**Spatiotemporal mapping of invasive yellow sweetclover blooms using Sentinel-2 and high-resolution drone imagery**

Sakshi Saraf1,\*, Ranjeet John1,2,, Venkatesh Kolluru2, Khushboo Jain2, Geoffrey Henebry3,4, Jiquan Chen3,4, Raffaele Lafortezza5

\*Corresponding author: Sakshi Saraf; Email: Sakshi.Saraf@coyotes.usd.edu; Orcid: 0000-0002-6785-8381

1. Department of Biology, University of South Dakota, Vermillion, SD 57069, USA

2. Department of Sustainability and Environment, University of South Dakota, Vermillion, SD 57069, USA

3. Department of Geography, Environment, and Spatial Sciences, Michigan State University, East Lansing, MI 48824, USA

4. Center for Global Change and Earth Observations, Michigan State University, East Lansing, MI 48823, USA

5. Department of Soil, Plant and Food Sciences (DISSPA), University of Bari “Aldo Moro”, Via Amendola 165/A, 70126, Bari, Italy

**Tables and Figures**

Table S1. Source and year-wise distribution of the yellow sweetclover sample points retrieved and synthesized across four states from 2016-2023 (n = 22,972).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Sources** | **2016** | **2017** | **2018** | **2019** | **2020** | **2021** | **2022** | **2023** |
| BLM-LMF | 31 | 23 | 25 | 34 | 124 |  |  |  |
| MTNHP | 76 | 110 | 39 | 107 |  |  |  |  |
| NEON | 64 | 143 | 146 | 121 |  | 10 | 16 |  |
| NGPN | 32 | 34 | 41 | 34 |  |  |  |  |
| BLM-TERRA | 40 | 72 | 90 | 103 | 80 | 28 |  |  |
| USGS\_MT |  |  |  | 16,000 |  |  |  |  |
| USGS\_EROS |  |  |  | 31 |  | 3 |  |  |
| FC |  |  |  |  |  |  | 32 |  |
| FC-UAV |  |  |  |  |  |  |  | 2,736 |
| FC-Validation |  |  |  |  |  |  |  | 2,547 |
| **Total** | **243** | **382** | **341** | **16,430** | **204** | **41** | **48** | **5,283** |

Footnote: USGS\_MT - United States Geological Survey Northern Rocky Mountain Science Center, Montana, USGS\_EROS- USGS Center for Earth Resources Observation & Science, USD - University of South Dakota, AIM- Assessment Inventory and Monitoring, BLM - Bureau of Land Management, NPS - National Park Service, MTNHP - Montana Natural Heritage Program.

Table S2. Abbreviations for the sources of the data retrieved and synthesized from 2016-2023.

|  |  |  |
| --- | --- | --- |
| No | Sources of the data | Codes |
| 1 | RCMAP data - USGS Center for Earth Resources Observation & Science | USGS\_EROS |
| 2 | USGS Northern Rocky Mountain Science Center, Montana | USGS\_MT |
| 3 | Terrestrial AIM database – BLM | TERRA |
| 4 | Landscape Monitoring Framework database - BLM | LMF |
| 5 | USD summer field data collection | FC |
| 6 | Northern Great Plains Inventory & Monitoring Network data - NPS | NGPN |
| 7 | National Ecological Observatory Network database funded by the National Science Foundation | NEON |
| 8 | Montana Natural Heritage Program | MTNHP |

Footnote: USGS - United States Geological Survey, USD - University of South Dakota, AIM- Assessment Inventory and Monitoring, BLM - Bureau of Land Management, NPS - National Park Service, RCMAP-Rangeland Condition Monitoring Assessment and Projection

Table S3. Summary statistics of yellow sweetclover percent cover (n = 22,972) used in this study collected from 2016-2023

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Year** | | **Minimum** | **1st Quartile** | **Median** | **Mean** | **3rd Quartile** | **Maximum** |
| 2016 | 1.0 | | 1.0 | 3.5 | 7.4 | 9.0 | 51.0 |
| 2017 | 0.5 | | 1.0 | 2.0 | 4.4 | 5.0 | 33.0 |
| 2018 | 0.5 | | 1.0 | 3.7 | 8.5 | 10.1 | 82.0 |
| 2019 | 0.5 | | 40.0 | 65.0 | 65.5 | 85.0 | 100.0 |
| 2020 | 0.5 | | 1.0 | 3.3 | 6.9 | 8.0 | 64.0 |
| 2021 | 0.5 | | 0.7 | 1.3 | 1.7 | 2.0 | 5.0 |
| 2022 | 2.0 | | 5.0 | 9.0 | 8.2 | 10.0 | 15.0 |
| 2023 | <0.1 | | 14.9 | 38.6 | 42.7 | 67.1 | 100.0 |
| **Total samples** | | 0.0 | 20.0 | 65.0 | 55.5 | 85.0 | 100.0 |

Table S4. Description of the multispectral indices and tasseled cap used in this study **(**abbreviations in Table S4).

|  |  |  |  |
| --- | --- | --- | --- |
| **Vegetation Index** | **Formula** | **Sentinel Bands** | **References** |
| Normalized Difference Vegetation Index (NDVI) | *NDVI* | and  are bands 8 and 4 | (Rouse et al., 1973) |
| Normalized Difference Moisture Index (NDMI) | *NDMI* = | and are bands 8 and 11 | (Xiao et al., 2005) |
| Normalized Difference Yellowness Index (NDYI) | *NDYI* = | andare bands 3 and 2 | (Sulik & Long, 2016) |
| Land Surface Water Index (LSWI) | *LSWI* = | andare bands 8A and 11 | (Xiao et al., 2005) |
| Normalized Difference Water Index (NDWI) | *NDWI* = | and are bands 8 and 12 | (Henrich et al., 2009) |
| Tasseled Cap Brightness (TCB) |  | as band 2,  as band 3,  as band 4,  as band 8,  as band 11 and  as band 12 | (Shi & Xu, 2019) |
| Tasseled Cap Greenness (TCG) |  |
| Tasseled Cap Wetness (TCW) |  |

Table S5. Description and source details of all the 64 independent variables considered for the study.

|  |  |  |
| --- | --- | --- |
| **Data Source with their spatial resolution** | **Variables** | **Codes and units** |
| Daymet dataset (1 km) | Mean (coefficient of variation - cv) annual precipitation | MAP (MAPcv) [mm] |
|  | Mean (cv) annual temperature | MAT (MATcv) [°C] |
|  | Biennial mean annual precipitation (cv) | MAP2 (MAP2cv) [mm] |
|  | Biennial mean (cv) annual temperature | MAT2 (MAT2cv) [°C] |
|  | Summer mean (cv) precipitation (June-Aug) | PJJA (PJJAcv) [mm] |
|  | Spring mean (cv) precipitation (Mar-May) | PMAM (P\_MAMcv) [mm] |
|  | Summer mean (cv) temperature (June-Aug) | TJJA (T\_JJAcv) [°C] |
|  | Spring mean (cv) temperature (Mar-May) | TMAM (T\_MAMcv) [°C] |
| MODIS MOD10A1 V6.1 derived Normalized Difference Snow Index (NDSI) percent snow cover (500m) | Winter mean (cv) snow cover (Dec-Feb) | SNOWc (SNOWc\_cv) |
| Seasonal mean composites of snow depth derived from NOAA National Weather Service's National Operational Hydrologic Remote Sensing Center (NOHRSC) SNOw Data Assimilation System (SNODAS) (1 km) | Winter mean (cv) snow depth (Dec-Feb) | Sdepth (Sdepth\_cv) (m) |
| Winter mean (cv) snow water equivalent (Dec-Feb) | SWE (SWEcv) |
| USGS National Elevation Dataset (10m) | Elevation | Elevation [m] |
| Slope | Slope [degrees] |
| Aspect | Aspect [degrees] |
| Hillshade | Hillshade |
| Terrain Roughness Index | TRI |
| Topographic Wetness Index | TWI |
| Maximum value composites (MVC) of each vegetation indices were derived independently from Sentinel 2A (10m) | Normalized Difference Vegetation Index | NDVI (NDVIcv) |
| Normalized Difference Moisture Index | NDMI (NDMIcv) |
| Normalized Difference Yellowness Index | NDYI (NDYIcv) |
|  | Land Surface Water Index | LSWI (LSWIcv) |
| Normalized Difference Water Index | NDWI (NDWIcv) |
| S2A-Green | Green (Green\_cv) |
| S2A- Red | Red (Red\_cv) |
| S2A-Near InfraRed | NIR (NIR\_cv) |
| S2A-Short Wave InfraRed 1 | SWIR1 (SWIR1\_cv) |
| S2A-Short Wave InfraRed 2 | SWIR2 (SWIR2\_cv) |
| Dimensionality reduction of Sentinel 2A multispectral bands to create an n-band image with the first 3 bands containing 3 principal components represented as Brightness, Greenness and Wetness and their standard deviations. MVC were derived each tasseled cap independently. (10m) | Tasseled Cap Brightness | TCB (TCBcv) |
| Tasseled Cap Wetness | TCW (TCWcv) |
| Tasseled Cap Greenness | TCG (TCGcv) |
| Polaris database developed by National Cooperative Soil Survey under USDA-Natural Resources Conservation Service (30m) | Sand | Sand [%] |
| Silt | Silt [%] |
| Clay | Clay [%] |
| Soil pH | Soil\_pH [1-14] |
|  | Soil Organic Matter | SOM [log₁₀(%)] |
|  | Bulk Density | BD [g/cm³] |
|  | Residual soil water content | Theta\_r [m³/m³] |
|  | Saturated soil water content | Theta\_s [m³/m³] |
| National Land Cover Database (NLCD) imperviousness (CONUS) product (30m) | Proximity to roads | Dist\_roads [m] |
| National Hydrography Dataset developed by USGS National Geospatial Program (30m) | Distance to streams | Dist\_streams [m] |

Table S6 Confusion matrix for Random Forest Classification model

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Predicted Absence | Predicted Presence | Row Total | User’s Accuracy (%) |
| Observed Absence | 605 | 9 | 614 | 98.5% |
| Observed Presence | 6 | 585 | 591 | 99% |
| Column Total | 611 | 594 | 1205 |  |
| Producer’s Accuracy (%) | 99% | 98.5% |  |  |
| True Positive Rate | 0.99 |  |  |  |
| False Positive Rate | 0.015 |  |  |  |
| Area under curve (AUC) | 0.987 |  |  |  |
| F1 Score | 0.987 |  |  |  |

Table S7 Estimated area covered by the Yellow sweetclover presence derived from the predicted Yellow sweetclover (MEOF) classification model.

|  |  |  |  |
| --- | --- | --- | --- |
| **Sites** | **MEOF presence (ha)** | **Total area (ha)** | **MEOF presence area (%)** |
| 1 | 1.3 | 10.5 | 12.47 |
| 2 | 0.6 | 1.9 | 29.51 |
| 3 | 0.8 | 4.9 | 17.12 |
| 4 | 0.6 | 4.1 | 14.50 |
| 5 | 6.6 | 0.4 | 21.64 |
| 6 | 1.3 | 3.2 | 40.32 |
| 7 | 1.1 | 7.2 | 15.54 |
| 8 | 1.7 | 3.0 | 55.23 |
| 9 | 1.1 | 4.9 | 21.77 |
| 10 | 0.8 | 4.6 | 17.44 |
| 11 | 1.7 | 4.2 | 39.40 |
| 12 | 2.4 | 7.2 | 32.74 |
| 13 | 2.4 | 10.5 | 22.66 |
| 14 | 1.0 | 4.7 | 20.51 |

Table S8 Annual Moran’s I test results with a threshold distance of 50m for spatial autocorrelation for yellow sweetclover data from years 2016-2023. \* indicates p <0.05.

|  |  |
| --- | --- |
| **Years** | **Moran’s I** |
| 2016 | -0.02 |
| 2017 | 0.14\* |
| 2018 | 0.00 |
| 2019 | 0.34\* |
| 2020 | 0.13\* |
| 2021 | 0.16\* |
| 2022 | -0.07 |
| 2023 | 0.29\* |

Table S9 Year-wise normalized Root Mean Square Error (nRMSE) value for predicted versus observed MEOF cover. Normalization was performed using the mean observed cover per year to account for scale differences in percent cover across years.

|  |  |
| --- | --- |
| Year | nRMSE |
| 2016 | 0.39 |
| 2017 | 0.23 |
| 2018 | 0.19 |
| 2019 | 0.23 |
| 2020 | 0.55 |
| 2021 | 0.25 |
| 2022 | 0.12 |
| 2023 | 0.65 |

Figure S1 Correlation matrix showing the top 16 selected predicting variables having correlation value less than the threshold value of 0.8.



Figure S2 Histogram showing the predicted percent cover range for the years 2019 and 2023.

![A graph of different sizes and colors

AI-generated content may be incorrect.]()

Figure S3 Variable importance from the Random Forest regression model for predicting the yellow sweetclover percent cover.

A graph with lines and dots

AI-generated content may be incorrect.

\* NDMI- Normalized Difference Moisture Index, Dist\_Roads – Proximity to roads, NDWIcv – Variability in Normalized Difference Water Index, TCWcv- Variability in Tasseled Cap Wetness, LSWIcv – Variability in Land Surface Water Index, SnowDepth\_cv – Variability in snow depth, SnowDepth – Snow Depth, MATcv – Variability in mean annual temperature, MAP – Mean annual precipitation, MAPcv – Variability in mean annual precipitation, MAT – Mean annual temperature.

Figure S4 Precipitation anomalies for 2016-2023 from 30 years (1990-2020) average annual precipitation derived using the Daymet dataset.

A collage of maps of different colors

AI-generated content may be incorrect.

Figure S5. Predicted MEOF percent cover estimates for the year 2024 using the Random Forest model developed using observed MEOF percent cover samples collected from 2016-2023.

A map of the state of minnesota

AI-generated content may be incorrect.

Figure S6. Predicted percent cover of yellow sweetclover (MEOF) at the four validation sites obtained from the random forest model against the percent cover derived through weighted average of MEOF presence from UAV-imagery (resampled to 10 m). The shaded area represents the95% confidence interval for the predicted values at each value of the observed sample.

A blue line with a line going up