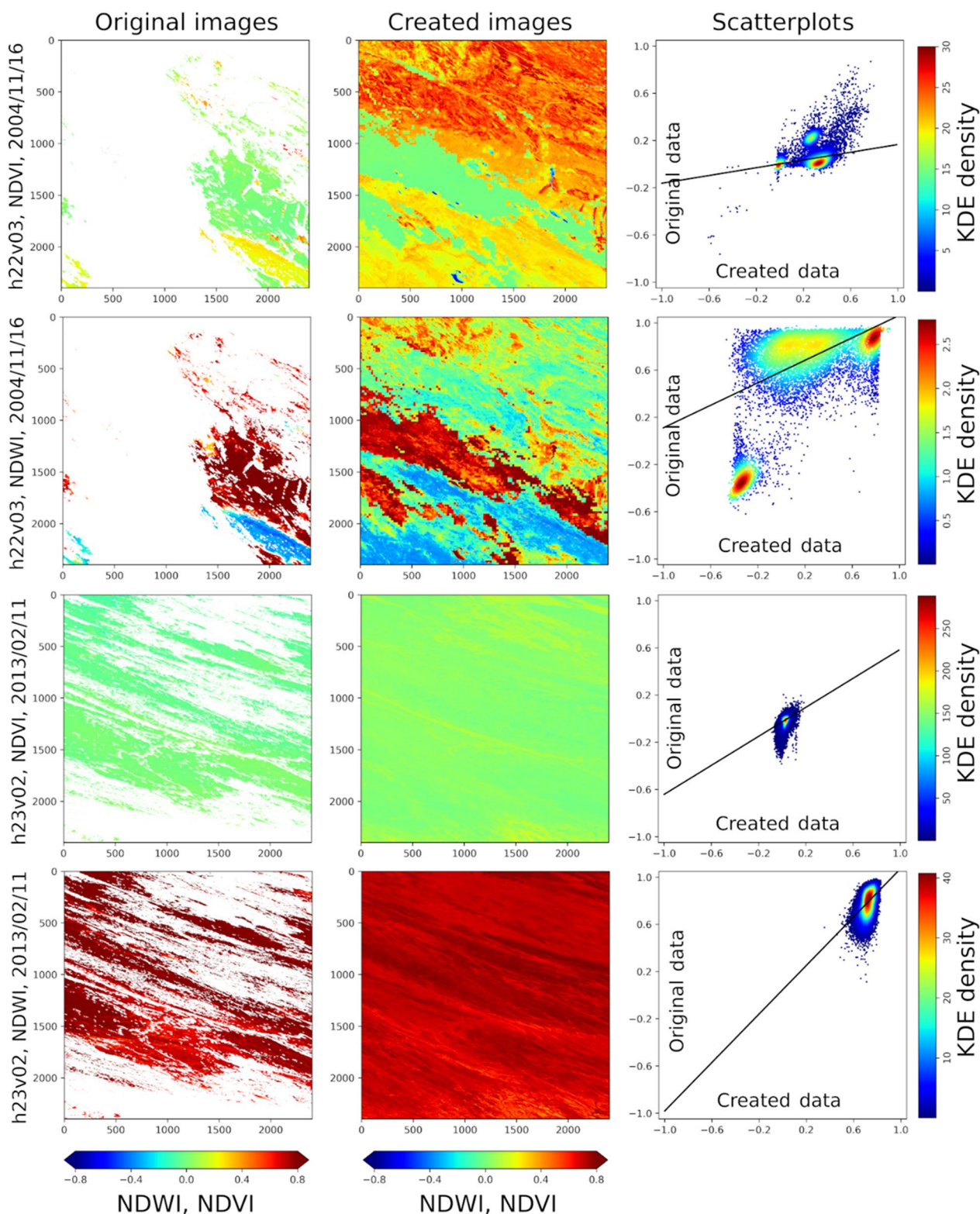


Fig. 6 Snow-affected images. Images showing regressions adversely affected by a mixture of snow and no-snow areas (h22v03), and areas with fewer spectral features due to snow cover (h23v02)

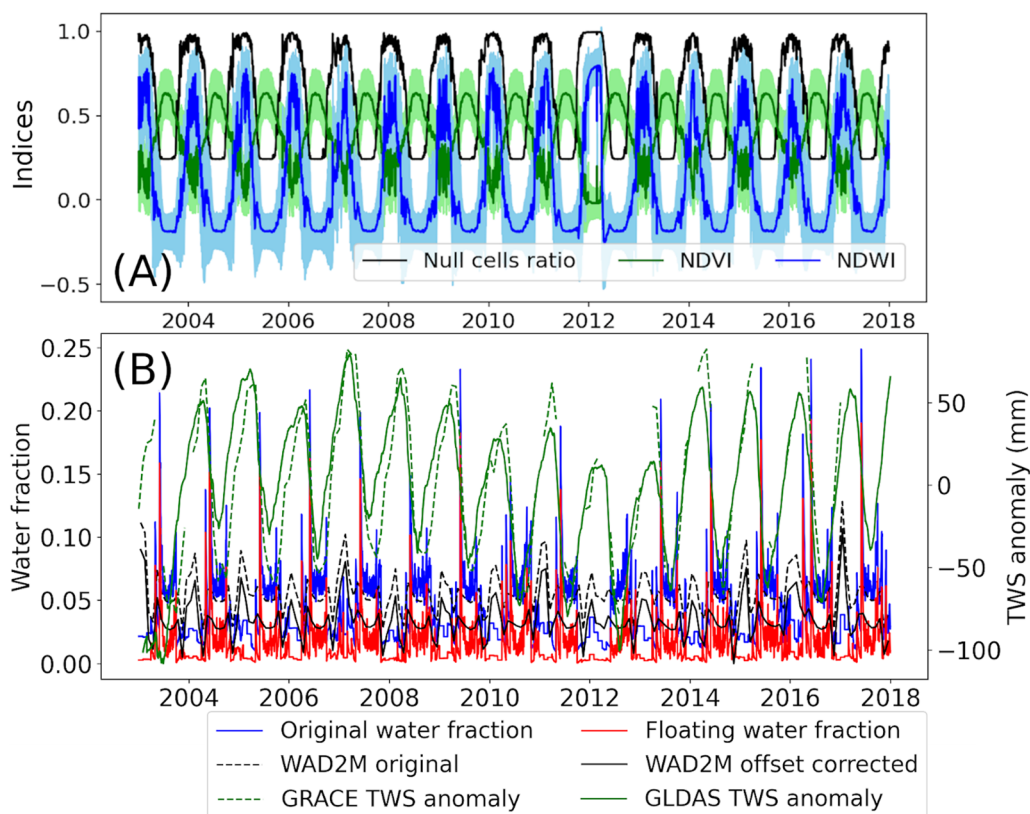


Fig. 7 Time series of 0.1° ARD product spatially averaged over the entire Siberian region. **A** Snow-masked NDVI and NDWI time-series with ratio of null cells averaged over entire study region. **B** Comparison between original (blue line) and floating (red line) water fraction extracted from our product, WAD2M wetland maps (Zhang et al. 2021), and time-series anomaly of terrestrial water storage (TWS) volumes (mm) from GRACE monthly water mass dataset (Swenson 2012; Landerer and Swenson 2012) and Global Land Data Assimilation System (GLDAS) version 2 (Li et al. 2019). For comparison with our floating water fraction, the minimum value of the WAD2M time-series data was subtracted from the original WAD2M to adjust the baseline to zero

WAD2M (adjusted to make the minimum values zero) were close to the upper limit of our floating water fraction. The WAD2M did not sense the strong peaks in the water fraction associated with soil thawing or snow melting in spring, probably due to its lower observation frequency (monthly) and coarser spatial resolution (0.25°) compared to our product (daily, 0.1° in the ARD product and 500 m in the original product). Furthermore, differences in the variable range of the water fraction between the ARD product and WAD2M also seemed to be linked to differences in their data sources. Unlike WAD2M, our optical sensor-based output observes the tree canopy in densely vegetated areas, but cannot explicitly delineate inundation under the vegetation. However, our product does implicitly utilize sources such as microwave data (i.e., AMSR series) and other data sources (i.e., ERA5-Land, VOD, DOY) that may be sensitive to inundation under the vegetation to some extent.

The interpretations above are also supported by TWS from GRACE and GLDAS, both of which account for snowpacks and show large values in winter. Snowmelt from snowpacks increases river runoff in the spring (Suzuki et al. 2018), which is consistent with the peak observed in our water fraction data. The lowest TWS during the summer–fall period also likely reflected our water fraction, as seen in the local minimum in the summer season.

Figure 8 shows temporally averaged NDWI and NDVI maps, which are useful for visualizing the overall distribution of surface water and vegetation (Fig. 8A, B). The original water fraction map largely contained permanent water bodies, such as rivers, lakes, and coastal areas (Fig. 8C), which were mostly excluded in the floating water fraction map (Fig. 8D). The difference between the original and floating water maps exposes the intermittent wetlands. Comparisons with WAD2M showed a

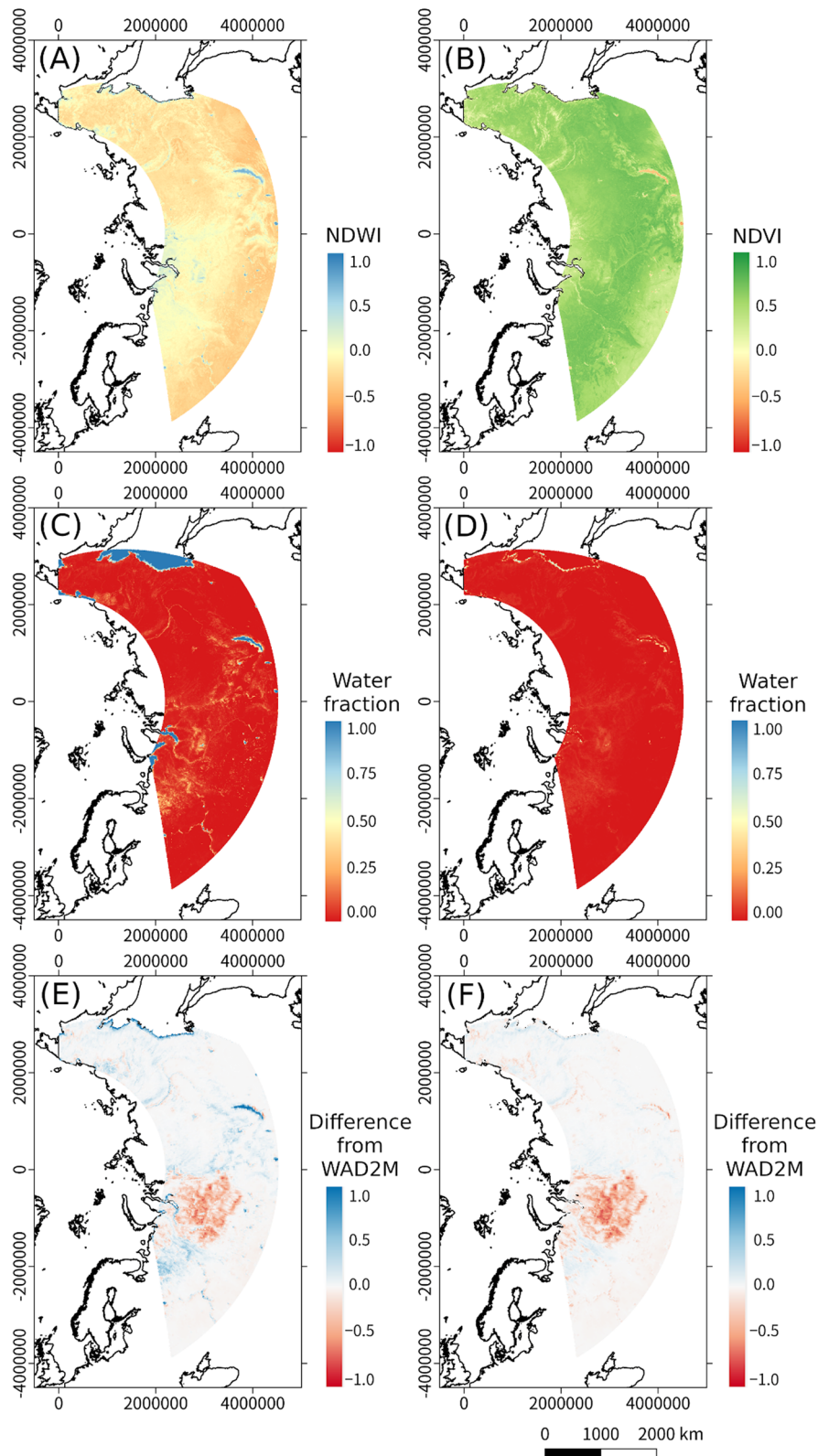


Fig. 8 Temporally averaged maps during period 2003–2017. **A** NDWI, **B** NDVI, **C** original water fraction, **D** floating water fraction, and difference between wetland map (WAD2M) provided by Zhang et al. (2021) and **E** original water fraction and **F** floating water fraction in this study at monthly intervals. The maps visualized on the Arctic Polar Stereographic projection (EPSG: 3995) are convenient for comparison with previous results for long-term terrestrial trends in the polar region

substantial underestimation of the water fraction on the west Siberian plain (Fig. 9E, F). High-resolution satellite images (Google Earth) confirmed the existence of abundant small-scale wetlands, which may not be detected by our moderate-resolution product (i.e., 500 m). Different effects of topography and vegetation cover on optical and microwave data may also cause discrepancies between WAD2M and our product. However, apart from this area, the exclusion of permanent water bodies from the original water map improved the consistency of the water fraction data between our product and WAD2M (Fig. 8F).

3.2 Application: trend analysis and phenological feature extraction

Our wall-to-wall products enable us to perform time-series trend analyses for vegetation and surface water (Fig. 9). Specifically, the NDVI slope map showed an overall decreasing trend especially in the southwest area, with sporadic hotspots in western to central Siberian wetlands and water bodies such as Lake Baikal. Several studies have reported an overall greening trend (i.e., increasing NDVI) in the Arctic region in recent decades, which they have attributed to increased temperatures and a longer growing season (May et al. 2020). However, there have been also reports of browning (i.e., decreasing NDVI) (Myers-Smith et al. 2020), with the distribution varying depending on the data sources and methodologies used. The browning patterns (i.e., negative NDVI values) shown in Fig. 9 are closely aligned with the spatial distribution of the annual maximum NDVI trend (Fig. 1 in Myers-Smith et al. 2020) based on the GIMMS3g AVHRR dataset (2000–2015) (Tucker et al. 2005; Pinzon and Tucker 2014), which supports the validity of our NDVI product. Extensive greening is expected over the northernmost Arctic Tundra region, but this could not be investigated as this region is not covered by our product. Consequently, expanding the map coverage area will be undertaken in future research. Expanding the study duration with similar data sources with a longer period is also important future work to monitor recent climate extremes after 2017 and to track the historical record before 2003.

The NDWI slope map showed an overall increasing trend across the study area. Given that the NDWI and NDVI are based on the inverse differences between visible and infrared reflectance, their overall contrasting spatial trends are expected. Exceptions were found in small hotspots around Lake Baikal, the lake around the Zeya Nature Reserve, and coastal areas facing the Sea of Okhotsk, where decreasing trends were observed even in NDWI slope maps. These hotspots were also observed in

the water fraction map for WAD2M, although WAD2M showed an overall decreasing trend across the study area. Rather, our results were more consistent with previous findings of significant increasing trends in the water fraction from 2003 to 2010 (Watts et al. 2012).

The dense time series (i.e., daily) nature of our product is well suited to extracting phenological parameters in the spring and the fall seasons through double-sigmoid fitting (Tables 8 and 9, Fig. 10). In both Spasskaya Pad and Elgeei, D_1 showed a negative trend (i.e., SOS occurring earlier), whereas D_3 and D_4 showed positive trends (i.e., SOF and EOF occurring later), although only the D_3 in Spasskaya Pad was statistically significant. Therefore, the duration of the growing season at the two larch forest sites increased from 2003 to 2017.

Nagai et al. (2020) determined the start of the growing season (SGS) and the end of the growing season (EGS) for the Spasskaya Pad and Elgeei sites, based on in-situ observations and a degree-day model. Although the parameters that they used differed from those used in this study, the SGS was significantly correlated with our D_1 (SOS) ($r=0.55$ with $p<0.05$ for Spasskaya Pad; $r=0.93$ with $p<0.01$ for Elgeei). These findings show that our SOS extraction is potentially well suited for use as a proxy for SGS. Conversely, unlike SGS, EGS showed a weaker correlation with our D_4 (EOF) ($r=0.34$ with $p=0.21$ for Spasskaya Pad; $r=0.54$ with $p=0.17$ for Elgeei).

The D_1 time series data for Spasskaya Pad seem to have changed at around 2005–2007, with the SOS occurring earlier thereafter (Fig. 10). This change was likely associated with waterlogging events and changes in the composition of the understory/overstory vegetation from 2005 to 2008 at the site (Kotani et al. 2014; 2019; Ohta et al. 2014; Nagano et al. 2022). The DOY of the Peak NDVI (D_5) showed larger interannual variability than the other parameters, and the temporal changes in D_5 were similar between Spasskaya Pad and Elgeei. Similarly, D_1 and D_2 showed time-series similarity between Spasskaya Pad and Elgeei, suggesting that these parameters were sensitive to large-scale meteorological factors such as temperature. Nagano et al. (2022) also observed a strong correlation between temperature anomalies and precipitation between Spasskaya Pad and Elgeei, supporting the importance of temperature as a large-scale phenological driver.

Nagai et al. (2020) also stressed the complexity of the sensitivity of leaf senescence to air temperature, noting that various internal (hormones and timing of spring budburst) and external (precipitation, photoperiod, drought, and heat stress) factors need to be considered in order to better understand leaf senescence, as doing so will improve predictions of the EOF and/or EGS.

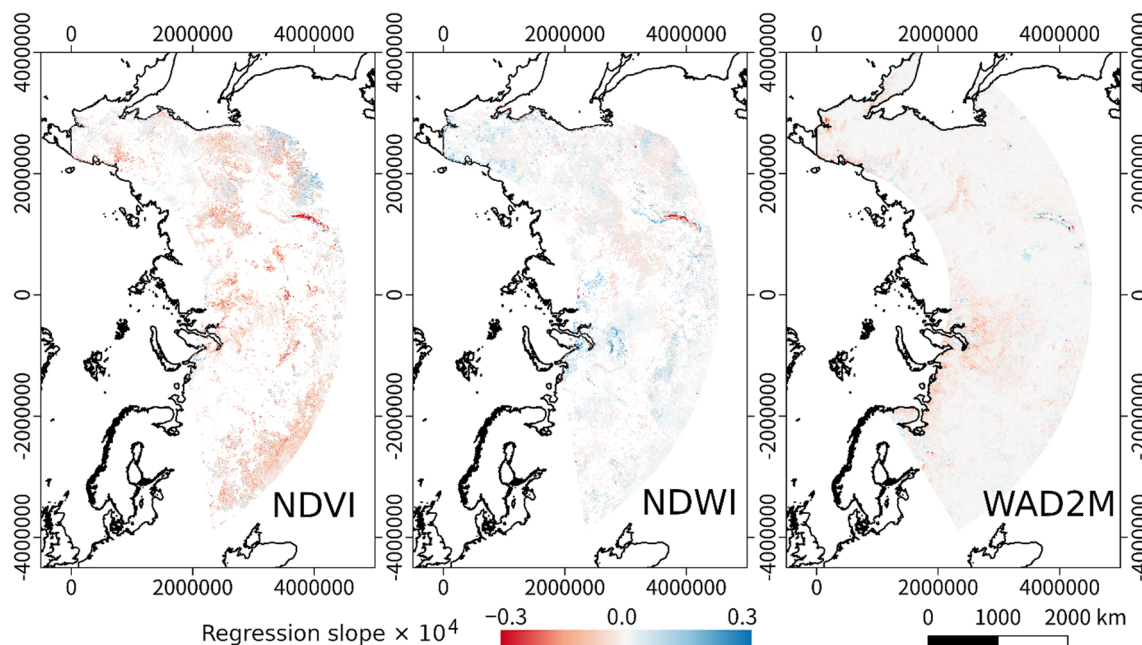


Fig. 9 Maps showing time-series regression. Regression slope maps for snow-masked NDVI, NDWI, and water fraction provided by Zhang et al. (WAD2M 2021) are shown. Only statistically significant ($p < 0.05$) pixels are shown in color for the NDVI and NDWI slope maps. The map projection is similar to that in Fig. 8

In this study, our original and ARD products are considered to be valuable for monitoring water and vegetation distributions, and for performing trend analysis and phenological research. Further potential applications may include contributions to land surface modeling (Guimberteau et al. 2018) as input and/or reference data. In addition, enhancing time-series land cover maps (Brown et al. 2022) to explicitly quantify temporal changes in vegetation and water distributions could also be undertaken in the future.

4 Conclusions

This research provided 15-years water and vegetation maps with daily frequency and a 500-m spatial resolution over the Siberia region on a continental scale, integrating optical and microwave satellite data, and meteorological data. A systematic treatment based on pixel-based RF techniques and further postprocessing produced maps of the NDWI, NDVI, and water fraction (original and floating), without any gaps for the period from 2003 to 2017.

In addition to the original sinusoidal maps with 500-m resolution, an analysis-ready product that merges all of the sinusoidal maps on a latitude/longitude projection was also rescaled to a resolution of 0.1° pixel spacing. The treated NDWI and NDVI images showed no overall systematic biases, with a RMSE of approximately 0.1 for all tiles. Based on the products, trend analysis showed the distribution of NDVI browning and increasing NDWI hotspots for the period from 2003 to 2017. Phenological features were extracted for each year at two larch forest sites (Spasskaya Pad and Elgeei) based on double-sigmoid fitting of NDVI time-series data. The findings revealed a recent lengthening tendency in the growing period at the sites and confirmed that waterlogging caused an earlier start to spring at the Spasskaya Pad site. Our product can potentially be applied to spatiotemporal monitoring of water and vegetation, including trend analysis, phenological research, model integration and contribution to land cover mapping.

Table 8 Phenological parameters estimated by double sigmoid fitting of NDVI over Spasskaya Pad

| YR | D_1 (SOS) | D_2 (EOS) | D_3 (SOF) | D_4 (EOF) | D_5 (peak) | SGS (Nagai et al. 2020) | EGS (Nagai et al. 2020) |
|--------------------------|-------------|-------------|-------------|-------------|--------------|-------------------------|-------------------------|
| 2003 | 151 | 167 | 244 | 264 | 201 | 153 | 281 |
| 2004 | 158 | 170 | 242 | 252 | 210 | 146 | 274 |
| 2005 | 151 | 157 | 236 | 267 | 172 | 129 | 277 |
| 2006 | 145 | 150 | 236 | 261 | 166 | 143 | 279 |
| 2007 | 134 | 155 | 243 | 271 | 194 | 137 | 278 |
| 2008 | 143 | 148 | 246 | 256 | 184 | 143 | 284 |
| 2009 | 146 | 151 | 243 | 264 | 170 | 137 | 285 |
| 2010 | 135 | 156 | 244 | 257 | 209 | 134 | 275 |
| 2011 | 136 | 148 | 242 | 253 | 197 | 134 | 270 |
| 2012 | 137 | 151 | 246 | 280 | 180 | 135 | 281 |
| 2013 | 140 | 157 | 244 | 258 | 203 | 130 | 273 |
| 2014 | 133 | 159 | 244 | 264 | 206 | 131 | 272 |
| 2015 | 149 | 157 | 247 | 260 | 192 | 143 | 276 |
| 2016 | 145 | 151 | 245 | 267 | 174 | 141 | 273 |
| 2017 | 142 | 160 | 248 | 266 | 204 | 143 | 277 |
| Slope (anomaly/ year) | -0.75 | 0.00 | 0.33 | 0.37 | 0.25 | -0.67 | -0.36 |
| p -value | 0.16 | 1.00 | 0.01* | 0.49 | 0.77 | 0.32 | 0.25 |

D_1 : start of spring (SOS), D_2 : end of spring (EOS), D_3 : start of fall (SOF), D_4 : end of fall (EOF), and D_5 : peak of NDVI. Slope and p -values were calculated based on Theil-Sen's regression (** represents statistical significance $p < 0.05$). Start of growing season (SGS) and end of growing season (EGS) determined by a degree-day model were obtained from Nagai et al. (2020) for comparison with our phenological parameters

NDVI: normalized difference vegetation index

Table 9 Phenological parameters estimated by double sigmoid fitting of NDVI over Elgeeei

| YR | D_1 (SOS) | D_2 (EOS) | D_3 (SOF) | D_4 (EOF) | D_5 (peak) | SGS (Nagai et al. 2020) | EGS (Nagai et al. 2020) |
|--------------------------|-------------|-------------|-------------|-------------|--------------|-------------------------|-------------------------|
| 2003 | 134 | 169 | 238 | 265 | 206 | – | – |
| 2004 | 151 | 170 | 243 | 256 | 212 | – | – |
| 2005 | 143 | 159 | 247 | 255 | 214 | – | – |
| 2006 | 139 | 154 | 235 | 256 | 188 | – | – |
| 2007 | 128 | 157 | 244 | 264 | 206 | – | – |
| 2008 | 144 | 150 | 248 | 257 | 191 | – | – |
| 2009 | 144 | 150 | 249 | 257 | 194 | – | – |
| 2010 | 134 | 150 | 245 | 255 | 208 | 134 | 276 |
| 2011 | 128 | 145 | 243 | 255 | 203 | 132 | 270 |
| 2012 | 134 | 151 | 234 | 263 | 184 | 131 | 282 |
| 2013 | 124 | 162 | 242 | 262 | 211 | 129 | 275 |
| 2014 | 123 | 158 | 248 | 262 | 219 | 130 | 273 |
| 2015 | 148 | 158 | 246 | 260 | 195 | 137 | 278 |
| 2016 | 145 | 154 | 239 | 257 | 185 | 138 | 273 |
| 2017 | 137 | 162 | 248 | 263 | 215 | 135 | 276 |
| Slope (anomaly/ year) | -0.67 | -0.33 | 0.24 | 0.10 | 0.00 | – | – |
| p -value | 0.52 | 0.62 | 0.55 | 0.58 | 1.00 | – | – |

D_1 : start of spring (SOS), D_2 : end of spring (EOS), D_3 : start of fall (SOF), D_4 : end of fall (EOF), and D_5 : peak of NDVI. Start of growing season (SGS) and end of growing season (EGS) determined by a degree-day model were read from Nagai et al. (2020), for comparison with our phenological parameters. Slope and p -value of the SGS and EGS were not calculated since the available duration of the data differed from our phenological parameters

NDVI: normalized difference vegetation index

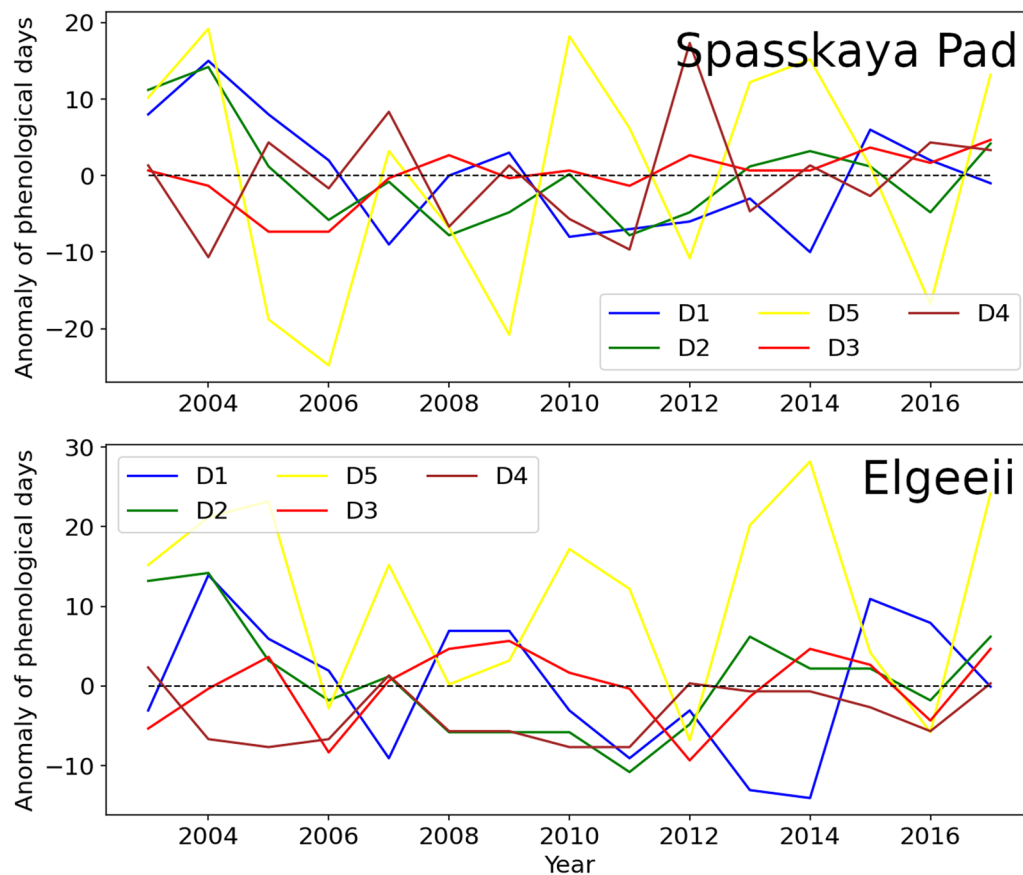


Fig. 10 Anomalous time series of phenological parameters (D_1 to D_5) at the Spasskaya Pad and Elgeeeii experimental sites. The parameters are described in Table 6. The black dashed line indicates the zero reference line

Abbreviations

| | |
|--------|--|
| AVHRR | Advanced Very High Resolution Radiometer |
| AMSR-E | Advanced Microwave Scanning Radiometer-EOS |
| ARD | Analysis ready data |
| DOY | Day of year |
| ECMWF | European Centre for Medium-Range Weather Forecasts |
| EOS | Earth Observing System |
| GCOM-W | Global Change Observation Mission-Water |
| GIMMS | Global Inventory Modeling and Mapping Studies |
| GIS | Geographic information system |
| GLDAS | Global Land Data Assimilation System |
| GRACE | The Gravity Recovery and Climate Experiment |
| HDF | Hierarchical data format |
| JAXA | Japan Aerospace Exploration Agency |
| ML | Machine learning |
| MODIS | Moderate Resolution Imaging Spectroradiometer |
| NDVI | Normalized difference vegetation index |
| NDWI | Normalized difference water index |
| RF | Random forest |
| RMSE | Root-mean-square error |
| SSM/I | Special Sensor Microwave/Imager |
| SWAMPS | Surface Water Microwave Product Series |
| SWE | Snow water equivalent |
| TMI | TRMM Microwave Imager |
| TWS | Terrestrial water storage |
| USGS | United States Geological Survey |
| VOD | Vegetation optical depth |

WAD2M Wetland Area and Dynamics for Methane Modeling

Acknowledgements

The authors thanks all members of the PAWCs project for research discussions.

Author contributions

HM, YI, and TH designed the research. HM and TS collected source data and implemented data fusion, validation, and visualization. HM conducted postprocessing for inundation mapping, added time-series trend analysis and phenological feature extraction, and wrote the first draft. All of the authors contributed to interpreting the results, discussing the findings, and revising the draft.

Funding

This research was supported by JSPS KAKENHI: 19H05668 (Pan-Arctic Water–Carbon Cycles; PAWCs).

Availability of data and materials

The datasets generated during the current study are available in the Pan-Arctic Water–Carbon Cycles (PAWCs) repository, at <https://enpawcs.home.blog/map/>. The datasets analyzed during the current study are available in the MODIS repository at <https://e4ftl01.cr.usgs.gov/MOLA>, the AMSR repository at <https://gportal.jaxa.jp>, and the VODCA repository at <https://zenodo.org/record/2575599#.YbFIX5HP36c>, last accessed on 2023/05/23. ERA5-Land, GRACE, and GLDAS can be downloaded via Google Earth Engine <https://earthengine.google.com/>, last accessed on 2023/07/28.

Declarations

Competing interests

The authors declare that they have no competing interests.

Author details

¹Geological Survey of Japan, National Institute of Advanced Industrial Science and Technology, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8567, Japan. ²Institute for Space-Earth Environmental Research (ISEE), Nagoya University, Nagoya 464-8601, Japan. ³Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa City, Chiba 277-8561, Japan. ⁴Graduate School of Science and Technology, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8577, Japan. ⁵Graduate School of Life and Agricultural Sciences, the University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan. ⁶Department of Geography, Tokyo Metropolitan University, 1-1 Minami Osawa, Hachioji, Tokyo 192-0397, Japan. ⁷Institute of Arctic Climate and Environment Research, Japan Agency for Marine-Earth Science and Technology, 2-15, Natsushima-cho, Yokosuka, Kanagawa 237-0061, Japan. ⁸Institute of Science and Technology, Niigata University, Niigata 950-2181, Japan. ⁹Center for Environmental Remote Sensing (CERES), Chiba University, Chiba 263-8522, Japan.

Received: 22 August 2023 Accepted: 17 February 2024

Published online: 23 February 2024

References

- AMAP (2021) Arctic climate change update 2021: key trends and impacts. Summary for Policy-makers. Arctic Monitoring and Assessment Programme (AMAP), Tromsø, Norway. 16
- Armstrong RL, Brodzik MJ (2001) Recent northern hemisphere snow extent: a comparison of data derived from visible and microwave satellite sensors. *Geophys Res Lett* 28(19):3673–3676
- Bartsch A, Balzter H, George C (2009) The influence of regional surface soil moisture anomalies on forest fires in Siberia observed from satellites. *Environ Res Lett* 4:045021
- Beamish A, Reynolds MK, Epstein H, Frost GV, Macander MJ, Bergstedt H, Bartsch A, Kruse S, Miles V, Tanis CM, Heim B, Fuchs M, Chabrilat S, Shevtsova I, Verdonen M, Wagner J (2020) Recent trends and remaining challenges for optical remote sensing of Arctic tundra vegetation: a review and outlook. *Remote Sens Environ* 246:111872
- Belgiu M, Stein A (2019) Spatiotemporal image fusion in remote sensing. *Remote Sens* 11:818
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32
- Brown CF, Brumby SP, Guzder-Williams B et al (2022) Dynamic World, near real-time global 10 m land use land cover mapping. *Sci Data* 9:251
- Buitenwerf R, Rose L, Higgins S (2015) Three decades of multi-dimensional change in global leaf phenology. *Nat Clim Change* 5:364–368
- Chasmer L, Cobbaert D, Mahoney C, Millard K, Peters D, Devito K, Brisco B, Hopkinson C, Merchant M, Montgomery J, Nelson K, Miemann O (2020) Remote sensing of boreal wetlands 1: data use for policy and management. *Remote Sens* 12(8):1320
- ECMWF (2023) ERA5-Land: data documentation. <https://confluence.ecmwf.int/display/CKB/ERA5-Land%3A+data+documentation> (last accessed on 2023.02.22)
- Fily M, Royer A, Goita K, Prigent C (2003) A simple retrieval method for land surface temperature and fraction of water surface determination from satellite microwave brightness temperatures in sub-arctic areas. *Remote Sens Environ* 85:328–338
- Fluet-Chouinard E, Lehner B, Rebelo LM, Papa F, Hamilton SK (2015) Development of a global inundation map at high spatial resolution from topographic downscaling of coarse-scale remote sensing data. *Remote Sens Environ* 158:348–361
- Guimberteau M, Zhu D, Maignan F, Huang Y, Yue C, Dantec-Nédélec S, Ottlé C, Jornet-Puig A, Bastos A, Laurent P et al (2018) ORCHIDEE-MICT (v8.4.1), a land surface model for the high latitudes: Model description and validation. *Geosci Model Dev* 11:1:121–163
- Hantemirov RM, Corona C, Guillet S, Shiyatov SG, Stoffel M, Osborn TJ, Melvin TM, Gorlanova LA, Kukarskih VV, Surkov AY, von Arx G, Fonti P (2022) Current Siberian heating is unprecedented during the past seven millennia. *Nat Commun* 13:4968
- Hatfield JL, Asrar G, Kanemasu ET (1984) Intercepted photosynthetically active radiation estimated by spectral reflectance. *Remote Sens Environ* 14(1–3):65–75
- HDF-EOS Tools and Information Center (2023) HDF-EOS to Geotiff conversion tool (HEG). <https://hdfeos.org/software/heg.php> (last accessed on 2023.2.22)
- Ide R, Oguma H (2013) A cost-effective monitoring method using digital time-lapse cameras for detecting temporal and spatial variations of snowmelt and vegetation phenology in alpine ecosystems. *Eco Inform* 16:25–34
- Isola P, Zhu JY, Zhou T, Efros AA (2017) Image-to-image translation with conditional adversarial networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017, 5967–5976
- JAXA (2023) Globe Portal System. <https://gportal.jaxa.jp/gpr/?lang=en> (last accessed on 2023.02.20)
- Jensen K, McDonald K (2019) Surface Water Microwave Product Series version 3: a near-real time and 25-year historical global inundated area fraction time series from active and passive microwave remote sensing. *IEEE Geosci Remote Sens Lett* 16(9):1402–1406
- Kim Y, Kimball JS, Zhang K, McDonald KC (2012) Satellite detection of increasing northern hemisphere non-frozen seasons from 1979 to 2008: implications for regional vegetation growth. *Remote Sens Environ* 121:472–487
- Kimball JS, Jones LA, Zhang K, Heinsch FA, McDonald KC, Oechel WC (2009) A satellite approach to estimate land-atmosphere CO₂ exchange for boreal and arctic biomes using MODIS and AMSR-E. *IEEE Trans Geosci Remote Sens* 47(2):569–587
- Kotani A, Kononov AV, Ohta T, Maximov TC (2014) Temporal variations in the linkage between the net ecosystem exchange of water vapor and CO₂ over boreal forests in eastern Siberia. *Ecophysiology* 7:209–225
- Kotani A, Saito A, Kononov AV, Petrov RE, Maximov TC, Iijima Y, Ohta T (2019) Impact of unusually wet permafrost soil on understory vegetation and CO₂ exchange in a larch forest in eastern Siberia. *Agric for Meteorol* 265:295–309
- Kushida K, Isaev AP, Maximov TC, Takao G, Fukuda M (2007) Remote sensing of upper canopy leaf area index and forest floor vegetation cover as indicators of net primary productivity in a Siberian larch forest. *J Geophys Res Biogeogr* 112:G02003
- Landerer FW, Swenson SC (2012) Accuracy of scaled GRACE terrestrial water storage estimates. *Water Resour Res* 48(W04531):11
- Lehner B, Döll P (2004) Development and validation of a global database of lakes, reservoirs and wetlands. *J Hydrol* 296:1–22
- Leroy M, Bicheron P, Brockmann C, Kramer U, Miras B, Huc M, Nino F, Defourny P, Vancutsem C, Petit D, Amberg V, Berthelt Bm Arino O, Ranera F (2006) GlobCover: a 300 m global land cover product for 2005 using ENVISAT MERIS time series. ISPRS Commission VII Mid-Term Symposium: Remote Sensing: from Pixels to Processes, Enschede (NL), May 2006
- Li B, Rodell M, Kumar S, Beaudoin HK, Getirana A, Zaitchik BF, de Goncalves LG, Cossetin C, Bhanja S, Mukherjee A, Tian S, Tangdamrongsub N, Long D, Nanteza J, Lee J, Policelli F, Goni IB, Daira D, Bila M, de Lannoy G, Mocko D, Steele-Dunne SC, Save H, Bettadpur S (2019) Global GRACE data assimilation for groundwater and drought monitoring: advances and challenges. *Water Resour Res* 55:7564–7586
- Madaeni F, Lhissou R, Chokmani K, Raymond S, Gauthier Y (2020) Ice jam formation, breakup and prediction methods based on hydroclimatic data using artificial intelligence: a review. *Cold Reg Sci Technol* 174:103032
- Mavrovic A, Sonnentag O, Lemmetyinen J, Baltzer JL, Kinnard C, Roy A (2023) Reviews and syntheses: recent advances in microwave remote sensing in support of arctic-boreal carbon cycle science. EGU sphere (preprint): <https://doi.org/10.5194/egusphere-2023-137>
- May JL, Hollister RD, Betway KR, Harris JA, Tweedie CE, Welker JM, Gould WA, Oberbauer SF (2020) NDVI changes show warming increases the length of the green season at Tundra communities in Northern Alaska: a fine-scale analysis. *Front Plant Sci* 11:1174
- Mirza M, Osindero S (2014) Conditional generative adversarial nets. arXiv: 1411.1784
- Mizuochi H, Nishiyama C, Ridwansyah I, Nasahara KN (2018) Monitoring of an Indonesian tropical wetland by machine learning-based data fusion of passive and active microwave sensors. *Remote Sens* 10(8):1235

- Mizuochi H, Iijima Y, Nagano H, Kotani A, Hiyama T (2021a) Dynamic mapping of subarctic surface water by fusion of microwave and optical satellite data using conditional adversarial networks. *Remote Sens* 13(2):175
- Mizuochi H, Ducharme A, Cheruy F, Ghattas J, Al-Yaari A, Wigneron JP, Bastrikov V, Peylin P, Maignan F, Vuichard N (2021b) Multivariate evaluation of land surface processes in forced and coupled modes reveals new error sources to the simulated water cycle in the IPSL (Institute Pierre Simon Laplace) climate model. *Hydrol Earth Syst Sci* 25(4):2199–2221
- Moesinger L, Dorigo W, de Jeu R, van der Schalie R, Scanlon T, Teubner I, Forkel M (2020) The global long-term microwave vegetation optical depth climate archive (VODCA). *Earth Syst Sci Data* 12(1):177–196
- Moradzadeh M, Srivastava PK (2021) A new model for an improved AMSR2 satellite soil moisture retrieval over agricultural areas. *Comput Electron Agric* 186:106205
- Myers E, Kerekes J, Daughtry C, Russ A (2019) Assessing the impact of satellite revisit rate on estimation of corn phenological transition timing through shape model fitting. *Remote Sens* 11(21):2558
- Myers-Smith IH et al (2020) Complexity revealed in the greening of the Arctic. *Nat Clim Change* 10:106–117
- Myneni RB, Dong J, Tucker CJ, Kaufmann RK, Kauppi PE, Liski J, Zhou L, Alexeyev V, Hughes MK (2001) A large carbon sink in the woody biomass of Northern forests. *Proc Natl Acad Sci USA* 98(26):14784–14789
- Nagai S, Kotani A, Morozumi T, Kononov AV, Petrov RE, Shakhmatov R, Ohta T, Sugimoto A, Maximov TC, Suzuki R, Tei S (2020) Detection of year-to-year spring and autumn bio-meteorological variations in siberian ecosystems. *Polar Sci* 25:100534
- Nagai S, Kobayashi H, Suzuki R (2019) Remote sensing of vegetation. In: Ohta T, Hiyama T, Iijima Y, Kotani A, Maximov TC (ed) *Water-carbon dynamics in Eastern Siberia*, 1st edn. Springer Nature, Singapore
- Nagano H, Kotani A, Mizuochi H, Ichii K, Kanamori H, Hiyama T (2022) Contrasting 20-year trends in NDVI at two Siberian larch forests with and without multiyear waterlogging-induced disturbances. *Environ Res Lett* 17:025003
- Ohta T, Kotani A, Iijima Y, Maximov TC, Ito S, Hanamura M, Kononov AV, Maximov AP (2014) Effects of waterlogging on water and carbon dioxide fluxes and environmental variables in a Siberian larch forest, 1998–2011. *Agric for Meteorol* 188:64–75
- Owe M, de Jeu R, Holmes T (2008) Multisensor historical climatology of satellite-derived global land surface moisture. *J Geophys Res Earth* 113:F01002
- Papa F, Prigent C, Jimenez C, Aires F, Rossow WB (2010) Interannual variability of surface water extent at global scale, 1993–2004. *J Geophys Res* 115:D12111
- Pekel JF, Cottam A, Gorelick N, Belward AS (2016) High-resolution mapping of global surface water and its long-term changes. *Nature* 540:418–422
- Perry CR, Lautenschlager LF (1984) Functional equivalence of spectral vegetation indices. *Remote Sens Environ* 14:169–182
- Petty GW (2006) *A first course in atmospheric radiation*, 2nd edn. Sundog Publishing, Madison, WI, USA, pp 99–100
- Pinzon JE, Tucker CJ (2014) A non-stationary 1981–2012 AVHRR NDVI3g time series. *Remote Sens* 6:6929–6960
- Reiche J, Hamunyela E, Verbesselt J, Hoekman D, Herold M (2018) Improving near-real time deforestation monitoring in tropical dry forests by combining dense Sentinel-1 time series with Landsat and ALOS-2 PALSAR-2. *Remote Sens Environ* 204:147–161
- Sakai T, Hatta S, Okumura M, Hiyama T, Yamaguchi Y, Inoue G (2015) Use of Landsat TM/ETM+ to monitor the spatial and temporal extent of spring breakup floods in the Lena River, Siberia. *Int J Remote Sens* 36:719–733
- Sakumura C, Bettadpur S, Bruinsma S (2014) Ensemble prediction and inter-comparison analysis of grace time-variable gravity field models. *Geophys Res Lett* 41:1389–1397
- Sawada Y, Koike T, Aida K, Toride K, Walker JP (2017) Fusing microwave and optical satellite observations to simultaneously retrieve surface soil moisture, vegetation water content, and surface soil roughness. *IEEE Trans Geosci Remote Sens* 55(11):6195–6206
- Schroeder R, McDonald KC, Chapman BD, Jensen K, Podest E, Tessler ZD, Bohn TJ, Zimmermann R (2015) Development and evaluation of a multi-year fractional surface water data set derived from Active/Passive microwave remote sensing data. *Remote Sens* 7:16688–16732
- Shestakova AA, Fedorov AN, Torgovkin YI, Konstantinov PY, Vasyliov NF, Kalinicheva SV, Samsonova VV, Hiyama T, Iijima Y, Park H, Iwahana G, Goroknov AN (2021) Mapping the main characteristics of permafrost on the basis of a permafrost-landscape map of Yakutia using GIS. *Land* 10(5):462
- Suzuki K, Matsuo K, Yamazaki D, Ichii K, Iijima Y, Papa F, Yanagi Y, Hiyama T (2018) Hydrological variability and changes in the Arctic circumpolar tundra and the tree largest pan-Arctic river basins from 2002 to 2016. *Remote Sens* 10(3):402
- Suzuki K, Hiyama T, Matsuo IK, Iijima Y, Yamazaki D (2020) Accelerated continental-scale snowmelt and ecohydrological impacts in the four largest Siberian river basins in response to spring warming. *Hydrol Process* 34:3867–3881
- Suzuki K, Matsuo K (2019) Remote sensing of terrestrial water. In: Ohta T, Hiyama T, Iijima Y, Kotani A, Maximov TC (ed) *Water-Carbon Dynamics in Eastern Siberia*, 1st edn. Springer Nature, Singapore
- Swenson SC (2012) TELLUS LAND_NC_RL05. Ver. 5.0. PO.DAAC, CA, USA. Dataset accessed [2023-07-28] at <https://doi.org/10.5067/TELND-NC005>
- Tucker CJ, Pinzon JE, Brown ME, Slayback DA, Pak EW, Mahoney R, Vermote EF, Saleous N (2005) An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. *Int J Remote Sens* 26:4485–4498
- Twele A, Cao W, Plank S, Martinis S (2016) Sentinel-1-based flood mapping: a fully automated processing chain. *Int J Remote Sens* 37(13):2990–3004
- USGS (2023) NASA's Land Processes Distributed Active Archive Center (LP DAAC) data pool. <https://e4ftl01.cr.usgs.gov/MOLA/MYD09A1.061/> (last accessed on 2023.2.20)
- Velicogna I, Tong J, Zhang T, Kimball JS (2012) Increasing subsurface water storage in discontinuous permafrost areas of the Lena River basin, Eurasia, detected from GRACE. *Geophys Res Lett* 39:09403
- Watts JD, Kimball JS, Jones LA, Schroeder R, McDonald KC (2012) Satellite Microwave remote sensing of contrasting surface water inundation changes within the Arctic-Boreal Region. *Remote Sens Environ* 127:223–236
- Witze A (2020) Why arctic fires are bad news for climate change. *Nature* 585:336–337
- Xu H (2006) Modification of normalized difference water index (NDWI) to enhance open water features in remotely sensed imagery. *Int J Remote Sens* 27:3025–3033
- Yamazaki D, Trigg MA, Ikeshima D (2015) Development of a global ~90 m water body map using multi-temporal Landsat images. *Remote Sens Environ* 171:337–351
- Yan D, Zhang X, Nagai S, Yu Y, Akitsu T, Nasahara KN, Ide R, Maeda T (2019) Evaluating land surface phenology from the Advanced Himawari Imager using observations from MODIS and Phenological Eyes Network. *Int J Appl Earth Obs Geoinformation* 79:71–83
- Yang D, Zhao Y, Armstrong R, Robinson D, Brodzik MJ (2007) Streamflow response to seasonal snow cover mass changes over large Siberian watersheds. *J Geophys Res Space Phys* 112:F02522
- Yokohata T, Saito K, Ito A, Ohno H, Tanaka K, Hajima T, Iwahana G (2020) Future projection of greenhouse gas emissions due to permafrost degradation using a simple numerical scheme with a global land surface model. *Prog Earth Planet Sci* 7:56
- Zakharov I, Kapfer M, Hornung J, Kohlsmith S, Puestow T, Howell M, Henschel MD (2020) Retrieval of surface soil moisture from Sentinel-1 time series for reclamation of wetland sites. *IEEE J-STARS* 13:3569–3578
- Zhang Z, Fluet-Chouinard E, Jensen K, McDonald K, Hugelius G, Gumbrecht T, Carroll M, Prigent C, Bartsch A, Poulter B (2021) Development of the global dataset of Wetland Area and Dynamics for Methane Modeling (WAD2M). *Earth Syst Sci Data* 13(5):2001–2023
- Zhao Q, Yu L, Du Z, Peng D, Hao P, Zhang Y, Gong P (2022) An overview of the applications of Earth observation satellite data: impacts and future trends. *Remote Sens* 14:1863
- Zhu L, Li W, Wang H, Deng X, Tong C, He S, Wang K (2023) Merging microwave, optical, and reanalysis data for 1 km daily soil moisture by triple collocation. *Remote Sens* 15(1):159

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Hiroki Mizuochi received a Ph.D. degree in environmental studies from the University of Tsukuba in 2017. HM is currently (1) a researcher with the Geological Survey of Japan, National Institute of Advanced Industrial Science and Technology, (2) a visiting associate professor at ISEE, Nagoya University, and (3) a visiting associate professor in the Graduate School of Frontier Sciences, the University of Tokyo. HM is involved in the ASTER and HISUI Earth observation remote-sensing projects, and a JSPS Grant-in-Aid for Scientific Research project entitled, "Pan-Arctic Water-Carbon Cycles (PAWCs)". HM is a member of the Remote Sensing Society of Japan, the Japan Society of Photogrammetry and Remote Sensing, and the Japan Geoscience Union.

Taiga Sasagawa is a Ph.D. student at the University of Tsukuba and a research fellow of the Japan Society for the Promotion of Science (JSPS). TS is a member of the Japanese Forest Society, Japan Society of Forest Planning, the Society of Agricultural Meteorology of Japan, the Remote Sensing Society of Japan, the American Geophysical Union, and the Japan Geoscience Union.

Akihiko Ito received a Ph.D. in biological sciences from the University of Tsukuba in 2001. AI is currently (1) a professor at the Graduate School of Life and Agricultural Sciences, the University of Tokyo, (2) a senior researcher at the National Institute for Environmental Studies, and (3) an affiliated researcher at the Japan Agency for Marine-Earth Science and Technology. AI is involved in a JSPS Grant-in-Aid for Scientific Research project entitled, "Pan-Arctic Water-Carbon Cycles (PAWCs)". AI is a member of the Ecological Society of Japan, the Meteorological Society of Japan, and the Japan Geoscience Union.

Yoshihiro Iijima received a Ph.D. in science from Tokyo Metropolitan University in 2001. YI is currently a professor at the Department of Geography, Tokyo Metropolitan University. YI is involved in the Arctic Challenge for Sustainability II (ArCS II) project of MEXT, and Pan-Arctic Water-Carbon Cycles JSPS scientific research. YI is a member of the Association of Japanese Geographers, the Remote Sensing Society of Japan, the Japanese Society of Snow and Ice, and the Japan Geoscience Union.

Hotaek Park received a Ph.D. in agricultural science, Nagoya University in 2000 and is currently a senior scientist at the Japan Agency for Marine-Earth Science and Technology (JAMSTEC), a visiting professor at ISEE, Nagoya University, and is involved in Pan-Arctic Water-Carbon Cycles funded by JSPS.

Hirohiko Nagano received a Ph.D. in agriculture from Chiba University in 2013. HN is currently an assistant professor at the Institute of Science and Technology, Niigata University. HN is involved in a JSPS Grant-in-Aid for Scientific Research (KAKENHI) project entitled, "Pan-Arctic Water-Carbon Cycles (PAWCs)". HN is a member of the American Geophysical Union, the Japanese Society of Soil Science and Plant Nutrition, the Society of Agricultural Meteorology of Japan, the Japanese Forest Society, and the Japan Geoscience Union.

Kazuhito Ichii received a Ph.D. in science from Nagoya University in 2002. KI is currently a professor at the Center for Environmental Remote Sensing (CERES), Chiba University, Japan.

Tetsuya Hiyama received a Ph.D. in science from the University of Tsukuba in 1995. TH is the principal investigator of the JSPS Grant-in-Aid for Scientific Research (KAKENHI) project entitled "Pan-Arctic Water-Carbon Cycles (PAWCs)".