Supplementary materials and results for RF MSPF

Haiyang Wang, Yuge Bi , Sen Wang, Wenye Xie, Yingshuo Du, and Zhanzhao Liu

In this supplement, we report the fixed development comparison and all archived source-grouped spatial ablations for Mapping Surface-Cover Components in Desert Steppe from UAV RGB Imagery Using Interpretable Features and Multiscale Superpixel Probability Fusion. We provide the full-precision CSV files and frozen analysis code alongside these tables.

Overview of the reported results

We found that regional probability aggregation improved classification performance in this image sample. On the fixed split, RF-MSPF achieved OA 0.880634, compared with 0.841168 for RF. In grouped validation, equal fusion achieved mean OA 0.870373 and RF-MSPF 0.869145; consistency weighting decreased OA in three of five folds. These comparisons do not establish external-site generalization or independent ecological replication.

Fixed method	OA	AA	Kappa	Macro F1	mIoU
RF	0.841168	0.815923	0.748780	0.804492	0.687875
KNN	0.830698	0.781707	0.728704	0.788139	0.668919
XGBoost	0.788441	0.755017	0.664989	0.747530	0.607200
RF-MSPF	0.880634	0.877852	0.810314	0.874204	0.783972

RF-MSPF class	Reference pixels	Precision	Recall	F1	IoU
HV	39265	0.858862	0.939488	0.897368	0.813842
WV	34829	0.916402	0.810445	0.860173	0.754652
LT	5975	0.735384	0.804184	0.768247	0.623702
BS	7703	0.985170	0.957289	0.971030	0.943691

We used 15 training and 4 validation patches in the fixed split, with three source-image groups shared between the subsets. In grouped validation, we kept all patches from each source image together; different photographs may nevertheless depict overlapping or adjacent ground. We report arithmetic fold means and sample SD (ddof = 1). The archived benchmark recorded 0.794583 ± 0.045674 s across 12 runs, excluding decoding and model loading; this is not a newly measured deployment benchmark.

Table S1 Original images and acquisition metadata

We use Used to indicate that an image contributes patches to the frozen analysis and Unused to indicate that it contributes no such patches; the latter does not mean that we never consulted the image. We acquired the images on 4–5 July 2025 and report local acquisition times in Beijing time (UTC+8); EXIF itself contains no time-zone field. The coordinates are recorded camera positions, not surveyed plot boundaries.

Source JPEG	Date and time 2025	Status	Patch numbers	Latitude °N	Longitude °E	Relative altitude m
DJI_0744.JPG	July 04 11:30:14	Unused		41.789334	111.887819	6.7
DJI_0746.JPG	July 04 11:31:42	Used	008, 009, 010	41.789689	111.887886	6.9
DJI_0749.JPG	July 04 11:33:06	Used	006, 007	41.790015	111.887943	7.1
DJI_0750.JPG	July 04 11:33:51	Used	004, 005	41.790245	111.887951	6.9
DJI_0753.JPG	July 04 11:34:57	Used	003	41.790601	111.887808	6.7
DJI_0756.JPG	July 04 11:35:45	Used	002	41.790756	111.887801	7.2
DJI_0757.JPG	July 04 11:36:24	Used	016, 017, 018, 019	41.790945	111.887879	7.2
DJI_0807.JPG	July 05 15:27:48	Used	012, 014, 015	41.790211	111.887733	7.2
DJI_0812.JPG	July 05 15:28:53	Used	001	41.789862	111.887661	6.8
DJI_0814.JPG	July 05 15:29:21	Unused		41.789689	111.887671	7.0
DJI_0816.JPG	July 05 15:29:57	Used	011, 013	41.789542	111.887601	6.9
DJI_0819.JPG	July 05 15:30:52	Unused		41.789337	111.887598	8.7

All source images contain 4000 × 3000 pixels, and their metadata identify DJI camera model FC300S and focal length 3.61 mm.

Table S2 All feature combinations on fixed validation

We evaluated every nonempty subset of the five features using RF and complete multiscale consistency fusion on the same development split. We rank the rows by OA to describe the results, not as an independent feature-selection procedure. The internal names GR-NLT and Gabor correspond to the historical MSAVI and Texture variables; we retained their original calculations.

Feature combination	Number	OA	AA	Kappa	Macro F1	mIoU
NGRDI + H + SP300_Grad + GR-NLT + Gabor	5	0.880634	0.877852	0.810314	0.874204	0.783972
H + SP300_Grad + Gabor	3	0.861926	0.814259	0.781174	0.806178	0.689033
NGRDI + H + SP300_Grad + GR-NLT	4	0.858998	0.867825	0.776283	0.862050	0.765455
H + SP300_Grad + GR-NLT + Gabor	4	0.858064	0.849471	0.774624	0.841636	0.733955
H + GR-NLT + Gabor	3	0.852983	0.850134	0.766763	0.842276	0.735724
NGRDI + H + GR-NLT + Gabor	4	0.850510	0.830591	0.762229	0.835879	0.725402
H + Gabor	2	0.848380	0.812578	0.760713	0.800773	0.676976
NGRDI + H + Gabor	3	0.841783	0.814235	0.749896	0.801370	0.678885
NGRDI + H + SP300_Grad + Gabor	4	0.828818	0.806504	0.728304	0.794526	0.669124
H	1	0.820581	0.794054	0.716414	0.783726	0.651171
H + GR-NLT	2	0.820296	0.844829	0.715656	0.831834	0.716975
NGRDI + H	2	0.813209	0.797799	0.704906	0.786834	0.655662
NGRDI + H + GR-NLT	3	0.808891	0.829951	0.698609	0.821067	0.701021
H + SP300_Grad	2	0.807672	0.781231	0.695747	0.767944	0.633708
H + SP300_Grad + GR-NLT	3	0.799161	0.811643	0.680358	0.799842	0.675511
NGRDI + H + SP300_Grad	3	0.759855	0.758072	0.618142	0.738909	0.597040
NGRDI + SP300_Grad + GR-NLT + Gabor	4	0.698332	0.665646	0.538971	0.661093	0.559583
NGRDI + SP300_Grad + GR-NLT	3	0.661771	0.644500	0.489946	0.639116	0.536234
SP300_Grad + GR-NLT + Gabor	3	0.652452	0.639257	0.473673	0.603404	0.470436
SP300_Grad + GR-NLT	2	0.607187	0.604675	0.405302	0.574974	0.443166
SP300_Grad + Gabor	2	0.540799	0.485182	0.309220	0.452707	0.315021
NGRDI + SP300_Grad + Gabor	3	0.540138	0.570034	0.324072	0.504632	0.360264
GR-NLT + Gabor	2	0.536697	0.498964	0.306623	0.474920	0.334615
NGRDI + GR-NLT + Gabor	3	0.534009	0.451623	0.297247	0.465065	0.324768

Feature combination	Number	OA	AA	Kappa	Macro F1	mIoU
NGRDI + GR-NLT	2	0.530784	0.474765	0.306267	0.473787	0.340253
NGRDI + SP300_Grad	2	0.500251	0.539030	0.277228	0.453207	0.309291
GR-NLT	1	0.458962	0.541307	0.206711	0.449610	0.313832
NGRDI + Gabor	2	0.428964	0.390006	0.192270	0.357411	0.232262
NGRDI	1	0.417901	0.457287	0.178768	0.357950	0.229223
SP300_Grad	1	0.414620	0.397511	0.124467	0.355138	0.224040
Gabor	1	0.296222	0.217483	-0.000730	0.183789	0.117306

Table S3a Patch provenance and validation membership

We list all 19 patches in Table S3a. Crop coordinates are zero-based columns and rows in the source JPEG, and each crop is 512 × 512 pixels. We verified that the archived crops reproduce the original crop pixels exactly. The fold number identifies the validation fold; we use all other folds for training.

Patch	Source image	Fixed split	Fold	Crop x	Crop y	Valid pixels
patch_001	DJI_0812.JPG	train	5	1925	1261	21455
patch_002	DJI_0756.JPG	train	2	1753	2288	29418
patch_003	DJI_0753.JPG	validation	3	1705	451	19437
patch_004	DJI_0750.JPG	validation	5	1496	1111	24957
patch_005	DJI_0750.JPG	train	5	2036	428	20869
patch_006	DJI_0749.JPG	train	4	1950	1414	26500
patch_007	DJI_0749.JPG	train	4	2034	358	34516
patch_008	DJI_0746.JPG	validation	3	1832	1616	18126
patch_009	DJI_0746.JPG	train	3	2934	1270	14129
patch_010	DJI_0746.JPG	train	3	1934	2226	27628
patch_011	DJI_0816.JPG	train	4	1568	1315	27651
patch_012	DJI_0807.JPG	train	2	1770	1182	9392
patch_013	DJI_0816.JPG	train	4	1967	720	18539
patch_014	DJI_0807.JPG	train	2	1360	1423	4318
patch_015	DJI_0807.JPG	train	2	2624	1375	2986
patch_016	DJI_0757.JPG	validation	1	993	40	25252
patch_017	DJI_0757.JPG	train	1	1593	2442	6907
patch_018	DJI_0757.JPG	train	1	256	2361	1503
patch_019	DJI_0757.JPG	train	1	221	1340	2092

Across all patches, we used 335,675 labeled positions (6.74% of the patch grids): HV 113,855; WV 138,365; LT 29,023; and BS 54,432. These pixels are spatially dependent. We use −1 for Ignore in the signed numeric masks and exclude these positions from training and evaluation.

Table S3b Fold counts and spatial dependence

Fold	Training patches	Validation patches	Training valid pixels	Augmented training samples	Validation valid pixels	Validation source images
1	15	4	299921	1199684	35754	DJI_0757.JPG
2	15	4	289561	1158244	46114	DJI_0756.JPG; DJI_0807.JPG
3	15	4	256355	1025420	79320	DJI_0746.JPG; DJI_0753.JPG
4	15	4	228469	913876	107206	DJI_0749.JPG; DJI_0816.JPG
5	16	3	268394	1073576	67281	DJI_0750.JPG; DJI_0812.JPG

The grouped validation folds contain 35,754, 46,114, 79,320, 107,206, and 67,281 reference pixels. We give each fold equal weight when calculating fold averages; these averages are not equivalent to metrics calculated by pooling all out-of-fold pixels.

Table S3c Shared sources in the fixed development split

Source image	Fixed training patches	Fixed validation patches
DJI_0746.JPG	patch_009, patch_010	patch_008
DJI_0750.JPG	patch_005	patch_004
DJI_0757.JPG	patch_017, patch_018, patch_019	patch_016

Table S3d Same-source crop overlap

Source image	Patch pair	Overlapping source pixels	Fixed membership	Grouped fold
DJI_0807.JPG	patch_012 / patch_014	27,642 (10.54% of one patch)	Both training	2

We found no same-source crop intersection crossing the fixed train–validation split and kept the overlapping pair in one grouped fold. This check cannot exclude ground overlap or spatial dependence between different photographs.

Table S4a Complete grouped spatial ablations

We report all six variants for every fold in Table S4a, evaluating each from that fold's fitted RF probabilities and including all five metrics. RF-MSPF corresponds to the internal code name Proposed. We did not retune any weighting or scale parameter after inspecting these results.

Fold	Spatial method	OA	AA	Kappa	Macro F1	mIoU
1	RF	0.810427	0.900738	0.702811	0.842473	0.746291
1	S150	0.889075	0.939746	0.816634	0.890624	0.809671
1	S300	0.808525	0.906987	0.701394	0.832934	0.728541
1	S500	0.827180	0.919451	0.727637	0.853192	0.760225
1	MS_equal	0.854170	0.929592	0.764976	0.870731	0.782863
1	RF-MSPF	0.849024	0.927928	0.757578	0.869580	0.782626
2	RF	0.887171	0.807991	0.724453	0.784491	0.668900
2	S150	0.904129	0.791546	0.756667	0.778720	0.657991
2	S300	0.922366	0.835587	0.801301	0.831648	0.727562
2	S500	0.924817	0.837156	0.807411	0.835064	0.732649
2	MS_equal	0.929848	0.838581	0.818467	0.842283	0.743164
2	RF-MSPF	0.928156	0.838034	0.814652	0.840233	0.740216
3	RF	0.739259	0.840560	0.458905	0.695526	0.571163
3	S150	0.852610	0.871606	0.634076	0.755914	0.628111
3	S300	0.844012	0.853606	0.613092	0.734427	0.605605
3	S500	0.844957	0.900719	0.621761	0.762761	0.635378
3	MS_equal	0.851135	0.876666	0.631530	0.771221	0.648258
3	RF-MSPF	0.850517	0.876756	0.630499	0.771239	0.648247
4	RF	0.760256	0.771172	0.671823	0.748823	0.608935
4	S150	0.835075	0.844841	0.773494	0.825141	0.708573
4	S300	0.820579	0.827102	0.752609	0.809493	0.688886
4	S500	0.799256	0.803752	0.723262	0.786982	0.660252
4	MS_equal	0.836819	0.843726	0.775328	0.825198	0.711456
4	RF-MSPF	0.837155	0.843802	0.775920	0.825202	0.711333

Fold	Spatial method	OA	AA	Kappa	Macro F1	mIoU
5	RF	0.812830	0.771832	0.733807	0.764025	0.640388
5	S150	0.862235	0.847540	0.805484	0.831939	0.724718
5	S300	0.852068	0.826838	0.790442	0.814754	0.704556
5	S500	0.844012	0.812908	0.778916	0.801332	0.688205
5	MS_equal	0.879892	0.855324	0.829607	0.844067	0.746583
5	RF-MSPF	0.880873	0.855725	0.830941	0.844634	0.747698

Table S4b Grouped summaries and paired differences

In Table S4b, we report the arithmetic mean ± sample SD across the five folds. In this descriptive analysis, we do not treat the five partially overlapping training sets as independent experimental replicates.

Spatial method	OA	AA	Kappa	Macro F1	mIoU
RF	0.801988 ± 0.057274	0.818458 ± 0.054263	0.658360 ± 0.114017	0.767068 ± 0.053502	0.647135 ± 0.066299
S150	0.868625 ± 0.027846	0.859056 ± 0.053760	0.757271 ± 0.072954	0.816468 ± 0.052214	0.705813 ± 0.069778
S300	0.849510 ± 0.044333	0.850024 ± 0.033649	0.731767 ± 0.077004	0.804651 ± 0.040576	0.691030 ± 0.050563
S500	0.848044 ± 0.046732	0.854797 ± 0.052346	0.731797 ± 0.070950	0.807866 ± 0.036426	0.695342 ± 0.051185
MS_equal	0.870373 ± 0.036696	0.868778 ± 0.037017	0.763982 ± 0.078975	0.830700 ± 0.037025	0.726465 ± 0.050512
RF-MSPF	0.869145 ± 0.036726	0.868449 ± 0.036405	0.761918 ± 0.079106	0.830178 ± 0.036610	0.726024 ± 0.050349

In Table S4c, we report the paired OA differences for the weighting and scale-integration steps. Differences are accuracy fractions; multiply by 100 to express them in percentage points. We provide the full paired differences for all five metrics in the CSV file.

Fold	RF-MSPF minus MS_equal	MS_equal minus S150
1	-0.005146	-0.034905
2	-0.001691	+0.025719
3	-0.000618	-0.001475
4	+0.000336	+0.001744
5	+0.000981	+0.017657
Mean	-0.001228	+0.001748
Sample SD	0.002412	0.023354

We observed that consistency weighting reduced mean OA by 0.122775 percentage points and produced lower OA in three folds. Equal multiscale fusion exceeded S150 mean OA by 0.174805 percentage points. We therefore identify regional aggregation as the principal reproducible benefit; these results do not demonstrate a reliable benefit from weighting.

Table S5 SLIC iteration sensitivity

We varied the maximum iteration count of aggregation SLIC (10, 20, 30), keeping all other segmentation settings and feature-extraction SLIC fixed. We compared segmentations for 19 patches × three scales using adjusted Rand index (ARI). Classification reused the original RF/scaler bundle on the four fixed validation patches and evaluated the same 87,772 reference pixels. All six 10-iteration predictions matched the archived arrays. This controlled comparison isolates the response of spatial probability aggregation to the iteration budget; the primary configuration remains at 10 iterations.

Table S5a Segmentation agreement across iteration settings

Iterations compared	Mean ARI	Minimum ARI
10 vs 20	0.791569	0.733047
10 vs 30	0.742037	0.663523
20 vs 30	0.885975	0.814032

Each comparison includes 57 pairs of segmentations. ARI is independent of region-ID numbering; values below one indicate different partitions.

Table S5b Fixed-validation classification metrics

Iterations	Method	OA	AA	Kappa	Macro-F1	mIoU
10	MS_equal	0.880395	0.877750	0.810042	0.873218	0.782700
10	RF-MSPF	0.880634	0.877852	0.810314	0.874204	0.783972
20	MS_equal	0.891651	0.887309	0.828002	0.883054	0.797686
20	RF-MSPF	0.893930	0.888549	0.831399	0.887091	0.803290
30	MS_equal	0.891776	0.889689	0.828317	0.884188	0.799613
30	RF-MSPF	0.890592	0.884974	0.826332	0.880663	0.793837

The segmentation and classification results quantify sensitivity to the iteration budget. The weighted-versus-equal ranking changes at 30 iterations, showing how this setting affects the relative behavior of the fusion strategies. The principal manuscript comparisons retain 10 iterations. Full-precision CSV files, realized region counts, per-patch comparisons, confusion counts, environment details, and source hashes accompany this table.