The Use of Satellite Information (MODIS/Aqua) for Phenological and Classification Analysis of Plant Communities

Yulia Ivanova 1,*, Anton Kovalev 2, Oleg Yakubailik 3 and Vlad Soukhovolsky 4

1 Institute of Biophysics SB RAS, Federal Research Center “Krasnoyarsk Science Center SB RAS”, Academgorodok 50-50, 660036 Krasnoyarsk, Russia
2 Federal Research Center “Krasnoyarsk Science Center SB RAS”, Academgorodok 50, 660036 Krasnoyarsk, Russia
3 Institute of Computational Modeling SB RAS, Federal Research Center “Krasnoyarsk Science Center SB RAS”, Academgorodok 50-44, 660036 Krasnoyarsk, Russia
4 Sukachev Institute of Forest SB RAS, Federal Research Center “Krasnoyarsk Science Center SB RAS”, Academgorodok 50-28, 660036 Krasnoyarsk, Russia
* Correspondence: lulja@yandex.ru; Tel.: +73-9124-9432-8

Abstract: Vegetation indices derived from remote sensing measurements are commonly used to describe and monitor vegetation. However, the same plant community can have a different NDVI (normalized difference vegetation index) depending on weather conditions, and this complicates classification of plant communities. The present study develops methods of classifying the types of plant communities based on long-term NDVI data (MODIS/Aqua). The number of variables is reduced by introducing two integrated parameters of the NDVI seasonal series, facilitating classification of the meadow, steppe, and forest plant communities in Siberia using linear discriminant analysis. The quality of classification conducted by using the markers characterizing NDVI dynamics during 2003–2017 varies between 94% (forest and steppe) and 68% (meadow and forest). In addition to determining phenological markers, canonical correlations have been calculated between the time series of the proposed markers and the time series of monthly average air temperatures. Based on this, each pixel with a definite plant composition can be characterized by only four values of canonical correlation coefficients over the entire period analyzed. By using canonical correlations between NDVI and weather parameters and employing linear discriminant analysis, one can obtain a highly accurate classification of the study plant communities.

Keywords: boreal forests and ecosystems; NDVI (normalized difference vegetation index); classification of plant communities; linear discriminant analysis

1. Introduction

Plant phenology is determined by the dates of the seasonal biological events in the plant life cycle such as the emergence of leaves, the emergence of flowers, etc. The occurrence of phenological events is influenced by plant interactions with the air, soil, and water flows, and, thus, phenology is one of the most reliable indicators of plant seasonal dynamics. On the one hand, phenology is a sensitive indicator of a changing climate and, on the other hand, phenology is related to the productivity and biophysical properties of the ecosystem [1–4]. Thus, changes in phenology dynamics can correspond to the effects of both local and global climate change on plant ecology [5,6]. Traditionally, observations of phenological events are performed by people, but in the last four decades, these have been studied using remote sensing-based products such as different vegetation indices [7–10]. Satellite data have
considerable potential for monitoring vegetation dynamics on the regional and global levels, as airborne sensors enable coordinated measurements on the spatial and temporal scales. Land surface phenology (LSP) is the study of the spatial and temporal development of vegetation surface using satellite measurements [11]. LSP is indirectly related to plant phenology through radiation absorption and reflection, but it is influenced by changes in the state of the atmosphere, cloud and snow covers, effects of reflection, and non-climatic factors such as biogenic or anthropogenic impacts [12]. Satellite-derived vegetation indices are spectral indicators of plant photosynthesis and metabolic rates [13–16]. It is, however, difficult to compare the data of the ground and satellite phenology because of the discrepancy between their spatial-temporal scales. The main source of uncertainty is the errors arising from comparing targeted ground observations with pixel calculations in remote sensing data. For instance, traditional phenology monitors growth phases determined for a limited number of certain plants, while airborne sensors record electromagnetic reflection from a large land plot [12,17,18].

Like ground phenological observations, satellite-derived data focus on the main dates of phenological events, which are plotted on the annual curves of vegetation indices, e.g., the dates of the beginning and end of the plant growing period, maxima of vegetation indices, etc. Methods are being developed that could be used to extract phenological markers from the time series of vegetation indices [11,19–22].

Studies of the dynamics of vegetation indices take into account both intra-annual and interannual changes related to climate variations [23,24] or natural or human-induced changes in the vegetation [25,26]. Thus, temporal trajectory of vegetation indices is used to describe ecosystem dynamics, study vegetation types, detect damages, and monitor vegetation.

Different types of plants in communities show their individual responses to changes in the temperature, humidity, and other weather phenomena. The individual sensitivity and the time of response to climate change of each plant type constitute the specific response of the whole plant community, which determines the dynamics of phenological sequences for each plant species and each plant community [27–30].

Over the past four decades, the number of studies devoted to classification of plant communities based on analysis of the time series of vegetation indices has been increasing steadily [31]. Vegetation is categorized using various methods, e.g., supervised machine learning algorithms [32–36]. Other methods include singular value decomposition [37,38] and methods based on regression and correlation analysis, which are used more often [39–47]. To detect changes in the vegetation based on analysis of long-term satellite datasets, a number of authors propose using wavelet decomposition [48,49] and Fourier analysis [50,51]. In these studies, analysis of time series involves separation of the signal from noise and, depending on the method employed, researchers perform a certain transformation aimed at selecting the dominating components of the time series signal variation. These methods are used to mark changes in the time series that are caused by seasonality and interannual variations in weather conditions for data analysis. The existing methods of detecting changes minimize seasonal variations and enable researchers to focus on certain periods of the year (such as the growing period) or investigate the properties of the obtained time series of the vegetation indices [52–54] instead of taking seasonality into account explicitly. Maximum likelihood estimation [55,56] and artificial neural networks [57,58] are used to increase the accuracy of classifications. There are also different software products for observing phenological dynamics [59,60].

A number of LSP studies show that not only climatic factors but also fires, land degradation, attacks of insects, floods, deforestation, etc. considerably affect phenology of plant communities [12,61–64]. Therefore, a more reliable approach is needed to reveal long-term phenological variations, which are determined by changes in the composition and boundaries of the plant communities caused by the effects of climate or other factors. The concept of a significant change in plant community should be defined. It is important for researchers to be able to track and reveal changes in space and time.

The purpose of this study is to develop methods of classification of plant communities based on long-term NDVI data. In order to reduce the number of variables, we introduce two
integrated parameters of the seasonal NDVI series as markers of phenological dynamics characterizing plant community.

2. Materials and Methods

2.1. Data

We investigated plant communities growing in five study areas. The study areas were sufficiently large to enable decoding of the data with 250 m spatial resolution provided by MODIS imaging sensor mounted on TERRA & AQUA satellites.

Study area no. 1 (Figure 1) represents forest-steppe vegetation. The area is located 40 km away from the city of Krasnoyarsk and is geographically positioned in the Krasnoyarsk Forest-Steppe, which is located in the south of Middle Siberia. The study area is flat terrain with micro-depressions. The elevation of the area at the northern point is 200 m above sea level, declining gradually. Mixed forest, consisting largely of pine and birch trees (Betula pendula Roth, Pinus sylvestris L.), grows in the north and south of the area, and herbaceous plants occupy its central part (Bromus inermis (Leyss.) Holub, Stipa pennata L., Poa pratensis L., Potentilla tanacetifolia Wild. Ex Schltdl., Elytrigia repens (L.) Nevski).

Figure 1. Location of the study areas.

Study areas no. 2–4 are located in the central part of the Shira District, in the steppe and forest-steppe zones of the Republic of Khakassia in the south of the Krasnoyarsk Territory, in the Minusinsk Depression. The area is undulating-flat terrain, with flat regions separated from each other by monoclinic cuesta ridges with strongly asymmetric. The Minusinsk Depression has an extreme
continental climate, with considerable yearly and daily temperature variations, low precipitation, strong winds, and low humidity. The amount of precipitation differs depending on the season: about 50%-60% of the total amount occurs in summer and no more than 10% in winter. The plant growing period is 130–160 days (academic and practical guide on climate of the USSR [65]).

The study areas are occupied by steppe plant communities dominated by herbaceous perennials (Table 1) with long growing periods. The Khakassia steppes, being very old and providing favorable conditions for plants, show high species saturation (study areas no. 2 and 3). Study area no. 4 represents lowland meadows located in riverine valleys and topographic lows. The northern slopes of the hills and low mountains are covered by mixed birch-larch forests (Larix sibirica L., Betula pendula Roth.), study areas no. 3 and 4.

Study area no. 5, which is located 45 km away from Krasnoyarsk (Figure 1), is covered by coniferous forest (fir and pine trees (Abies sibirica Ledeb., Pinus sylvestris L.). This is the forested part of the Krasnoyarsk Forest-Steppe. The plant growing period in the Krasnoyarsk Forest-Steppe is 130–150 days, and the freeze-free period is 90 days. The annual amount of precipitation is 366 mm, 70% of which occurs from May through September.

In each study area, we considered several pixels of MODIS data (Tables 1 and 2) to be further used to obtain time series of NDVI values, conduct discriminant analysis, and classify plant communities growing in these areas.

Table 1. Description of plant communities growing in study areas.

<table>
<thead>
<tr>
<th>Area No.</th>
<th>Latitude/Longitude</th>
<th>Plant Communities</th>
<th>Number of Pixels</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>56.352/93.013</td>
<td>Steppes (Bromus inermis (Leyss.) Holub, Stipa pennata L., Poa pratensis L., Potentilla tanacetifolia Wild. Ex Schidbl., Elytrigia repens (L.) Nevski); Coniferous-deciduous forests (Betula pendula Roth., Pinus sylvestris L.)</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>54.486/90.425</td>
<td>Steppes (Stipa capillata L., Carex pediformis C.A. Mey., Koeleria cristata (L.) Pers., Festuca valesiaca Gaudin, Taraxacum officinale Wigg.); Meadows (Iris biglumis Vahl, Elytrigia repens (L.) Nevski, Festuca pseudovina Hack. Ex Wiesb., Plantago media L.)</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>54.702/90.527</td>
<td>Steppes (Stipa capillata L., Carex pediformis C.A. Mey., Koeleria cristata (L.) Pers., Festuca valesiaca Gaudin, Artemisia frigida Willd., Taraxacum officinale Wigg.); Coniferous-deciduous forests (Larix sibirica L., Betula pendula Roth.)</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>54.520/89.744</td>
<td>Meadows (Iris ruthenica Ker-Gawl., Ranunculus polyanthemos L., Phleum phleoides (L.) Karst, Helictotrichon schellianum (Hack.) Kitag., Artemisia tanacetifolia L.); Coniferous-deciduous forests (Larix sibirica L., Betula pendula Roth.)</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>56.146/92.204</td>
<td>Coniferous forests (Abies sibirica Ledeb., Pinus sylvestris L.)</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 2. Photos of plant communities growing in study areas.

| Area No.1 | Steppes; Coniferous-deciduous forests. |
Vegetation indices are commonly used to study the state and integrated changes in the vegetation cover on different spatial scales. In the present study, plant communities were classified using normalized difference vegetation index (NDVI), which is based on the difference between the red and near infrared light reflected by vegetation detected by satellite sensors [66]. NDVI is successfully used to estimate the state and changes of the vegetation cover, as NDVI values are related to photosynthetically active radiation. In the present study, NDVI was calculated from the standard formula

\[ NDVI = \frac{NIR - Red}{NIR + Red} \]  

(1)
where \(NIR\) and \(Red\) are reflectance of sunlight values in the near infrared and red spectral bands for a given point on Earth’s surface. The biophysical interpretation of \(NDVI\) is the fraction of absorbed photosynthetically active radiation.

MOD09Q1/MYD09Q1 data of MODIS Terra/Aqua sensor were used to calculate NDVI vegetation indices. These data are the reflectance values of NIR and Red spectral bands with spatial resolution (pixel size) of 250 meters, with a time step of 8 days. For each 8-day period, the best possible value is selected from the source information of the regular daily survey according to the maximum quality criterion taking into account atmospheric conditions (absence of clouds and shadows from clouds, transparency of the atmosphere). The data were obtained from the Earth Observing System Data and Information System EOSDIS [67]. NDVI values are given for the period from 2003 through 2017. A typical seasonal NDVI time series is shown in Figure 2.

![Figure 2](image.png)

**Figure 2.** A typical NDVI \((n,i)\) time series (where \(n\) is the number of the day in a year; \(i\) is the number of the year, study area no. 4, meadow, \(i = 2003\)). (1) a seasonal time series of NDVI, (2) filtered series 1, (3) series 2 clipped at critical value \(NDVIF_{crit} = NDVIF_{min} + 0.2(NDVI_{max} - NDVI_{min})\), (4) parabolic approximation of series 3; (5) critical value of NDVI \((n)\).

The curve of NDVI \((n)\) seasonal dynamics has a complex shape (Figure 2), and to simplify the description of this curve, we propose the following procedure of the 'reduction' of the NDVI \((i)\) seasonal time series in year \(i\):

1. Determine the NDVI \((i)_{max}\) and NDVI \((i)_{min}\) for season \(i\).
2. Filter the NDVI \((n)\) time series in the range of values between \(n_0(i)\) and \(n_f(i)\) with a high-frequency moving average filter including five points:

\[
NDVIF (n) = \frac{1}{5} (NDVI (n - 2) + NDVI (n - 1) + NDVI (n) + NDVI (n + 1) + NDVI (n + 2))
\]  

(2)

3. Determine the \(NDVIF(i)_{crit}\) for the season: \(NDVIF(i)_{crit} = NDVIF_{min} + 0.20 \times (NDVIF(i)_{max} - NDVIF(i)_{min})\).
4. Determine the \(n_0(i)\)—the starting time point, before which all NDVIF values of the current year are below \(NDVIF(i)_{crit}\), and \(n_f(i)\)—the end point, after which all NDVIF values of the current year are below \(NDVIF(i)_{crit}\).
5. Approximate the filtered series \(NDVIF(n)\) with the parabolic Equation (3) using the nonlinear least squares method. Now the NDVI \((n)\) seasonal dynamics in the \((n_0, n_f)\) range of values is characterized by three parameters: \(a(i), b(i), c(i)\).

\[
NDVIF (n, i) = a (i)n^2 + b (i)n + c (i)
\]  

(3)
6. Instead of parameters \(a(i), b(i), c(i)\), use two new variables determined from the NDVIF \((n, i)\) Equation (4) (Figure 3): the maximum seasonal value \(\text{NDVIF}(i)_{\max}\), determined using a common procedure from equation:

\[
\frac{d\text{NDVIF}(n, i)}{dn} = 2a(i)n + b(i) = 0
\]

\[
n_{\max} = -\frac{b(i)}{2a(i)}; \text{NDVIF}(i)_{\max} = a(i)n_{\max}^2 + c(i)
\]

The rate \(\frac{\text{NDVIF}(i)}{dn}\) = \(2a(i)n_1^2 + b(i)\) of the NDVIF increase at point \(n_{1/2}(i) = \frac{(n_{\max}(i) - n_0(i))}{2} + n_0(i)\), located in the middle of the range between points \(n_0(i)\) and \(n_{\max}(i)\). This parameter corresponds approximately to the date of emergence of leaves [68]. The procedure for recalculating of the NDVI seasonal dynamics curve is shown in Figure 3.

![Figure 3. Procedure for recalculating of the NDVI seasonal dynamics curve. Point 1 \((n_{\max}, \text{NDVIF}_{\max})\); point 2 \((\frac{n_{\max} - n_0}{2}, n_0)\); point 3 \((n_0, \text{NDVIF}(i)_{\crit})\), \(\alpha = \frac{d\text{NDVIF}(i)}{dn} = 2a(i)n_1^2 + b(i)\)](image)

In the first stage of the analysis, plant communities were classified using the NDVIF \(_{\max}\) and \(\text{dNDVIF}/\text{dn}\) datasets for 2003–2017 (i.e., 15 data pairs were used for each NDVI image). The NDVIF \(_{\max}\) and \(\text{dNDVIF}/\text{dn}\) datasets for the areas with different plant communities (meadow, steppe, forest) were compared using linear discriminant analysis [69] in order to differentiate the ranges of NDVIF \(_{\max}\) and \(\text{dNDVIF}/\text{dn}\) values for different plant communities and to rate the classification of plant communities based on these parameters using the classification template.

Canonical correlation analysis was used to relate the remote sensing data to weather parameters [70–72]. This method is used to measure the strength of association between different types of value sets \([x_n]\) and \([y_n]\) and determine correlations (canonical correlations) between linear combinations \(a_1x_1 + \ldots + a_nx_n\) and \(b_1y_1 + \ldots + b_ny_n\) of each of the sets analyzed. To determine canonical correlations between remote sensing data and weather data for each NDVI image, we created a 11×2 template \(N\), composed of 11 lines of yearly values of two parameters-NDVIF \(_{\max}\) and \(\text{dNDVIF}/\text{dn}\)—for the period of 2006–2015 and for 2017 (the data for these years were available at [73]), and 11 × 2 templates \(T(k)\), composed of the average temperatures and amounts of precipitation of the kth month (April, May, June, and July) of the same years. The weather data were obtained at the Shira Weather Station (54°29’ N, 89°58’ E) — the station located nearest to the study areas. The calculations for each NDVI pixel provided four values of the canonical correlation coefficient for the four months of the growing season, which were then used to perform discriminant analysis.
Two procedures were proposed to test the approach based on using the variables NDVIF\((i)_{\text{max}}\) and dNDVIF/dn (Figure 3).

Procedure 1 included the following steps:

a. Choose the area with known types of plant communities;
b. Choose seasonal NDVI data for m seasons from earthdata.nasa.gov;
c. Determine parameters \(n_0, a, b, c, \text{NDVI}_{\text{max}}, \text{dNDVI}/dn\) on the selected areas for all years using the curve of NDVI seasonal dynamics;
d. Carry out discriminant analysis for different types of plant communities based on NDVI parameters;
e. Estimate the accuracy of the classification of plant communities based on NDVIF\((i)_{\text{max}}\) and dNDVIF/dn using the discriminant analysis classification template.

The accuracy of classification of forest, steppe, and meadow plant communities using Procedure 1 was between 68% and 94%.

Procedure 2 included the following steps:

a. Calculate the average temperatures of the air and amounts of precipitation for April, May, June, and July for the m analyzed seasons using the weather database;
b. Determine coefficients of canonical correlation between the weather data template over m seasons for each of the four months and the NDVIF\((i)_{\text{max}}\), dNDVIF/dn data template over the same years;
c. Carry out discriminant analysis based on coefficients of canonical correlation for the study areas;
d. Estimate the accuracy of the classification based on parameters of canonical analysis using the discriminant analysis classification template.

The accuracy of classification of forest, steppe, and meadow plant communities in Khakassia using Procedure 2 reached 100%.

3. Results

Thus, two ‘integrated’ parameters of phenological dynamics of plant communities—NDVIF\((i)_{\text{max}}\) and dNDVIF/dn—were proposed for classifying plant communities. These phenological indicators of vegetation dynamics of plant communities can be used to reduce the number of variables for analysis and classification. If classification is based on remote sensing data for m years, 2m integrated parameters instead of 46m variables will be used, i.e., the number of variables will be 23-fold lower.

For analysis of NDVI images of different plant communities, the values of the parameters were analyzed in the (NDVIF\(_{\text{max}}\), dNDVIF/dn) plane. The data for the steppe and forest communities in Khakassia in the seasons of 2003–2017 are shown in Figure 4.

The values of NDVIF\(_{\text{max}}\) and dNDVIF/dn for different communities varied widely in different seasons (the NDVIF\(_{\text{max}}\) for the steppe plant communities varied between 0.3 and 0.8 in different years and the NDVIF\(_{\text{max}}\) for the forest communities—between 0.6 and 0.9) (Figure 4).

To distinguish between the steppe and forest communities based on the values of NDVIF\(_{\text{max}}\) and dNDVIF/dn, we conducted discriminant analysis (Table 3).

The sufficiently low values of the \(\lambda_W\) and \(p\) statistics indicated the high level of discrimination between the forest communities and the steppe ones based on NDVI parameters. The error of classification of the steppe and forest communities was 4% (Table 3). Rather good discrimination based on the NDVIF\(_{\text{max}}\) and dNDVIF/dn was achieved between the steppe and meadow communities in Khakassia (Table 3). The parameters for the meadow and steppe study areas for 2003–2017 are given in the (NDVIF\(_{\text{max}}\), dNDVIF/dn) plane (Figure 5).
Figure 4. Characterization of the NDVIF seasonal curves in the (NDVIF$_{\text{max}}$, dNDVIF/dn) plane for the steppe (1) and forest (2) communities in Khakassia.

Table 3. Parameters of discriminant analysis for plant communities *

<table>
<thead>
<tr>
<th>Community</th>
<th>m*</th>
<th>k</th>
<th>$\lambda_W$</th>
<th>F</th>
<th>$p^{**}$</th>
<th>S, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forest–steppe</td>
<td>148</td>
<td>0.243</td>
<td>228.66</td>
<td>0.00001</td>
<td>96</td>
<td></td>
</tr>
<tr>
<td>Meadow–steppe</td>
<td>134</td>
<td>0.571</td>
<td>49.49</td>
<td>0.00001</td>
<td>81.48</td>
<td></td>
</tr>
<tr>
<td>Meadow–forest</td>
<td>148</td>
<td>0.825</td>
<td>15.61</td>
<td>0.001</td>
<td>68.7</td>
<td></td>
</tr>
<tr>
<td>Forest (Khakassia)–forest (Krasnoyarsk)</td>
<td>106</td>
<td>0.872</td>
<td>7.47</td>
<td>0.0009</td>
<td>71.42</td>
<td></td>
</tr>
<tr>
<td>Coniferous forest–mixed forest</td>
<td>46</td>
<td>0.209</td>
<td>79.70</td>
<td>0.00001</td>
<td>97.8</td>
<td></td>
</tr>
<tr>
<td>Steppe–forest (10 years)</td>
<td>100</td>
<td>0.253</td>
<td>143.11</td>
<td>0.00001</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td>Steppe–forest (5 years)</td>
<td>50</td>
<td>0.220</td>
<td>83.53</td>
<td>0.0001</td>
<td>98</td>
<td></td>
</tr>
<tr>
<td>Steppe–forest (canonical correlation)</td>
<td>11</td>
<td>0.13</td>
<td>9.61</td>
<td>0.0089</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Steppe–forest–meadow (canonical correlation)</td>
<td>14</td>
<td>0.102</td>
<td>4.26</td>
<td>0.0066</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

* m—the number of the seasons analyzed, k—the number of NDVI images; $\lambda_W$—Wilk's Lambda statistic; F—Fisher's F test; $p$—significance level; S—percentage of the accurately determined seasonal remote sensing data. ** The critical value was chosen at $p = 0.05$.

Figure 5. NDVIF$_{\text{max}}$ and dNDVIF/dn for the meadow (1) and steppe (2) communities in Khakassia.

Discriminant analysis was carried out to distinguish between the steppe and meadow communities, too (Table 3). The values of the $\lambda_W$ statistic were greater than the corresponding values obtained by the
discriminant analysis of the forest and steppe communities in Khakassia, suggesting the poorer quality of classification of the meadow and steppe plant communities based on NDVI parameters.

The average percentage of the steppe and meadow plant communities accurately classified using remote sensing data was 81.48%. The classification of the meadow and forest communities based on long-term remote sensing measurements was somewhat less accurate (Table 3).

How close are the NDVIF$_{\text{max}}$ and dNDVIF/dn for the plant communities of the same type but with different geographical positions? To answer this question, we compared clusters in the (NDVIF$_{\text{max}}$, dNDVIF/dn) plane (Figure 6) for mixed forests with different species compositions (Table 1) located in Khakassia (Study areas no. 3 and 4) and at Krasnoyarsk (Study area no. 1).

Figure 6. NDVIF$_{\text{max}}$ and dNDVIF/dn for forest communities in the Krasnoyarsk Territory (1) and Khakassia (2) for 2003–2017.

Discriminant analysis produced accurate classification of 71% of the data for these study areas (Table 3). At the same time, only 20% of the forest communities in the Krasnoyarsk Territory were classified accurately.

Finally, analysis based on NDVIF$_{\text{max}}$ and dNDVIF/dn (Figure 7) discriminated between the coniferous forest (Study area no. 5) and the mixed forest (Study area no. 1).

Figure 7. NDVIF$_{\text{max}}$ and dNDVIF/dn for 2003–2017 for (1) coniferous forest; (2) mixed forest.
4. Discussion

The plant communities inhabiting the study areas differ in their phenological dynamics during the growing season. Each of the plant species that constitute plant communities show different sensitivity and time of response to changes in the temperature of the air, amount of precipitation, and other weather phenomena. The species composition of each plant community determines the specific features of its phenological dynamics in each growing season, which differ from the seasonal phenological dynamics of other plant communities. However, the ranges of values of these variables for different plant communities had different positions on the (NDVIF$_{max}$, dNDVIF/dn) plane. Is the classification approach proposed in the present study competitive with other well-known approaches? A study by Miklashevich and Bartalev [73] suggested using the start of the growing season, SOS, as a phenological parameter for classifying vegetation. In research based on remote sensing data, the start of the growing season is often defined as the first day of the year corresponding to the start of the ascending trend in the time series of the normalized difference vegetation index (NDVI) values [74]. The SOS values were used during one year to classify vegetation types such as forests, open stands, tundra, herbaceous vegetation of wetlands, and vegetation-free areas in West Siberia, Russia [73]. That study showed the following SOS values: between 150 and 216 days per year for tundra, between 135 and 151 days per year for forest tundra, and between 105 and 136 days per year for taiga.

To compare that classification approach with the methods used in the present study, several plots with the known SOS data for the period between 2003 and 2017 were chosen for each of the vegetation types considered here (meadow, steppe, coniferous-deciduous forest (Betula pendula Roth, Pinus sylvestris L.), and coniferous forest (Abies sibirica Ledeb., Pinus sylvestris L.). SOS distribution functions for every year of the study period were constructed for each vegetation type (Figure 8).

![Figure 8](image_url)

**Figure 8.** SOS distribution functions for the period between 2003 and 2017. (1) meadow, (2) steppe, (3) deciduous forest (Betula pendula Roth, Pinus sylvestris L.), (4) coniferous forest (Abies sibirica Ledeb., Pinus sylvestris L.).

SOS-based classification could be effective if intersection regions of SOS distribution functions for different vegetation types were minimal. However, there are many intersections of SOS distribution functions (Figure 8), the modes of SOS distribution functions and ranges of SOS values almost coincide, and, thus, it is impossible to distinguish between the vegetation types using long-term SOS data. A possible reason for the disagreement between our results and the data presented in the study [73] is that the authors of that study considered vegetation types that differed considerably in their species composition, such as forest and tundra. For the vegetation types that grow on a rather limited area, which are examined in our study, the differences in the phenology of plant communities are far less
significant. Therefore, the approach proposed in the present study is effective for distinguishing between the adjacent plant communities growing in the same territory.

Thus, the long-term integrated parameters of $\text{NDVIF}_{\text{max}}$ and $\text{dNDVIF}/\text{dn}$ seasonal dynamics used to classify the types of plant communities can effectively (with a probability no less than 0.95) distinguish between forest and steppe and between coniferous forest and mixed forest. The proposed parameters were used somewhat less effectively (with a probability of about 0.69) to distinguish between meadow and forest vegetation types.

In the present study, a number of calculations were undertaken to answer the question: Will the quality of classification change if, instead of the whole dataset, the data for, say, 10 years are used? Discriminant analysis was carried out based on the 10-year data (2008–2017) for the steppe and forest communities in Khakassia ($\lambda_W = 0.25$ and $S = 95\%$) (Table 3). The steppe and forest communities in Khakassia can be classified using the 5-year NDVI data (2013–2017) ($\lambda_W = 0.25$ and $S = 98\%$) (Table 3). Thus, if the depth of the analysis based on the proposed phenological markers is reduced from 15 to 5 years, the quality of classification is not affected.

Different types of plant communities show dissimilar responses to weather conditions, and, thus, $\text{NDVIF}_{\text{max}}$ and $\text{dNDVIF}/\text{dn}$ response to weather changes can be used to improve classification of plant communities. Relationships between the variables characterizing NDVI and weather variables (average temperatures and amounts of precipitation in April, May, June, and July) were found using the method of canonical correlations. Plants are obviously unaware of calendar dates, and dissimilar conditions in the same months of different years will produce different effects on plant phenology, but one can try to reveal relationships between the templates of NDVI parameters for m years and the templates of weather parameters for these years.

The canonical correlation coefficient will characterize the level of the relationship between weather parameters of the month and NDVI parameters for each type of the plant communities studied here. To calculate canonical correlations between weather parameters and $\text{NDVIF}_{\text{max}}$ and $\text{dNDVIF}/\text{dn}$ for the steppe and forest plant communities, we used the data from the Shira Weather Station, which is situated close to the plant communities in Khakassia (Study areas no. 2–4).

The coefficients of canonical correlation for April, May, June, and July between weather parameters (the average temperatures of the air and amounts of precipitation) and integrated parameters $\text{NDVIF}_{\text{max}}$ and $\text{dNDVIF}/\text{dn}$ for the forest and steppe plant communities are listed in Table 4.

<table>
<thead>
<tr>
<th>Plant Communities</th>
<th>Stat. Parameters</th>
<th>April</th>
<th>May</th>
<th>June</th>
<th>July</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meadow</td>
<td>average</td>
<td>0.800</td>
<td>0.563</td>
<td>0.310</td>
<td>0.777</td>
</tr>
<tr>
<td></td>
<td>st. dev.</td>
<td>0.114</td>
<td>0.188</td>
<td>0.056</td>
<td>0.140</td>
</tr>
<tr>
<td>Forest</td>
<td>average</td>
<td>0.616</td>
<td>0.616</td>
<td>0.488</td>
<td>0.702</td>
</tr>
<tr>
<td></td>
<td>st. dev.</td>
<td>0.153</td>
<td>0.166</td>
<td>0.192</td>
<td>0.121</td>
</tr>
<tr>
<td>Steppe</td>
<td>average</td>
<td>0.493</td>
<td>0.588</td>
<td>0.732</td>
<td>0.758</td>
</tr>
<tr>
<td></td>
<td>st. dev.</td>
<td>0.082</td>
<td>0.099</td>
<td>0.071</td>
<td>0.042</td>
</tr>
</tbody>
</table>

Long-term annual average coefficients of canonical correlation between the proposed markers of phenological dynamics, $\text{NDVIF}_{\text{max}}$ and $\text{dNDVIF}/\text{dn}$, and weather parameters in April were significantly higher for the forest and meadow communities than for the steppe communities (Table 4). By contrast, in June, long-term annual average coefficients of canonical correlation between the proposed NDVI markers of phenological dynamics and weather parameters were significantly lower for the forest and meadow communities than for the steppe communities. These relationships can help discriminate further between plant communities by using both integrated NDVI variables and the coefficients of canonical correlation between these variables and weather parameters in different months.
Canonical correlations between the integrated NDVI data and weather parameters over the growing season can be a tool for perfect discrimination between the steppe, forest, and meadow plant communities ($S = 100$).

5. Conclusions

Using remote sensing data to describe vegetation is an attractive idea, as data on the state of vegetation all over the world can be collected rapidly and inexpensively. NDVI parameters, however, vary rather widely both in one season and between seasons, as weather factors considerably influence the current values obtained by remote sensing techniques. The same plant community can have a different NDVI depending on weather conditions, and this complicates classification of plant communities.

New integrated variables, NDVIF$_{max}$ and dNDVIF/dn, determined from the NDVI time series, were proposed in the present study in order to achieve effective classification of plant communities using remote sensing data. These markers of phenological dynamics of plant communities are regarded as the key parameters of a plant community. These parameters were tested as the basis for classifying plant communities. The data obtained for the study areas with plant communities were considered as 'clouds', and the 'clouds' for different communities were separated using linear discriminant analysis. The quality of classification (the average percentage of the yearly data for study areas accurately classified by the plant composition) achieved using this approach varied between 94% (in discriminating between forest and steppe communities) and 68% (in discriminating between meadow and forest communities). Thus, using the parameters describing NDVI seasonal dynamics proposed in this study, one can reduce the time series of NDVI seasonal dynamics to two parameters with limited variance.

A definite advantage of the approach to classification of plant communities proposed in the present study is the transition from using only remote sensing data to using the method including both remote sensing data and weather parameters. Weather is clearly a key phenological factor, and NDVI of different vegetation types should be sensitive to weather changes. The use of the canonical correlations to estimate differences between vegetation types such as steppe, forest, and meadow enabled classification of these vegetation types with a probability of 1. Integration of remote sensing and land data may be effective for classifying vegetation types.

To reduce the NDVI measurement data more substantially, this study proposed using the method of canonical correlations between weather parameters and the integrated markers of phenological dynamics of NDVI. As a result of using this approach, each study area was characterized by only four values of canonical correlation coefficients over the entire period analyzed. Discriminant analysis of the parameters of canonical correlation demonstrated that these parameters could be used to classify plant communities.

Author Contributions: Y.I. conceived and designed the study. A.K. performed the calculations. O.Y. participated in data processing. V.S. performed the data analysis. All the authors analyzed the results and wrote the paper.

Funding: This research was funded by RFBR and Russian Geographical Society according to the research project number 17-05-41012.

Acknowledgments: The authors are grateful to Elena Krasova for translating the article.

Conflicts of Interest: The authors declare no conflict of interest.

References


58. Han, M.; Cheng, L.; Meng, H. Application of four-layer neural network on information extraction. *Neural Netw.* 2003, 16, 547–553. [CrossRef]
65. *Academic and Practical Guide on Climate of the USSR (Nauchno-priladnoy spravochnik po klimatu SSSR);* Series 3. Long-Term Data. Parts 1–6, No. 21, Book 1; Gidrometeoizdat: The Krasnoyarsk Territory, the Tuvan ASSR, St. Petersburg, 1990; Volume 6, 625p. (In Russian)

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).