**КОМПЬЮТЕРНЫЕ ИССЛЕДОВАНИЯ
И МОДЕЛИРОВАНИЕ 2022 Т. 14 № 6 С. 1357–1370**
DOI: 10.20537/2076-7633-2022-14-6-1357-1370

Ки&М

АНАЛИЗ И МОДЕЛИРОВАНИЕ СЛОЖНЫХ ЖИВЫХ СИСТЕМ

УДК: 528.854

Мониторинг распространения борщевика Сосновского с использованием алгоритма машинного обучения «случайный лес» в Google Earth Engine

Т. Т. Уифтер[a], Ю. Н. Разумный[b], А. В. Орловский[c], В. К. Лобанов[d]

Департамент механики и процессов управления, Инженерная академия,
Российский университет дружбы народов,
Россия, 117198, г. Москва, ул. Миклухо-Маклая, д. 6

E-mail: [a] teklay-ty@rudn.ru, [b] razoumny-yun@rudn.ru, [c] orlovskiy-av@rudn.ru, [d] lobanov-vk@rudn.ru

Получено 01.07.2022, после доработки — 10.11.2022.
Принято к публикации 16.11.2022.

Изучение спектрального отклика растений на основе данных, собранных с помощью дистанционного зондирования, имеет большой потенциал для решения реальных проблем в различных областях исследований. В этом исследовании мы использовали спектральные свойства для идентификации инвазивного растения — борщевика Сосновского — по спутниковым снимкам. Борщевик Сосновского — инвазивное растение, которое наносит много вреда людям, животным и экосистеме в целом. Мы использовали выборочные данные о геолокации мест произрастания борщевика в Московской области, собранные с 2018 по 2020 год, и спутниковые снимки Sentinel-2 для спектрального анализа с целью его обнаружения на снимках. Мы развернули модель машинного обучения Random Forest (RF) на облачной платформе Google Earth Engine (GEE). Алгоритм обучается на наборе данных, состоящем из 12 каналов спутниковых снимков Sentinel-2, цифровой модели рельефа и некоторых спектральных индексов, которые используются в алгоритме в качестве параметров. Используемый подход заключается в выявлении биофизических параметров борщевика Сосновского по его коэффициентам отражения с уточнением радиочастотной модели непосредственно по набору данных. Наши результаты наглядно демонстрируют насколько сочетание методов дистанционного зондирования и машинного обучения может помочь в обнаружении борщевика и контроле его инвазивного распространения. Наш подход обеспечивает высокую точность обнаружения очагов произрастания борщевика Сосновского, составляющую 96,93 %.

Ключевые слова: борщевик Сосновского, инвазивные растения, Google Earth Engine, машинное обучение, случайный лес

COMPUTER RESEARCH AND MODELING
2022 VOL. 14 NO. 6 P. 1357–1370
DOI: 10.20537/2076-7633-2022-14-6-1357-1370

ANALYSIS AND MODELING OF COMPLEX LIVING SYSTEMS

UDC: 528.854

Monitoring the spread of Sosnowskyi's hogweed using a random forest machine learning algorithm in Google Earth Engine

T. T. Yifter[a], Y. N. Razoumny[b], A. V. Orlovsky[c], V. K. Lobanov[d]

Department of Mechanics and Control Processes, Academy of Engineering, Peoples' Friendship University of Russia (RUDN University),
6 Miklukho-Maklaya st., Moscow, 117198, Russia

E-mail: [a] teklay-ty@rudn.ru, [b] razoumny-yun@rudn.ru, [c] orlovskiy-av@rudn.ru, [d] lobanov-vk@rudn.ru

Received 01.07.2022, after completion — 10.11.2022.
Accepted for publication 16.11.2022.

Examining the spectral response of plants from data collected using remote sensing has a lot of potential for solving real-world problems in different fields of research. In this study, we have used the spectral property to identify the invasive plant *Heracleum sosnowskyi* Manden from satellite imagery. *H. sosnowskyi* is an invasive plant that causes many harms to humans, animals and the ecosystem at large. We have used data collected from the years 2018 to 2020 containing sample geolocation data from the Moscow Region where this plant exists and we have used Sentinel-2 imagery for the spectral analysis towards the aim of detecting it from the satellite imagery. We deployed a Random Forest (RF) machine learning model within the framework of Google Earth Engine (GEE). The algorithm learns from the collected data, which is made up of 12 bands of Sentinel-2, and also includes the digital elevation together with some spectral indices, which are used as features in the algorithm. The approach used is to learn the biophysical parameters of *H. sosnowskyi* from its reflectances by fitting the RF model directly from the data. Our results demonstrate how the combination of remote sensing and machine learning can assist in locating *H. sosnowskyi*, which aids in controlling its invasive expansion. Our approach provides a high detection accuracy of the plant, which is 96.93 %.

Keywords: Heracleum Sosnowski Manden, invasive plants, Google Earth Engine, machine learning, random forest

Citation: *Computer Research and Modeling*, 2022, vol. 14, no. 6, pp. 1357–1370.

Introduction

Plant invasions, if not monitored and contained, can pose a threat to humans, animals and the environment and ecosystem as a whole. Non-native plant invasions are among of the most disruptive forms of ecological change. *Heracleum sosnowskyi* Manden is an invasive plant that was discovered in 1772, and described in 1944 by I. P. Mandenova as a separate species. Its name is derived from the surname of Prof. D. I. Sosnowskyi, a botanist who studied Caucasian flora [Jakubska-Busse, Śliwiński, Kobylka, 2013]. The plant was introduced as a fodder plant to many countries during the Soviet Union. But later, the plant was found to be invasive and spreads in a large area and adapts itself even to cold weather to a certain degree [Dalke et al., 2020]. It keeps causing a lot of harm, and detecting the plant from satellite imagery will enable a mitigation of these negative impacts by focusing on those areas and applying different methods of eradication. It was promoted as a crop for northwest Russia, where it was first introduced in 1947. From the 1940s onwards, it was introduced as a fodder plant to Latvia, Estonia, Lithuania, Belarus, Ukraine and the former German Democratic Republic [Nielsen et al., 2005; Kabuce et al., 2010]. In Europe, there are two other species of Heracleum, Heracleum sphondylium, and Heracleum mantegazzianum. All of them contain furanocoumarins with photoallergic properties in the aerial parts. *H. sosnowskyi* is among the most dangerous invasive plant species in Eastern Europe [Panasenko, 2017]. *H. sosnowskyi* originates in the central and eastern Caucasus and western, central, eastern and southwestern Transcaucasia and in northeastern Turkey [Kabuce et al., 2010; Jahodová et al., 2007].

In parts of Europe, *H. sosnowskyi* has been grown as a crop for silage production. After this practice had been given up, the plant established itself in very extensive and dense stands taking advantage of widespread abandonment of arable land and a significant reduction in the abundance of grazing animals. In Russia the first herbarium sample was collected in 1948 from the Serpukhov region of the Moscow district. Before the 1970s *H. sosnowskyi* was rare, but since 1970s it has been recorded in many sites. Apart from crop reason (livestock fodder) it was cultivated in many botanical gardens and sometimes as ornamental plant in gardens [Kabuce et al., 2010].

Various papers studied the characteristics, impact and spread of *H. sosnowskyi* in Europe. For example, the paper [Jahodová et al., 2007] gives a historical perspective of the introduction of the species of the genus Heracleum (Umbelliferae). It discusses the genetic pattern relatedness between and within species. It concludes that multiple introductions of the different species of the plants are likely to have occurred from south-west Asia to Europe. The paper [Ozerova, Krivosheina, 2018] also discusses the introduction of *H. sosnowskyi* for the accumulation of seeds and cultivation of this plant in the 1940s–1950s, which occurred in five main centers: Murmansk, Moscow, and Leningrad oblasts, the Komi Republic, and Kabardino-Balkaria. Another paper [Baležentiene, Stankeviciene, Snieškiene, 2013] compared the seed ecology of *H. sosnowskyi* and Heracleum mantegazzianum, and the authors found out that *H. sosnowskyi* expansion is not dependent on weather, but rather on human activities, namely, the massive planting in eastern Europe, than on ecological constraints. About the relationship between these two species, the paper [Jakubska-Busse, Śliwiński, Kobylka, 2013] compares them in terms of their bioactive components of essential oils, from the seed samples collected, and finds no significant differences in their chemical compositions, suggesting that the species can be closely related. Other papers in the literature studied the impact of *H. sosnowskyi* on the ecology. This paper explores *H. sosnowskyi*'s ability to become monostand on the area that it invaded, and uses this observation to investigate traits (demography, canopy structure, morphology and physiology) of the plant in relation to other surrounding plants. For example, the authors found, that *H. sosnowskyi*'s canopy captured approximately 97 % of the light and also photosynthetic water use efficiency (6–7μMCO2/mM H2O), which is several orders of magnitude higher than average taiga zone grasses [Dalke et al., 2015]. *H. sosnowskyi* being a dominant invasive plant, the authors found out that the active colonization of meadow phytocenoses by *H. sosnowskyi* in the territory of the Moscow agglomeration leads to a decrease in the biodiversity of plant species and microorganisms and impoverishment and disturbance of the developed biotic cycle. And they suggest monitoring of microbial communities and their changes and disturbances to correct and prevent adverse irreversible consequences [Glushakova, Kachalkin, Chernov, 2015]. The effect of *H. sosnowskyi* on the soil was also studied in the paper [Renčo et al., 2019]. The authors analyzed the effect of *H. sosnowskyi* on soil nematode communities by investigating areas with and without the invasion. And they concluded that *H. sosnowskyi* eventually became dominant impacted soil nematode communities but did not affect nematode diversity.

In relation to dataset usage for the analysis of *H. sosnowskyi* in large geographical areas, the work [Ozerova et al., 2017] uses the geodatabase of the general characteristics of the geographical location of *H. sosnowskyi* with expeditionary investigations from 2008 to 2016. The locations where the data collected are mainly valleys of large and medium rivers, as the presence of the plant is high around this high moisture areas. Based on these field observations, they explained the conducive environment created for the invasion of the plant around these rivers. Another paper [Chadin et al., 2017] models the presence of *H. sosnowskyi* in the Komi Republic (north-eastern part of Russia) from data collected using the project called "citizen science project". The paper [Panasenko, 2017] assesses the presence of *H. sosnowskyi* in the Bryansk oblast. And their conclusion about the invasion is that the seed productivity and the formation of a large seed pool, together with the higher rate of spring development, in comparison to the indigenous plants of the examined region, are to blame for the rapid expansion. The work [Afonin et al., 2017] also utilizes the dataset to create a map of the predicted locations of *H. sosnowskyi*. They collected data from the European Russia in the years 2013–2014 and calculated the number of plants from a vehicle, which moved according to a preselected route along the moisture gradient. As a result of this analysis, they come up with a map of the European Russia and adjacent countries showing the expansion of the plant.

Some papers also discussed the efforts made to eradicate or control the expansion of *H. sosnowskyi*. The paper [Dalke, Chadin, Zakhozhiy, 2018] evaluates the government contracts made to mitigate the negative impact of this plant in different regions of Russia. The authors suggest that the control should be started only in one or two settlements as a pilot plan, and only then should the experience learned be applied to a larger region.

Identifying *H. sosnowskyi* from satellite imagery with high accuracy will facilitate the processes of eradicating the plant and mitigating its impact. Unlike pollution, the proliferation of alien species may not stop even after the source is controlled; unlike the effects of wildfire or logging, invaded vegetation may remain dominated by a single species for a long period of time [Huang, Asner, 2009]. This is because it is often not feasible (or possible) to completely remove the invading plant. In addition, it also should be investigated whether the removal of the invasive plant species harmed some native organisms that have already adapted to them [Grzedzicka, 2022]. This indicates that there is a need for a mechanism that enables continuous monitoring of the alien species. Remote sensing together with the latest advancements in cloud computing and machine learning can enable automatic detection and prediction of future expansion of the plant [Vicente et al., 2016]. Knowing the specific plant characteristics and how it behaves in its surroundings is a crucial factor in detecting it. Field observations and traditional survey methods, such as ground mapping using global positioning systems (GPS), have high accuracy and are species-specific methods for small management areas, but might be technically, logistically, and financially unfeasible for larger areas or multi-species invasions. Therefore, remote sensing has become an increasingly popular approach for identifying and monitoring invasive plant species [Paz-Kagan et al., 2019].

Remote sensing, which aims to extract information about the condition and/or state of an object without requiring physical contact, has become increasingly important for environmental conservation and ecological monitoring, including the management of invasive species [Vaz et al., 2018]. During the 1970s and 1980s, remote sensing reached important milestones with the launching of the first multi- and hyperspectral satellites [Kwok, 2018]. This allowed the acquirement of data applicable in conservation and ecology studies with global coverage and high spatial, temporal and spectral resolutions. However, the first remote sensing studies were mostly applied to agriculture monitoring and forest mapping, without focus on invasions. It has been used for decades to measure and map the biophysical characteristics of vegetation. Remote sensing has been an important tool for large-scale ecological studies in the past three decades, but it was not commonly used to study alien invasive plants until the mid1990s [Vaz et al., 2018]. Since then, it has been proposed and used as a tool for understanding of the drivers, processes and effects of plant invasion. And remote sensing data and tools are more accessible and abundant than ever, giving rise to “ecology’s remote sensing revolution” [Kwok, 2018]. Significant advances in RS spatial and spectral resolution, and in spatial and temporal coverage, have made it to be used in the identification of plant invasions and invaded ecosystems, map invasive plants in diverse eco-systems, prediction of the potential distribution of invasive species, in comprehending landscape invasion ability and associated ecological impacts and for assessing impacts on the ecosystem functional attributes, properties and services [Vaz et al., 2019]. Resulting distribution maps can be used to target management of early infestations and to model future invasion risk. Remote identification of invasive plants based on differences in spectral signatures is the most common approach, typically using hyperspectral data. The methodology used to detect different types of plants is rooted in the target plant’s unique characteristic [Xie, Sha, Yu, 2008].

The advancement of computation infrastructure and resources has also benefited RS in its capabilities for solving many real-world problems related to a variety of fields including agriculture, environment, and ecology. The cloud-based platform Google Earth Engine (GEE) is one of such facilities, which gives researchers the required tools for solving diverse problems. GEE offers a new choice for researchers interested in geospatial big remotely sensed data analysis by providing a multi-petabyte catalog of remote sensing data and machine learning algorithms. GEE provides free satellite imagery with a Git repository for storing, sharing and geo-big data processing facilities [Gorelick et al., 2017].

The primary objective of this study is to assess technological solutions for identification of *H. sosnowskyi* from satellite imagery. We performed a classification using a pixel-based supervised random forest machine learning algorithm using the Google Earth Engine (GEE) cloud computing platform. By performing this classification, we tried to answer the scientific research question: How accurately can *H. sosnowskyi* be classified using a random forest machine learning algorithm on Sentinel-2 multispectral images? This technological capability for identifying the location of the invasion will help to apply integrated methods of eradicating that include mowing/cutting, chemical control, soil cultivation and sowing of grass mixtures [Nielsen et al., 2005; Kabuce et al., 2010].

Materials and Methods

Study Area

The study area is the Moscow region with a total area of 47 thousand square kilometers. The climate of the Moscow region is moderately continental, which leads to cold winters and warm summers with moderate precipitation. This is a favorable climate for the growth of *H. sosnowskyi*.

Data

Based on its high temporal resolution together with good spatial and spectral resolution, Sentinel-2 was considered to be suitable for the study. We have used Cloud-free Sentinel-2 Level-2A surface reflectance data in the GEE (ImageCollectionID: COPERNICUS/S2_SR). Samples of the locations, where *H. sosnowskyi* exits, were collected from GEE. The ground truth data (the geolocation of *H. sosnowskyi*) is based on the field work done by the project “Stop Hogweed” [Prishchepov, 2022]

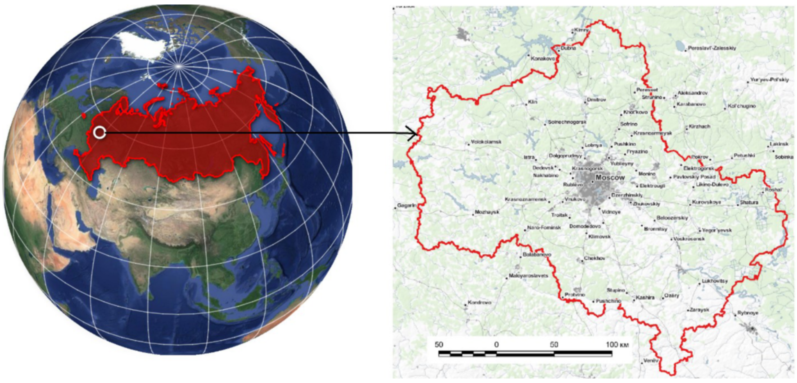

Figure 1. Location of the region of interest (ROI), Moscow region, Russia

and these locations are represented as a polygon shape files. Using these polygons of the locations of *H. sosnowskyi*, we used data from June 2018 to July 2020 for optimal temporal window selection analysis (Table 1). The wide-swath multispectral imager of the Sentinel-2A/2B satellites has 13 spectral bands with four bands at 10 m, six bands at 20 m, and three bands at 60 m spatial resolution. This study used blue, green, red, near-infrared (NIR), and short-wave infrared (SWIR) bands. The spatial resolution of blue, green, red, and NIR bands is 10 m. The SWIR band was resampled to 10-m pixel size.

Table 1. Sentinel-2 Bands and their Parameters

Sentinel-2 Band	Center Wavelength (nm)	Spectral Width (nm)	Spatial Resolution (m)
Band 1-Coastal aerosol	443	20	60
Band 2-Blue	490	65	10
Band 3-Green	560	35	10
Band 4-Red	665	30	10
Band 5-Vegetation	705	15	20
Band 6-Vegetation	740	15	20
Band 7-Vegetation	783	20	20
Band 8-NIR	842	115	10
Band 8a-Vegetation	865	20	20
Band 9 – Water vapour	945	20	60
Band 11-SWIR1	1610	90	20
Band 12-SWIR2	2190	180	20

Sample Data

We used Google Earth satellite images from 2018–2020 (Table 2) to create the ground control points as sample training and validation data.

Table 2. Information of the Sentinel-2 imagery used in this study

Sensor	Date
Sentinel2 MSI	1/6/2018 – 30/7/2018
Sentinel2 MSI	1/6/2019 – 30/7/2019
Sentinel2 MSI	1/6/2020 – 30/7/2020

The accuracy assessment is done based on these collected data. Sample pixels of *H. sosnowskyi* were picked based on the geolocation data collected from the field (Figure 2). These field data collections are performed by a team organized by the project “Stop Hogweed” [Prishchepov, 2022].

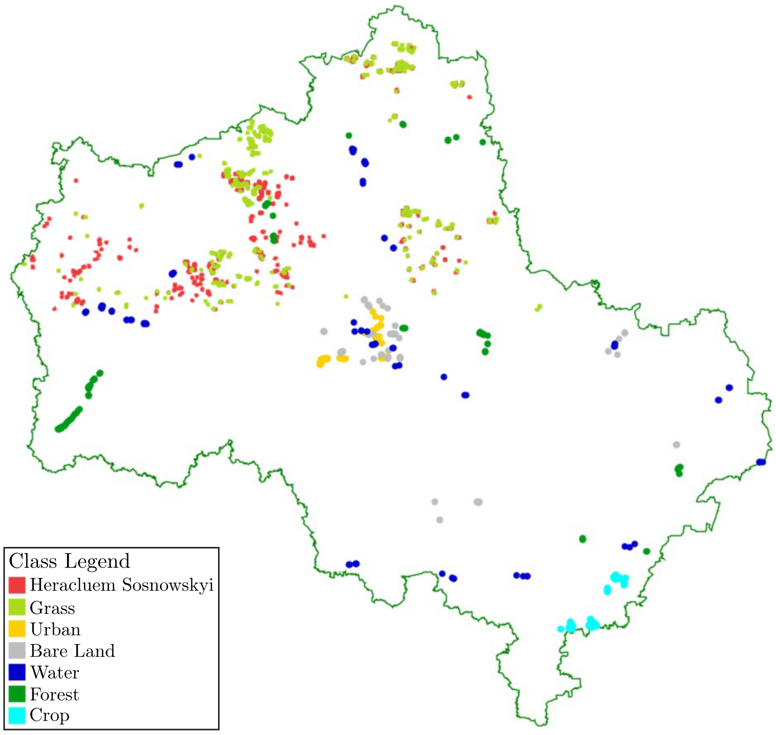

Figure 2. The study area Moscow region with sample plots

Digital Elevation Model

We have used the 30-m SRTM digital elevation model (DEM) data and its derived variable slope to be added as calculated values in the sample collection so as to give more features to the classification algorithm. These data are necessary as there could be a relationship between elevation level and where *H. sosnowskyi* thrives. We have utilized DEM as one of the features that the Random Forest algorithm uses to perform the classification.

Method

The RF classifier, developed by Breiman [Breiman, 2001], is an ensemble classifier that produces multiple decision trees, using a randomly selected subset of training samples features. It has been widely applied in different remote sensing research fields, in particular, to map Land Cover classes, urban buildings, mapping of invasive plants [Belgiu, Drăgu, 2016]. It has also been successfully used for classification in ecology [Cutler et al., 2007]. In this paper, we have used the pixel-based supervised RF algorithm (Figure 3) because it is less susceptible to data noise and overfitting and is extremely useful in classifying remote sensing data.

Random Forest classifier has been found to successfully handle high data dimensionality and multi-collinearity, being both fast and insensitive to overfitting [Belgiu, Drăgu, 2016]. We used spectral features as classification inputs. The spectral features include the red, green, blue, NIR, and SWIR bands, normalized difference vegetation index (ndvi), bare soil index (bsi), modified normalized difference water index (mndwi), normalized difference built-up index (ndbi), and an index

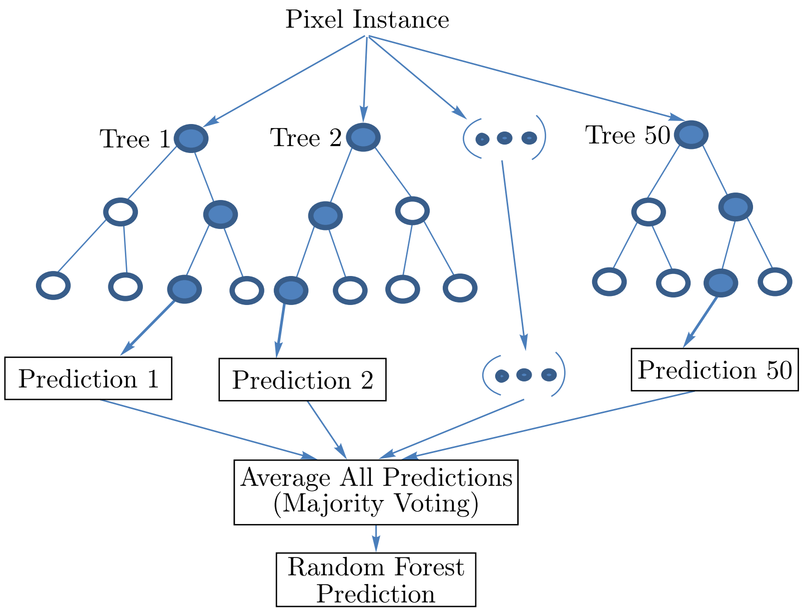

Figure 3. Random Forest Algorithm

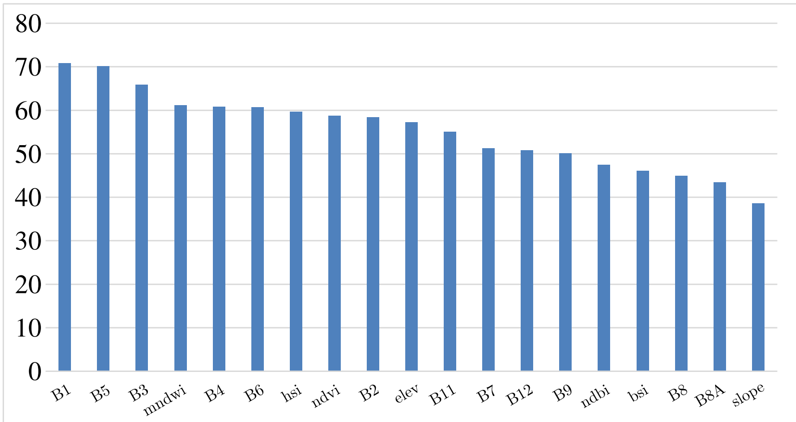

Figure 4. Importance of the features in Random Forest Classification

that we introduced and called *H. sosnowskyi* index (hsi) and it is computed as: hsi = (Green – – Blue) ∗ NIR. RF classifiers also provide a quantitative measurement of each variable's contribution to the classification output, which is useful in evaluating the importance of each variable (Figure 4).

RF classifiers available in GEE use six input parameters: (1) number of classification trees, (2) number of variables used in each classification tree, (3) minimum leaf population, (4) bagged fraction of the input variables per decision tree, (5) out-of-bag mode, and (6) random seed variable for decision tree construction. RF is undoubtedly the most widely applied ML algorithm in the GEE environment [Amani et al., 2020]. It has several advantages over other classification tree-based approaches. Pruning of trees is not necessary and the approach is robust to overfitting, a problem that plagues classification trees. It is easier to use than many other ensemble classification methods, with the only parameters to be set being the number of trees grown and the number of variables used at each tree split; however, it has been shown to be not very sensitive to the setting of either of these parameters [Breiman, 2001]. It is also claimed that RF can provide a reliable estimate of error using the data that are randomly withheld from each iteration of tree development (the “out-of-bag” portion), making it unnecessary to have an independent accuracy assessment data set [Breiman, 2001].

Heracleum sosnowskyi Detection

The proposed hierarchical classification framework mainly comprises two steps: (1) calculate the potential features that enable the detection *H. sosnowskyi*; (2) apply the RF classifier to detect *H. sosnowskyi*.

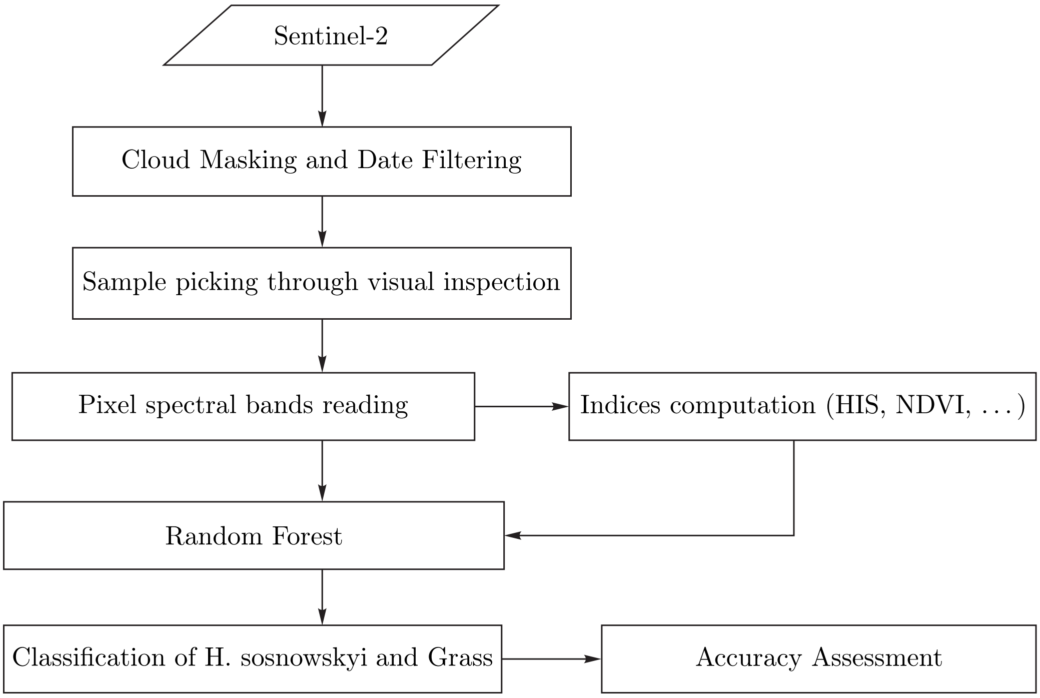

Figure 5. Algorithm flowchart

Results

Separability Analysis and Optimal Time Window Selection

In individual *H. sosnowskyi* plants, under the same conditions, changes in phenological phases can occur in different time periods. For example, for *H. sosnowskyi* growing in the Moscow region, the period of appearance of the first shoots begins in May, the seed maturation period lasts approximately from mid-June to August, the period of fruiting from early August to mid-September. We find that the best time for getting the highest classification accuracy for *H. sosnowskyi* is in the period between June and July. Because of this, we select all our training data from this period (Table 1) and we applied the RF algorithm on imagery for the same period. The result of the classification is displayed in Figure 6, which shows that 2.44 % of the region was covered by *H. sosnowskyi* (for the years 2018–2020 in which the data was collected).

Classification Accuracy Assessment

Classification accuracy is an important metric that enables one to measure classification correctness. The accuracy evaluation is usually performed using confusion matrix (Table 3).

In this study, the parameters for the RF classifier are selected for optimal results. The number of trees was 50, we run a hyperparameter tuning of RF for the optimal number of trees (Figure 7),

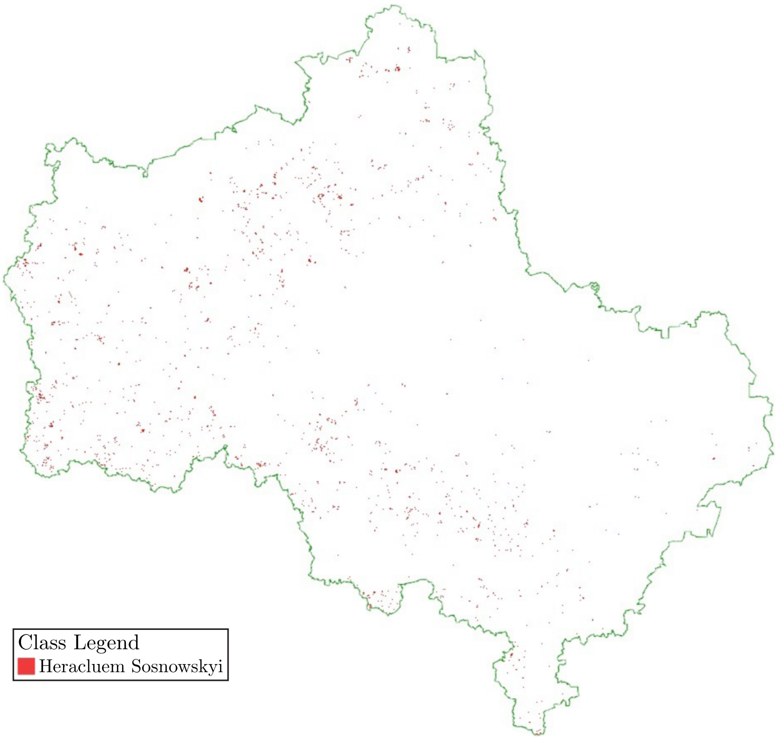

Figure 6. Land Cover Classification of *H. sosnowskyi* in Moscow region

Table 3. Confusion Matrix

	H. sosnowskyi	Others
H. sosnowskyi	7524	391
Others	347	15 844

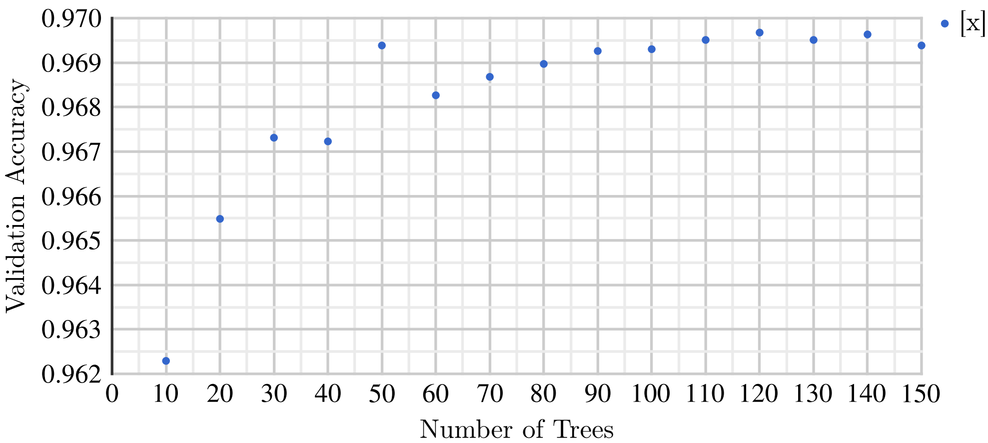

Figure 7. Optimal Number of Trees (50) with the highest Validation Accuracy for the RandomForest algorithm used in the study

and the number of variables per split was the square root of the number of variables. The sample data were randomly divided into training (70 %) and validation (30 %). We quantified classification accuracy using overall accuracy, consumer accuracy and the kappa coefficient.

With regard to feature selection, we used 18 features, namely, the bands B1–B8, B8A, B9–B12, bsi, elevation (elev), mndwi, ndbi and ndvi (Figure 4) for each sample that we collected in preparation for the RF classification.

As a result of the conducted tests, we get the overall accuracy of 96.93 % and the Kappa coefficient of 93.04 %.

Importance of features

We calculated the variable importance scores for all features with an RF classifier. Implementing an RF classifier with one mosaic image formed by cloud masking Sentinel-2 images from the months of June and July for the years 2018 to 2020. There is a little variation of this importance of the features for different years, but the overall trend looks similar. Although we used 19 features as inputs, the features with high importance scores changed when using imagery of different periods, especially when multi-period imagery was used. The variable importance score also decreased when the number of features increased, which was due to how the importance was measured.

The feature with the top importance in all the test runs that we made is the NIR band followed by the red band. This is expected as the NIR band is strongly reflected back from *H. sosnowskyi* and satellite sensors measure wavelengths of light absorbed and reflected by green plants. The leaves themselves strongly reflect wavelengths of near-infrared light (NIR) and certain pigments in plant leaves strongly absorb wavelengths of visible (red) light. One instance of the importance of features from our test runs for the year 2020 and the months June to July are displayed in Figure 4.

Discussion

Potential of Sentinel-2 Imagery to Monitor Heracleum Sosnowskyi

The potential of Sentinel-2 Imagery to Monitor *H. sosnowskyi* has been explored in this study. There is a similar paper for monitoring *H. sosnowskyi* using remote sensing [Visockiene, Tumeliene, Maliene, 2020]. It utilizes data collected from satellite imagery and applied supervised and unsupervised machine learning algorithms for detecting *H. sosnowskyi* from the imagery for Vilnius District Municipality, Lithuania. The authors used four spectral bands from Sentinel-2 with a spatial resolution of 10 m and 20 m for the classification, namely: blue, green, red, and near infrared (NIR). That paper is similar to ours in that it uses remote sensing for monitoring *H. sosnowskyi*, but differs in performing the classification as we have used the pixel based random forest classification instead of the supervised and unsupervised classification algorithms utilized in this paper.

The spectral signature of *H. sosnowskyi* is so similar to grass and, as a result, our classification yields a lot of false-positives. We decided to collect more sample data for grass and also, we have included sample data for other common classes, which are: Urban, Bare Land, Water, Forest, and Crop, and we grouped all under the class name *others*. Finally, we made a binary classification which consists of *H. sosnowskyi* on one side and all the reaming classes on the other side. This enabled us to minimize these false positives that most of the time occur between *H. sosnowskyi* and Grass. Based on this approach, we are able to more accurately identify the locations of *H. sosnowskyi*.

Generalization of the Model for Detecting Different Plants or in Different Areas?

The RF modeling applied in GEE was successful in predicting *H. sosnowskyi* occurrences. If the model is applied to other geographic areas, a similar successful classification is expected. One factor that could mislead the model is that, as the geographic extent changes across large distances, *H. sosnowskyi* might have different spectral signatures, which could decrease model performance or increase the number of false positives. This change could originate in the different yearly cycle as climate changes from place to place or other plants which are similar to *H. sosnowskyi* begin appearing.

Like any plant, *H. sosnowskyi* undergoes periodic changes in phenological phases (growth of green mass, flowering, seed maturation, wilting). This species has adapted well to the conditions of different climatic zones of the Russian Federation and can grow in almost all regions and weather conditions, as a result of which the periods of change of the phenological phases of *H. sosnowskyi* can differ for different territories. One should therefore be cautious in shifting too much from the geography of the training area, as the other territories have different ecosystems. If a more widespread application was to be made, this issue could be overcome by including training data from several different areas.

Conclusions

In this research, we employed GEE, which represents the programming environment that allowed complex image processing and analysis using the javascript language. It combines the data catalogue and computing environment, for detecting *H. sosnowskyi* in the Moscow region, Russia. GEE has achieved substantial success thanks to its free cloud-based platform that overcame the remote sensing application's intrinsic limitation of data storage, retrieval, and pre-processing. GEE has created a conducive environment in LULC classification. We have applied the pixel-based approach, using which we have visually selected the training data from GEE environment. The data for the classification is prepared by selecting the best time window of the year, applying a cloud-filtering, calculating various spectral indices, and integrating all these to be used as features in the RF classifier.

We created a classification of 2 classes using an RF algorithm on the data collected from Sentinel-2. We have used multispectral imagery as the input for the classification. In addition to the 13 bands in the imagery, we have added the elevation and its derived slope together with the computed indices to improve the classification accuracy. The overall classification accuracy attained is 96.93 % and this is encouraging because it enables prediction of the presence of *H. sosnowskyi* with that accuracy. The outcomes of this study could assist in analyzing the recent status of *H. sosnowskyi*. This kind of geographic information can be the basis of management and conservation practices to prevent future spread of *H. sosnowskyi* and other similar invasive plant species.

References

Afonin A., Luneva N., Li Y., Kotsareva N. Ecological-geographical analysis of distribution pattern and occurrence of cow-parsnip (Heracleum sosnowskyi Manden) with respect to area aridity and its mapping in European Russia // Russ. J. Ecol. — 2017. — Vol. 48, No. 1. — P. 86–89.

Amani M., Ghorbanian A., Ahmadi S., Kakooei M., Moghimi A., Mirmazloumi S., Moghaddam S., Mahdavi S., Ghahremanloo M., Parsian S., Wu Q., Brisco B. Google Earth Engine cloud computing platform for remote sensing big data applications: A comprehensive review // IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. — 2020. — Vol. 13. — P. 5326–5350.

Baležentiene L., Stankeviciene A., Snieškiene V. Heracleum sosnowskyi (Apiaceae) seed productivity and establishment in different habitats of central Lithuania // Ekologija. — 2013. — Vol. 59. — P. 123–133.

Belgiu M., Drăgu L. Random forest in remote sensing: A review of applications and future directions // ISPRS J. Photogramm. Remote Sens. — 2016. — Vol. 114, No. 4. — P. 24–31.

Breiman L. Random forests // Mach. Learn. — 2001. — Vol. 45. — P. 5–32.

Chadin I., Dalke I., Zakhozhiy I., Malyshev R., Madi E., Kuzivanova O., Kirillov D., Elsakov V. Distribution of the invasive plant species Heracleum sosnowskyi Manden in the Komi Republic (Russia) // PhytoKeys. — 2017. — Vol. 77. — P. 71–80.

Cutler D., Edwards T., Beard K., Cutler A., Hess K., Gibson J., Lawler J. Random forests for classification in ecology // Ecology. — 2007. — Vol. 88, No. 11. — P. 2783–2792.

Dalke I., Chadin I., Zakhozhiy I., Malyshev R., Maslova S., Tabalenkova G., Golovko T. Traits of heracleum sosnowskyi plants in monostand on invaded area // PLoS One. — 2015. — Vol. 11, No. 10.

Dalke I., Chadin I., Zakhozhiy I. Control of Sosnowskyi's hogweed (Heracleum sosnowskyi Manden) invasion on the territory of the Russian Federation // Russ. J. Biol. Invasions. — 2018. — Vol. 10, No. 9. — P. 331–344.

Dalke I. V., Chadin I. F., Malyshev R. V., Zakhozhiy I. G., Tishin D. V., Kharevsky A. A., Solod E. G., Shaikina M. N., Popova M. Y., Polyudchenkov I. P., Tagunova I. I., Lyazev P. A., Belyaeva A. V. Laboratory and field assessment of the frost resistance of Sosnowsky's hogweed // Russ. J. Biol. Invasions. — 2020. — Vol. 11, No. 1. — P. 9–20.

Glushakova A., Kachalkin A., Chernov I. Soil yeast communities under the aggressive invasion of Sosnowsky's hogweed (Heracleum sosnowskyi) // Eurasian Soil Sci. — 2015. — Vol. 48, No. 3. — P. 201–207.

Gorelick N., Hancher M., Dixon M., Ilyushchenko S., Thau D., Moore R. Google Earth Engine: Planetary-scale geospatial analysis for everyone // Remote Sens. Environ. — 2017. — Vol. 202, No. 3. — P. 18–27.

Grzedzicka E. Invasion of the giant hogweed and the Sosnowsky's hogweed as a multidisciplinary problem with unknown future — a review // Earth. — 2022. — Vol. 3. — P. 287–312.

Huang C., Asner G. Applications of remote sensing to alien invasive plant studies // Sensors. — 2009. — Vol. 9. — P. 4869–4889.

Jahodová Š., Trybush S., Pyšek P., Wade M., Karp A. Invasive species of Heracleum in Europe: An insight into genetic relationships and invasion history // Divers. Distrib. — 2007. — Vol. 13, No. 1. — P. 99–114.

Jakubowicz O., Żaba C., Nowak G., Jarmuda S., Żaba R., Marcinkowski J. Heracleum sosnowskyi Manden // Ann. Agric. Environ. Med. — 2012. — Vol. 19. — P. 327–328. — https://www.aaem.pl/Heracleum-sosnowskyi-Manden,71782,0,2.html

Jakubska-Busse A., Śliwiński M., Kobylka M. Identification of bioactive components of essential oils in Heracleum sosnowskyi and Heracleum mantegazzianum (apiaceae) // Arch. Biol. Sci. — 2013. — Vol. 65. — P. 877–883.

Kabuce N., Priede N., Priede A., Priede N. NOBANIS — Invasive alien species fact sheet Heracleum sosnowskyi // Online Database North Eur. Balt. Netw. Invasive Alien Species – NOBANIS. — 2010. — Vol. 8. — P. 2013.

Kwok R. Ecology's remote-sensing revolution // Nature. — 2018. — P. 137–138.

Nielsen C., Ravn H. P., Nentwig W., Wade M. (eds.) The giant hogweed best practice manual: guidelines for the management and control of invasive weeds in Europe. — 2005.

Ozerova N., Shirokova V., Krivosheina M., Petrosyan V. The spatial distribution of Sosnowsky's hogweed (Heracleum sosnowskyi) in the valleys of big and medium rivers of the East European Plain (on materials of field studies 2008–2016) // Russ. J. Biol. Invasions. — 2017. — Vol. 8, No. 12. — P. 327–346. — https://link.springer.com/article/10.1134/S2075111717040075

Ozerova N., Krivosheina M. Patterns of secondary range formation for Heracleum sosnowskyi and H. mantegazzianum on the territory of Russia // Russ. J. Biol. Invasions. — 2018. — Vol. 9, No. 4. — P. 155–162.

Panasenko N. On certain issues of biology and ecology of Sosnowsky's hogweed (Heracleum sosnowskyi Manden) // Russ. J. Biol. Invasions. — 2017. — Vol. 8, No. 7. — P. 272–281.

Paz-Kagan T., Silver M., Panov N., Karnieli A. Multispectral approach for identifying invasive plant species based on flowering phenology characteristics // Remote Sens. — 2019. — Vol 11. — P. 953.

Prishchepov A. Map of hogweed Sosnovsky to the Moscow Region — Stop hogweed. — Available at: https://borshevictory.ru/project/karta-borshhevika-na-moskovskuyu-oblast/ (accessed: September 17, 2022).

Renčo M., Kornobis F., Domaradzki K., Jakubska-Busse A., Jurová J., Homolová Z. How does an invasive Heracleum sosnowskyi affect soil nematode communities in natural conditions? // Nematology. — 2019. — Vol. 21. — P. 71–89.

Vaz A., Alcaraz-Segura D., Campos J., Vicente J., Honrado J. Managing plant invasions through the lens of remote sensing: A review of progress and the way forward // Sci. Total Environ. — 2018. — Vol. 642, No. 3. — P. 1328–1339.

Vaz A., Alcaraz-Segura D., Vicente J., Honrado J. The many roles of remote sensing in invasion science // Front. Ecol. Evol. — 2019. — Vol. 7. — P. 1–5.

Vicente J., Alagador D., Guerra C., Alonso J., Kueffer C., Vaz A., Fernandes R., Cabral J., Araújo M., Honrado J. Cost-effective monitoring of biological invasions under global change: a model-based framework // J. Appl. Ecol. — 2016. — Vol. 53. — P. 1317–1329.

Visockiene J., Tumeliene E., Maliene V. Identification of Heracleum sosnowskyi-invaded land using earth remote sensing data // Sustain. — 2020. — Vol. 12, No. 1. — P. 759.

Xie Y., Sha Z., Yu M. Remote sensing imagery in vegetation mapping: a review // J. Plant Ecol. — 2008.