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

Sakshi Saraf[1,*], Ranjeet John[1,2,], Venkatesh Kolluru[2], Khushboo Jain[2], Geoffrey Henebry[3,4], Jiquan Chen[3,4], Raffaele Lafortezza[5]

*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 collected 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.

26   Table S2. Abbreviations for the sources of the data collected 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

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

30

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

33

34 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)	$NDVI = \dfrac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}}$	ρ_{nir} and ρ_{red} are bands 8 and 4	(Rouse et al., 1973)
Normalized Difference Moisture Index (NDMI)	$NDMI = \dfrac{\rho_{nir} - \rho_{swir1}}{\rho_{nir} + \rho_{swir1}}$	ρ_{nir} and ρ_{swir1} are bands 8 and 11	(Xiao et al., 2005)
Normalized Difference Yellowness Index (NDYI)	$NDYI = \dfrac{\rho_{green} - \rho_{blue}}{\rho_{green} + \rho_{blue}}$	ρ_{green} and ρ_{blue} are bands 3 and 2	(Sulik & Long, 2016)
Land Surface Water Index (LSWI)	$LSWI = \dfrac{\rho_{nir\_re} - \rho_{swir1}}{\rho_{nir\_re} + \rho_{swir1}}$	ρ_{nir_re} and ρ_{swir} are bands 8A and 11	(Xiao et al., 2005)
Normalized Difference Water Index (NDWI)	$NDWI = \dfrac{\rho_{nir} - \rho_{swir2}}{\rho_{nir} + \rho_{swir2}}$	ρ_{nir} and ρ_{green} are bands 8 and 12	(Henrich et al., 2009)
Tasseled Cap Brightness (TCB)	$TC_{bright} = 0.3510 \times \rho_{blue} + 0.3813 \times \rho_{green} + 0.3437 \times \rho_{red} + 0.7196 \times \rho_{nir} + 0.2396 \times \rho_{swir1} + 0.1949 \times \rho_{swir2}$	ρ_{blue} as band 2, ρ_{green} as band 3, ρ_{red} as band 4, ρ_{nir} as band 8, ρ_{swir1} as band 11 and ρ_{swir2} as band 12	(Shi & Xu, 2019)
Tasseled Cap Greenness (TCG)	$TC_{green} = -0.3599 \times \rho_{blue} - 0.3533 \times \rho_{green} - 0.4734 \times \rho_{red} + 0.6633 \times \rho_{nir} + 0.0087 \times \rho_{swir1} - 0.2856 \times \rho_{swir2}$		
Tasseled Cap Wetness (TCW)	$TC_{wetness} = 0.2578 \times \rho_{blue} + 0.2305 \times \rho_{green} + 0.0883 \times \rho_{red} + 0.1071 \times \rho_{nir} - 0.7611 \times \rho_{swir1} - 0.5308 \times \rho_{swir2}$		

35
36

37

38   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 [$\log_{10}$(%)]
	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]

39

40

41   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			

42

43 Table S7 Estimated area covered by the Yellow sweetclover presence derived from the predicted
44 Yellow sweetclover (MEOF) classification model.
45

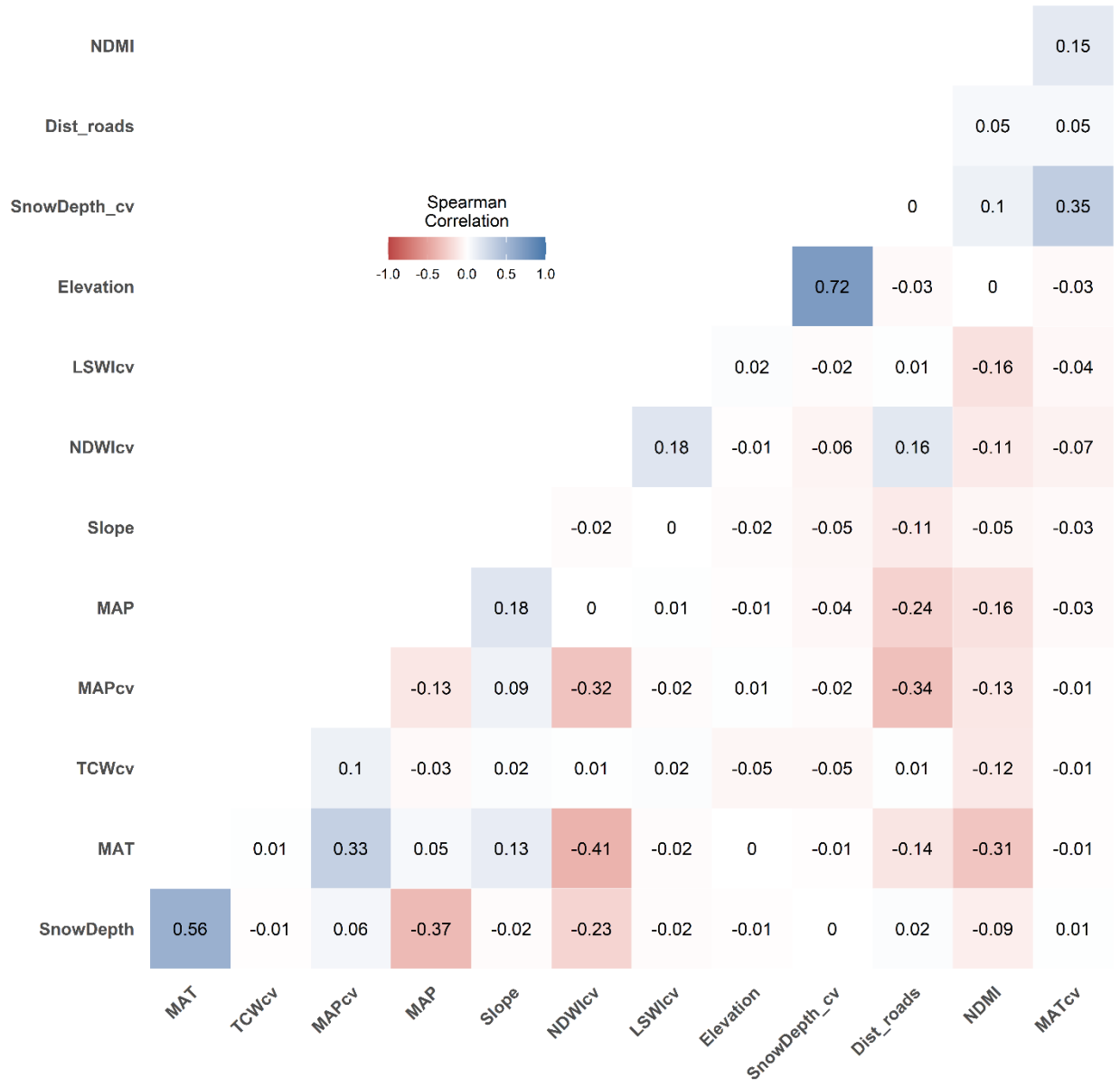
Sites	MEOF presence (ha)	Total area (ha)	MEOF presence area (%)
1	1.3	10.5	12.47
2	0.6	1.9	29.51
3	0.8	4.9	17.12
4	0.6	4.1	14.50
5	6.6	0.4	21.64
6	1.3	3.2	40.32
7	1.1	7.2	15.54
8	1.7	3.0	55.23
9	1.1	4.9	21.77
10	0.8	4.6	17.44
11	1.7	4.2	39.40
12	2.4	7.2	32.74
13	2.4	10.5	22.66
14	1.0	4.7	20.51

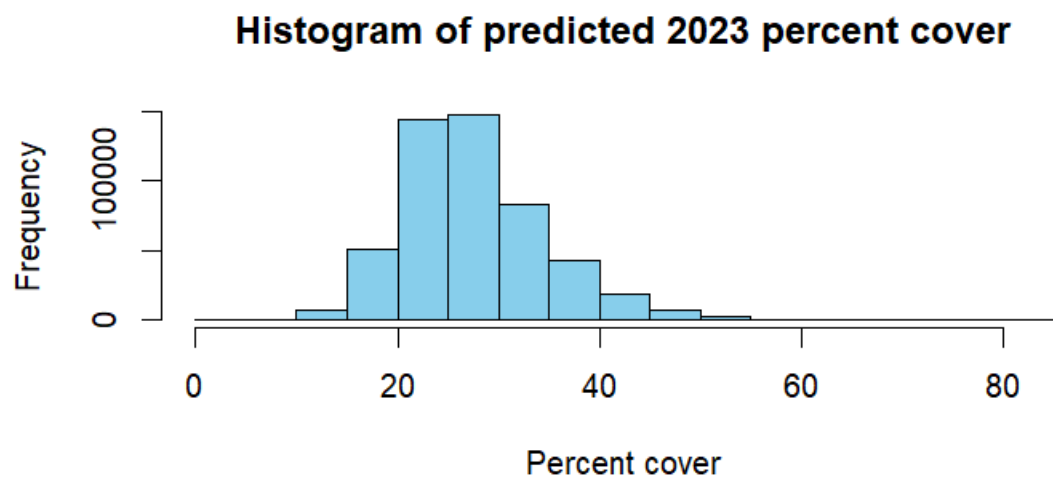
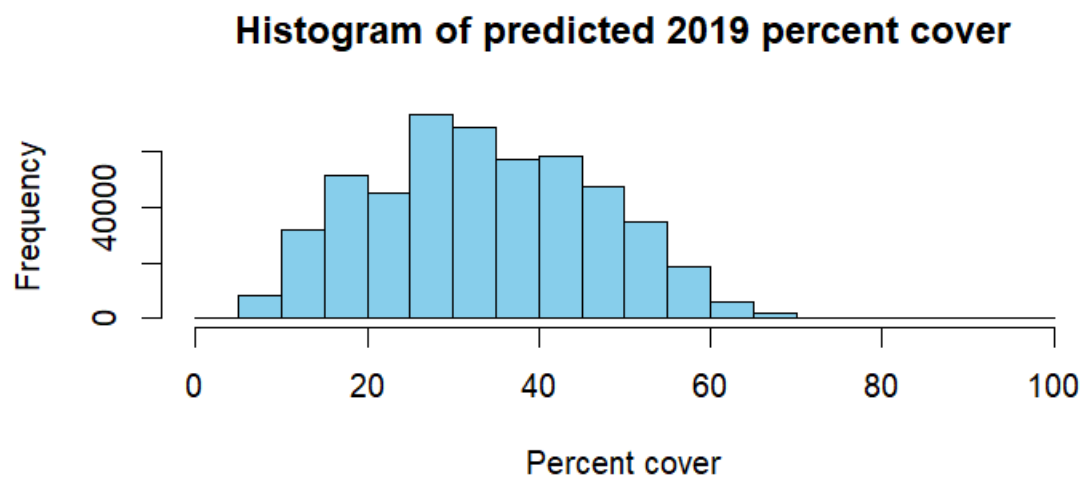
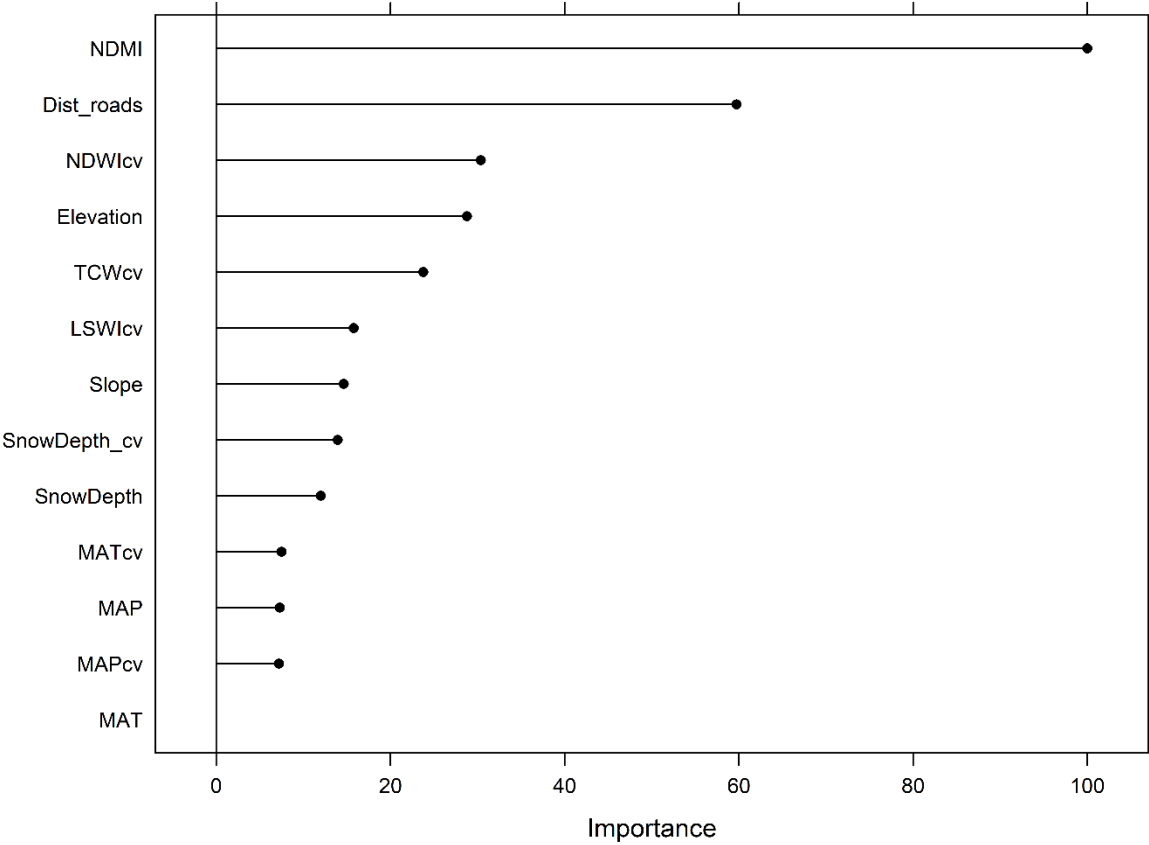
46
47
48

49 Table S8 Annual Moran's I test results with a threshold distance of 50m for spatial
50 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*

51

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

57   Figure S2 Histogram showing the predicted percent cover range for the years 2019 and 2023.
58

59

60 Figure S3 Variable importance from the Random Forest regression model for predicting the
61 yellow sweetclover percent cover.
62

63
64 * NDMI- Normalized Difference Moisture Index, Dist_Roads – Proximity to roads, NDWIcv – Variability in
65 Normalized Difference Water Index, TCWcv- Variability in Tasseled Cap Wetness, LSWIcv – Variability in Land
66 Surface Water Index, SnowDepth_cv – Variability in snow depth, SnowDepth – Snow Depth, MATcv – Variability
67 in mean annual temperature, MAP – Mean annual precipitation, MAPcv – Variability in mean annual precipitation,
68 MAT – Mean annual temperature.

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

71

72

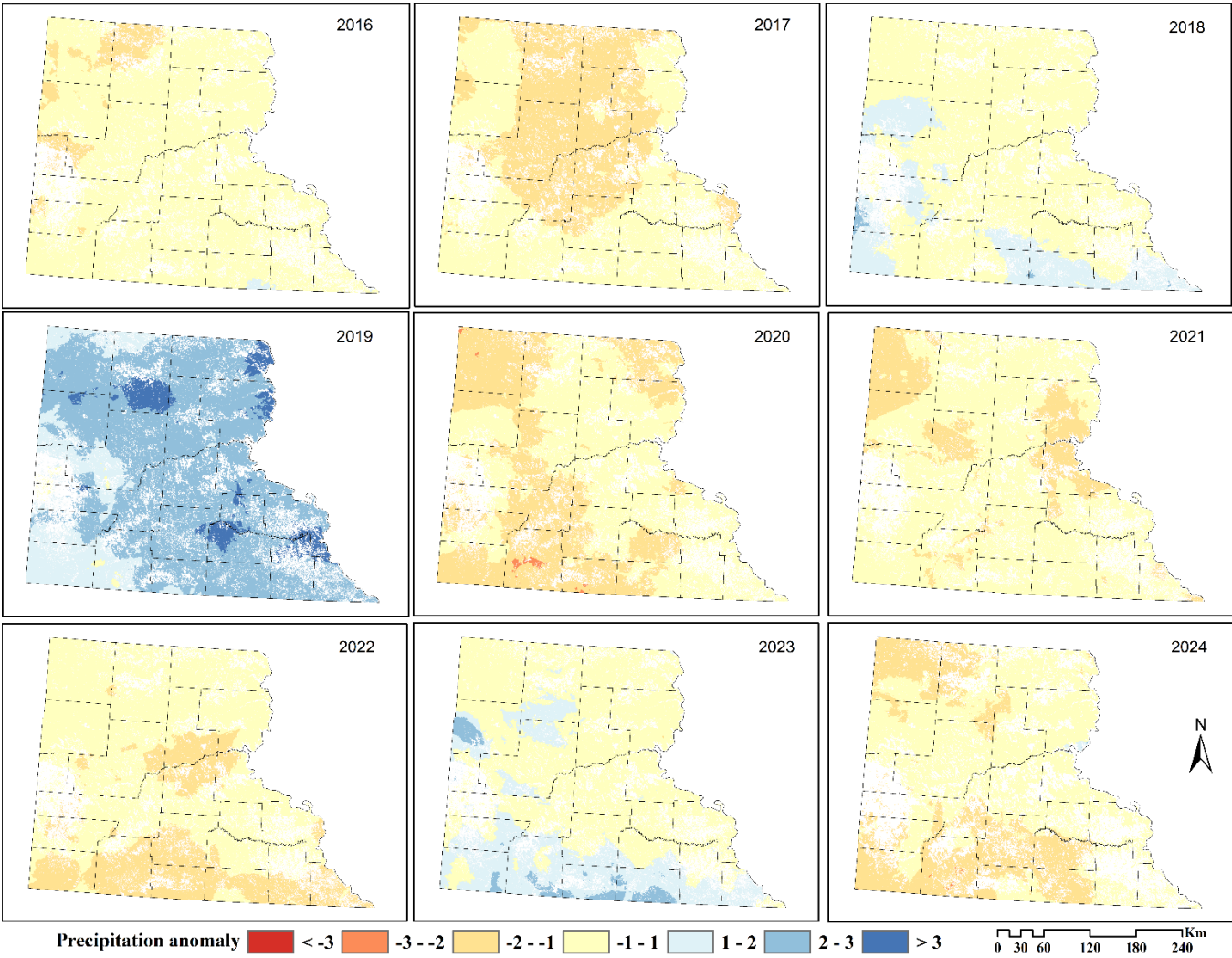
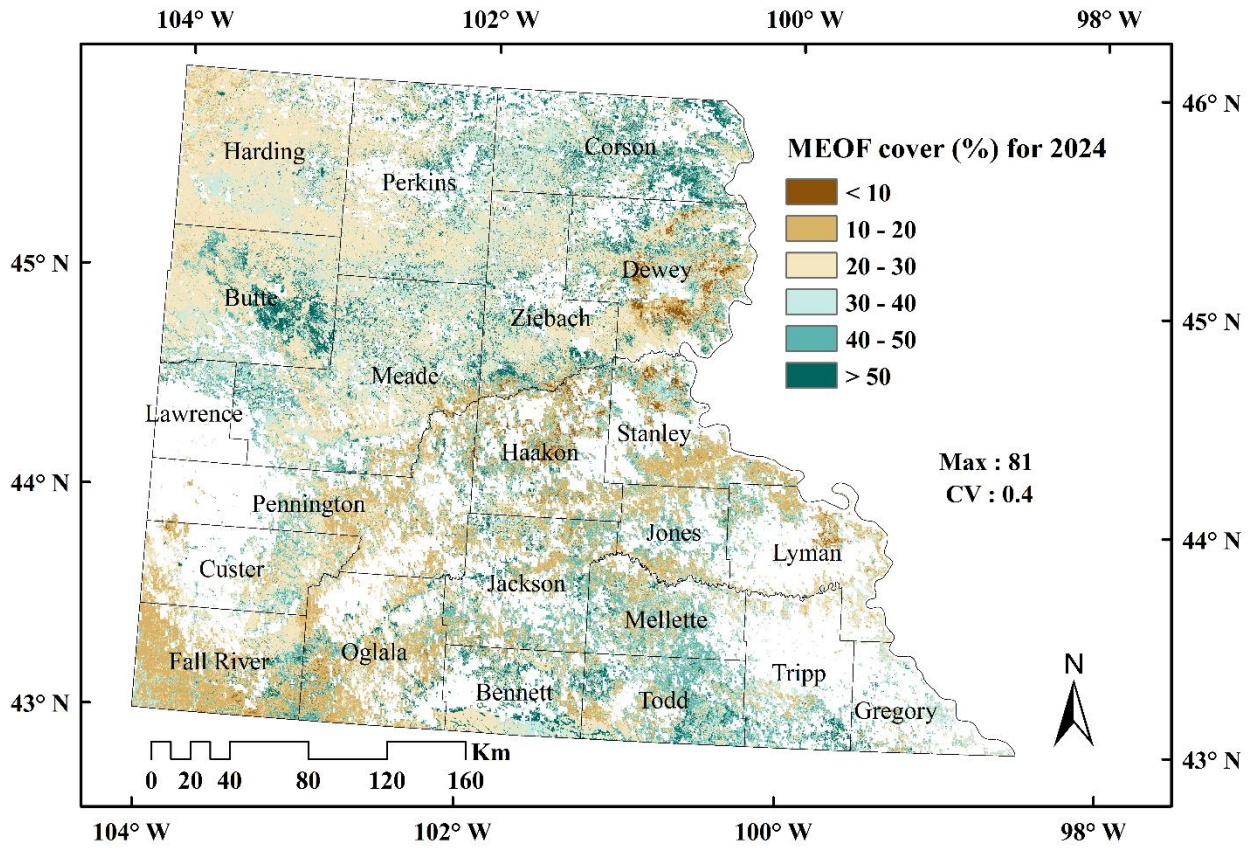



Figure S5. Predicted MEOF percent cover estimates for the year 2024 using the Random Forest model developed using observed MEOF percent cover samples collected from 2016-2023.

77 Figure S6. Predicted percent cover of yellow sweetclover (MEOF) at the four validation sites
78 obtained from the random forest model against the percent cover derived through weighted
79 average of MEOF presence from UAV-imagery (resampled to 10 m). The shaded area represents
80 the 95% confidence interval for the predicted values at each value of the observed sample.
81

82