LANDCOVER MAPPING OF CHARLES M. RUSSELL WILDLIFE REFUGE: SPOT VS. LANDSAT 8

Ute Langner, Claudine Tobalske and Linda Vance, Montana Natural Heritage Program, University of Montana. December 2014

PROJECT OBJECTIVES

This study had two objectives: first, to generate a landcover map for the Charles M. Russell Wildlife Refuge (CMR) emphasizing the distribution of land cover types in relation to greater sage grouse (*Centrocercus urophasianus*) habitat needs, and second, to provide data that would allow a determination of whether results were better with SPOT imagery or Landsat 8 imagery. SPOT imagery is provided at a 10m pixel resolution, while Landsat 8 is at 30m. Results from this classification will allow managers to determine which resolution provides the accuracy needed for habitat planning and management.

STUDY AREA

The study area was delineated using level 4 HUC boundaries clipped to the overlap of Landsat 8 P37R27 scene and SPOT imagery obtained through the USGS 2011 data buy (for classification comparison); it cover 4,919,395 acres and encompasses most of the Charles M. Russell National Wildlife Refuge (Figure 1).

METHODS

Training points

Training points collected in the field between 2012 and 2013 were grouped into 18 classes: Forested Burn (66), Foothill Woodland Steppe Transition (73), Greasewood Flat (73), Greasewood Steppe (239), Greasewood Sage Steppe (277), Great Plains Badlands (166), Great Plains Riparian (255), Low Density Sage Steppe (776), Medium Density Sage Steppe (783), Mixed Grass Prairie (555), Mixed Grass Prairie Burned (278), Ponderosa Pine Woodland and Shrubland (512), Riparian Floodplain (223), Semi-Desert Grassland (103), Sparsely Vegetated Mixed Shrub (252), Silver Sage Flat (70) , Silver Sage Steppe (64), and Water (246). When insufficient field data were available for a class, we augmented it through photointerpretation of 15 cm aerial imagery, using expert knowledge and field experience to guide us. The final dataset had 5,011 training points.

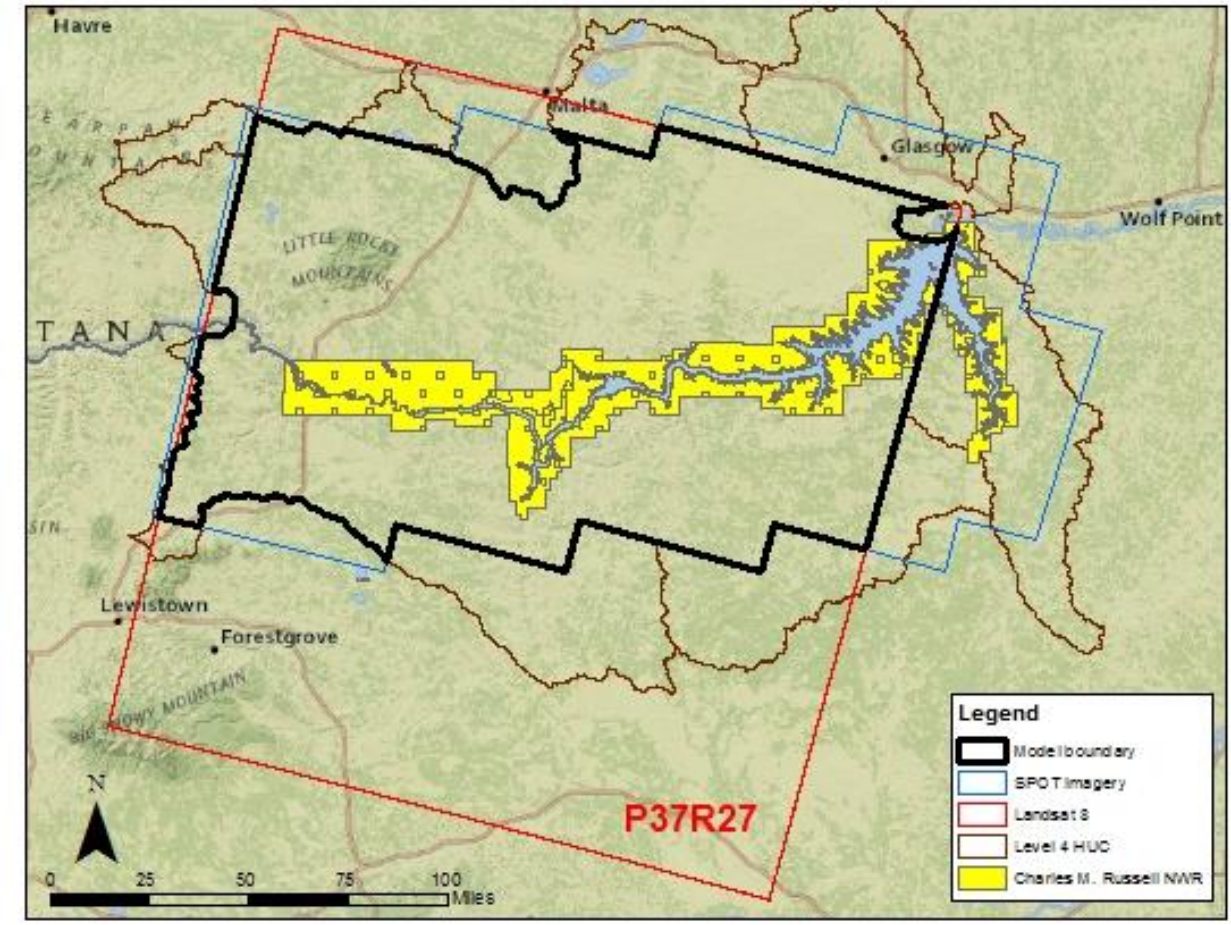

Figure 1. Study area location.

SPOT classification

Data preparation, which included data used in the Landsat classification, included downloading and preprocessing raw SPOT images (including ortho-rectification of images); using 10m Digital Elevation Models (DEM) to derive, slope, aspect; downloading soil layers (SSURGO), and calculate a normalized difference vegetation index (NDVI). We used an atmospheric correction module in the software package ENVI to test different parameters to see which produce best similar reflectance values among images. Parameters included

a) QUAC vs FLAASH vs. Dark Object Subtraction
b) MLS vs. SAS vs US standard models
c) 40, 60, 80, 100 km visibility

We opted for (a) FLAASH, (b) US Standard, (c) 80 km in this step.

We used the Agriculture cover class from the 2013 Landcover Theme in the Montana Spatial Data Infrastructure (MSDI); this cover class is digitized by the Department of Revenue from 1m National Agricultural Imagery Program aerial photos. For remaining cover classes, we segmented the SPOT imagery in the software package eCognition, using a scale of 10 and 15 (eCognition scale has no units), and color parameters of 10 vs 15. After obtaining field data training points, we combined image-derived variables with topographic variables (a 10m DEM, derived aspect, and derived slope (%)) and soils variables generated from the SSURGO soil dataset: clay class and content, sand content, organic matter content, particle size, available water capacity, carbon concentration, geomorphological class, and taxonomic suborder. We then classified the segments with Random Forest in weka (a machine learning algorithm). This proved to be unsatisfactory for sagebrush, greasewood and sparsely vegetated grassland types,, largely due to the sharply dissected topography of the CMR Missouri Breaks area and the extreme patchiness of surface soil colors. Accuracy never exceeded 50 to 60%. Consequently we used the same image derived and ancillary variables to run a pixel-based classification using Random Forest 4.6-7 in R 2.15.2, with the number of trees set at 500. An Erdas Imagine predictive raster layer was extrapolated from the model to the entire study area.

Landsat classification

Landsat 8 scene P37R27 for 8/21/2013 was downloaded from the USGS GloVis website and converted from Digital Numbers to Reflectance, then tasseled-cap transformed, using Landsat 8-specific algorithms in Erdas Imagine. Principal Component Analysis of the 6 reflectance bands was calculated in ArcGIS. Two indices were also computed from the reflectance bands: Normalized Difference Vegetation Index (NDVI: NIR – Red / NIR + Red), and Normalized Moisture Vegetation Index (NDMI: NIR – SWIR1 / NIR + SWIR1).

This resulted in 14 image-derived variables (6 reflectance, brightness, greenness, wetness from tasseled cap transformation, top 3 PCA bands, NDVI, NDMI). Topographic variables included a 10m DEM, derived aspect, and derived slope (%). Nine soils variables were generated from the NRCS SSURGO soil dataset: clay class and content, sand content, organic matter content, particle size, available water capacity, carbon concentration, geomorphological class, and taxonomic suborder. As with the SPOT imagery, All training points were entered in a Random Forest model using RandomForest 4.6-7 in R 2.15.2. Number of trees was set at 500. An Erdas Imagine predictive raster layer was extrapolated from the model to the entire study area.

Post-modelling

Post-modeling was similar to that done for both classifications:

1. Group Greasewood Steppe and Greasewood Sagebrush Steppe;
2. Use the MSDI Landcover Layer to replace the model for the Little Rocky Mountains. Because this is primarily a forested area, the time that would have been involved in modelling forest classes was not justified.
3. Update raster with Agriculture (resampled to 30m in the case of Landsat);
4. Update raster with human development (roads, settlements, mines) from MSDI Landcover;
5. Use river valley bottom delineation to limit the extent of Riparian Floodplain (outside of valley bottom, reclass as Great Plains Riparian);
6. To address the confusion shadow/water: if MSDI Landcover is forested and the pixel does not overlap NHD High Resolution stream, reclass Riparian as forest type (from Landcover).

RESULTS

User's, Producer's, and overall accuracies were generated internally through bootstrapping. Out of-bag estimate of error rate with SPOT classification was 24.5% (i.e. overall accuracy was 75.5%); when we removed the "Foothill Wooded Steppe" category, it was reduced to 22.1. With Landsat 7, out-of-bag error was 27.3% . Tables 1 and 2 show the confusion matrices for both classifications.[1]

For the SPOT classification, producer's accuracy values, as a measure of the accuracy of the classification scheme, were lowest for Greasewood Steppe (28.35%, confused with Greasewood Sage Steppe and Silver Sage Flat). Landsat classification was worst for Greasewood Flat (41.10%, confused with Great Plains Riparian) and Silver Sage Flat (28.13%, confused with Mixed Grass Prairie Burned, Low and Medium Density Sagebrush Steppe, and Semi-Desert Grassland). All these classes had among the lowest numbers of training points. Not surprisingly, the highest producer's accuracy values (> 85% in both classifications) were obtained for classes with a more unique spectral signature, such as Water, Ponderosa Pine, Riparian Floodplain, and Badlands. Silver Sage also classified easily in SPOT, with accuracy of 91.8%. We attempted to tease out Low Density Sage Steppe and Medium Density Sage Steppe, as the latter is more suitable nesting habitat for sage grouse. In the SPOT classification, the producer's accuracy was low for Low Density Sage Steppe (67.1%, confused with Mixed Grass Prairie and Medium Density Sage Steppe) but good (83.3%) for Medium Density Sage Steppe. In the Landsat classification, accuracy values for these two classes were 62.11% and 70.11% respectively. Grouping the two sagebrush classes was more helpful in the Landsat classification

[1] The accuracy assessment for the SPOT classification used only field data points. Silver Sage Steppe and Foothill Wooded Steppe classes were insufficiently represented in the dataset and so were not included in the accuracy assessment, although we did model them for the final raster.

than in the SPOT classification. Grouped, accuracy values for SPOT were 75.75% (producer's) and 68.69% (users). In contrast, grouping brought overall sagebrush accuracy in the Landsat classification to 78.88% (user's) and 84.80% (producer's). User's accuracies, the measure the reliability of an output map generated from a classification scheme, were lowest in both classifications for Low Density Sagebrush Steppe and again, highest for Water, Ponderosa Pine, Riparian Floodplain, and Badlands. A grouping of the two sagebrush classes would again significantly increase accuracy.

Table 3 compares the acreage of the final 28 landcover classes between the two approaches. The Spot classification has more Mixed Grass Prairie and Greasewood Steppe than the Landsat classification, which has more Sagebrush Steppe (both low and medium density). However, the SPOT classification was more accurate for Medium Density Sagebrush Steppe, which is the preferred habitat type for greater sage grouse.

DISCUSSION

The classification of SPOT imagery in this study was fraught with difficulties, which greatly increased the time it took to complete the project. Because we were relying on an existing data buy, rather than commissioning a specific image acquisition (not available within our budget), we had images of varying quality and cloud cover from several different dates, so that each image had to be processed separately. The imagery from the USGS data buy was not orthorectified (as is common when image acquisition is commissioned), so pixel overlay, especially between scenes from different years, was labor-intensive. Atmospheric correction, also something typically done by the SPOT corporation for commissioned imagery, was not done, so we had to acquire new software (ENVI) to fill the gap. However, this introduced the need to format files differently for each software program used in the segmentation and classification (ENVI, ERDAS IMAGINE, and eCognition). All classification was hindered by field data acquisition, which was delayed by an unusually wet spring in 2013.

Although we have long advocated for the advantages of object-oriented classification, we found that it was not suitable for the study area, where surface colors vary widely within areas of comparable vegetation cover. We found that continuous areas of Low Density Sage Steppe were being segmented into multiple "objects" by eCognition, apparently due to surface coloration. However, color differences appeared to be linked to small local phenomena, like blowouts, gravel outwash, patches of *Selaginella densa*, ant mounds, etc., rather than to vegetation variability. This introduced such a high rate of error into the object-oriented classification that we had to abandon it, and revert to a pixel-based classification for both SPOT and Landsat imagery. Unfortunately, pixel-based classification also has limitations, notably in the classification of mixed vegetation communities, like sagebrush steppe, which may have several other shrub species –all with different spectral signatures—mixed in.

We recommend that other sources of imagery be investigated for future classification. While our budgets did not allow for the acquisition of RapidEye, we think it is a promising high-resolution (5m), large-area alternative to SPOT. What this study did demonstrate is that it is possible to tease out medium-density sage steppe from low-density sage steppe, and to isolate areas where greasewood is the dominant shrub species. Refining this approach with more appropriate imagery that comes to the end user in a properly prepared format might provide the level of detail managers need.

As end products for this study, we have produced rasters from both the SPOT and Landsat classifications, which are being sent to land managers in the study area. We acknowledge that accuracy assessment based on training data is only one part of determining which product is "better." The final determination of classification utility and suitability will rest with those who use it on the ground.

	MGP	LD_SS	MD_SS	GWS_S	GW_F	PPWS	SV_MXSH	GPB	GW_S	SS_F	RFP	W	GPRP	burn	MGP-burn	sum	Pacc
MGP	228	36	15	0	0	0	0	0	0	0	1	0	0	0	2	282	80.85
LD_SS	22	221	66	11	0	2	1	0	4	0	0	0	0	0	2	329	67.17
MD_SS	10	39	310	8	1	4	0	0	0	0	0	0	0	0	0	372	83.33
GWS_S	4	16	28	74	2	1	0	3	7	1	2	0	0	1	0	139	53.24
GW_F	5	3	6	0	41	0	0	0	2	0	6	0	1	0	1	65	63.08
PPWS	1	2	2	2	0	92	0	0	0	0	0	0	0	0	0	99	92.93
SV_MXSH	0	2	1	2	0	1	24	1	1	0	0	0	0	0	0	32	75.00
GPB	0	0	0	9	0	0	0	27	0	0	0	0	0	0	0	36	75.00
GW_S	1	11	6	14	2	1	0	0	19	11	2	0	0	0	0	67	28.36
SS_F	2	0	0	0	2	0	0	0	1	56	0	0	0	0	0	61	91.80
RFP	0	0	0	0	2	0	0	0	0	0	134	0	0	0	0	136	98.53
W	0	0	0	0	0	0	0	0	0	0	0	78	0	0	0	78	100.00
GPRP	0	2	0	0	0	0	0	0	1	0	4	0	9	0	0	16	56.25
burn	0	0	0	0	0	3	0	0	0	0	0	0	0	35	0	38	92.11
MGP-burn	0	4	3	0	0	0	0	0	0	0	1	0	0	0	36	44	81.82
Sum	273	336	437	120	50	104	25	31	35	69	150	78	10	36	41	1795	
Uacc	83.5%	65.8%	70.9%	61.7%	82.0%	88.5%	96.0%	87.1%	54.3%	81.2%	89.3%	100.0%	90.0%	97.2%	87.8%		77.10

Table 1. Confusion matrix derived from bootstrapping of the original data (SPOT) Rows and column labels are: MGP=Mixed Grass Prairie; LD_SS=Low Density Sage Steppe; MD_SS= Medium Density Sage Steppe; GWS_S=Greasewood-Sage Steppe; GW_F= Greasewood Flat; PPWS_ Ponderosa Pine Woodland and Shrubland; SV_MXSH= Sparsely Vegetation Mixed Shrub; GPB=Great Plains Badland; GW_S=Greasewood Steppe; SS_F= Silver Sage Flat; RFP= Riparian Floodplain; W= Water; GPRP= Great Plains Riparian; burn=previously forested burned area; MGP-burn= Burned Grassland; Uacc=User's Accuracy; Pacc=Producer's Accuracy.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	sum	Pacc
1	231	0	0	4	2	0	0	1	3	3	0	1	1	0	0	0	0	0	246	93.90
2	0	185	0	0	3	6	9	4	24	10	5	0	1	5	0	0	0	0	252	73.41
3	1	2	54	1	0	0	0	0	2	0	2	4	1	0	1	0	2	0	70	77.14
4	2	0	1	199	2	0	1	0	1	0	0	1	16	0	0	0	0	0	223	89.24
5	0	2	0	0	485	6	0	2	1	1	3	0	0	1	7	2	0	2	512	94.73
6	0	6	0	0	3	177	12	19	38	2	7	0	1	0	0	2	10	1	278	63.67
7	0	2	0	0	0	5	412	23	81	10	4	2	6	0	1	0	7	2	555	74.23
8	0	1	0	0	6	12	37	549	143	16	7	0	1	0	3	1	1	6	783	70.11
9	0	12	0	0	5	19	73	148	482	11	8	0	0	0	7	0	7	4	776	62.11
10	0	7	0	0	3	3	1	29	23	171	22	2	4	4	7	0	0	1	277	61.73
11	1	6	5	0	2	2	5	12	23	18	152	4	1	0	3	2	0	0	236	64.41
12	1	0	5	2	0	0	2	8	7	1	4	30	10	0	2	0	1	0	73	41.10
13	0	3	6	10	2	6	4	3	9	2	10	4	195	0	0	0	0	1	255	76.47
14	0	6	0	0	1	0	0	0	2	8	3	0	0	146	0	0	0	0	166	87.95
15	0	1	0	0	11	0	1	6	7	7	7	1	0	0	30	2	0	0	73	41.10
16	0	0	0	0	7	0	0	2	1	0	2	0	0	0	1	48	4	1	66	72.73
17	0	0	0	0	3	4	7	0	7	0	0	0	1	0	0	1	80	0	103	77.67
18	0	1	1	0	5	9	2	7	9	1	0	1	1	0	0	1	8	18	64	28.13
Sum	236	234	72	216	540	249	566	813	863	261	236	50	239	156	62	59	120	36	5011	
Uacc	97.88	79.06	75.00	92.13	89.81	71.08	72.79	67.53	55.85	65.52	64.41	60.00	81.59	93.59	48.39	81.36	66.67	50.00		**72.76**

Table 2. Confusion matrix derived from bootstrapping of the original data (LANDSAT) Rows and column labels are: 1Water; 2 Sparsely Vegetated Mixed Shrub; 3 Silver Sage Flat; 4 Riparian Floodplain; 5 Ponderosa Pine Woodland and Shrubland; 6 Mixed Grass Prairie, Burned; 7 Mixed Grass Prairie; 8 Medium Density Sagebrush Steppe; 9 Low Density Sagebrush Steppe; 10 Greasewood Sagebrush Steppe; 11 Greasewood Steppe; 12 Greasewood Flat; 13 Great Plains Riparian; 14 Great Plains Badland; 15 Foothill Woodland Steppe Transition; 16 Forested Burn; 17 Semi-Desert Grassland; 18 Silver Sage Steppe. Uacc: User's Accuracy; Pacc: Producer's Accuracy.

	Acres			Percent		
Class	Landsat 8	Spot	Landsat minus Spot	Landsat 8	Spot	Landsat minus Spot
Water	201807	221208	-19401	4.10	4.50	-0.39
Sparsely Vegetated Mixed Shrub	276005	223217	52788	5.61	4.54	1.07
Silver Sage Flat	13586	14084	-497	0.28	0.29	-0.01
Riparian Floodplain	19416	11794	7621	0.39	0.24	0.15
Mixed Grass Prairie, Burned	234068	237347	-3279	4.76	4.82	-0.07
Mixed Grass Prairie	901353	1042149	-140796	18.32	21.18	-2.86
Medium Density Sage Steppe	487667	351361	136305	9.91	7.14	2.77
Low Density Sage Steppe	963234	887246	75988	19.58	18.04	1.54
Greasewood Steppe and Greasewood Sage Steppe	410044	665240	-255195	8.34	13.52	-5.19
Greasewood Flat	27676	26928	748	0.56	0.55	0.02
Riparian and Wet Grass	192051	140910	51141	3.90	2.86	1.04
Great Plains Badlands	73987	66816	7172	1.50	1.36	0.15
Burn (Forested)	31123	10188	20936	0.63	0.21	0.43
Semi-Desert Grassland	120066	145317	-25251	2.44	2.95	-0.51
Silver Sage Steppe	81319	34409	46910	1.65	0.70	0.95
Agriculture: Cultivated and Hay/pasture	385878	386521	-643	7.84	7.86	-0.01
Roads, Development, Mines	36514	36639	-125	0.74	0.74	0.00
Cliff, Canyon and Massive Bedrock	429	421	7	0.01	0.01	0.00
Aspen Forest and Woodland	348	254	94	0.01	0.01	0.00
Dry-Mesic Montane Mixed Conifer Forest	529	511	18	0.01	0.01	0.00
Foothill Limber Pine - Juniper Woodland	1242	543	698	0.03	0.01	0.01
Lodgepole Pine Forest	36423	35870	553	0.74	0.73	0.01
Ponderosa Pine Woodland and Savanna	345083	370613	-25530	7.01	7.53	-0.52
Poor Site Lodgepole Pine Forest	373	370	3	0.01	0.01	0.00
Aspen and Mixed Conifer Forest	897	795	102	0.02	0.02	0.00
Montane-Foothill Deciduous Shrubland	3090	2948	142	0.06	0.06	0.00
Foothill Woodland Steppe Transition	73833	4274	69559	1.50	0.09	1.41
Lower Montane, Foothill, and Valley Grassland	1355	1319	35	0.03	0.03	0.00

Table 3. Comparison of the Landsat 8-based classification with the Spot-based classification, in terms of predicted acreage of 28 landcover classes. Negative numbers (in red) correspond to classes that have a greater extent on the Spot than on the Landsat raster.