Long-Term Dynamics of Alpine Grassland Ecosystem Resilience in Northwestern Sichuan with Stagewise Shifts and Nonlinear Driver Responses

Haijun Yi and Cuicui Jiao *

School of Economics, Sichuan University of Science & Engineering, Yibin 644000, PR China; 325095114134@stu.suse.edu.cn (H.Y.); jiaocuicui@suse.edu.cn (C.J.)
* Correspondence: jiaocuicui@suse.edu.cn (C.J.); Tel.: +86-150-1158-3530

Software and Computational Environment

Data preprocessing, RI calculation, and spatial heterogeneity analyses were conducted in R 4.5.2. Raster and vector data import, cropping, masking, pixel-wise calculations, and true pixel-area calculations were implemented using terra (v1.8-93). Linear regression, statistical tests, and file handling used the base R packages stats, tools, and utils, respectively. The Pettitt change-point test and segmented regression were implemented using custom functions in the analysis scripts; the trend and segmented packages were not directly used. The endpoint-rule sensitivity and surrogate-null analyses reported in Table S13 were also implemented in R using custom functions and terra. Temporal autocorrelation was examined at lags 1–6. Regional trends were checked using Newey–West HAC standard errors, and pixelwise linear-slope sensitivity was assessed with HAC tests and BH/BY false-discovery-rate adjustments using custom R routines and stats::p.adjust (Table S12). These sensitivity checks did not confirm the significance of individual turning points.

Machine-learning modeling, OOF-SHAP, ALE, and their sensitivity analyses were performed in Python 3.11.15 using NumPy 2.4.6, pandas 3.0.3, scikit-learn 1.9.0, SciPy 1.17.1, Matplotlib 3.10.8, XGBoost 3.2.0, LightGBM 4.6.0, Optuna 4.9.0, SHAP 0.51.0, openpyxl 3.1.5, and joblib 1.5.3.

Table S1. Collinearity diagnostics, distributional characteristics, and correlation screening of candidate ecological drivers.

(A) Collinearity diagnostics and screening results

Variable	Final status	Collinearity diagnostics			Selection rationale
		VIF	Maximum |r|	Paired variable	
TEM	Removed	69.61	0.93	TMN	Iterative VIF elimination: the largest VIF in this iteration and ≥ 10; highly correlated with TMN (|r| = 0.93). Pearson correlation was used as a supplementary diagnostic.
PET	Removed	36.27	0.94	AET	Iterative VIF elimination: the largest VIF in this iteration and ≥ 10; highly correlated with AET (|r| = 0.94). Pearson correlation was used as a supplementary diagnostic.
AI	Removed	20.77	0.74	VPD	Iterative VIF elimination: the largest VIF in this iteration and ≥ 10. Pearson correlation was used as a supplementary collinearity diagnostic.

Variable	Final status	Collinearity diagnostics			Selection rationale
		VIF	Maximum |r|	Paired variable	
AET	Removed	7.56	0.94	PET	VIF < 10; removed during ecological-representativeness screening to reduce information overlap with water- and energy-process variables.
VPD	Retained	5.42	0.83	PET	VIF < 10; represents atmospheric moisture demand.
TMX	Retained	5.37	0.88	PET	VIF < 10; represents maximum temperature and potential heat stress.
SR	Retained	5.02	0.81	PRE	VIF < 10; represents solar radiation and energy input. Although strongly correlated with PRE (|r| = 0.81), the two variables represent different ecological processes.
PRE	Retained	3.90	0.81	SR	VIF < 10; represents precipitation input. Although strongly correlated with SR (|r| = 0.81), the two variables represent different ecological processes.
TMN	Retained	3.89	0.93	TEM	VIF < 10; represents minimum temperature and low-temperature limitation. Its correlation with TMX was 0.61.
SM	Retained	2.75	0.73	PRE	VIF < 10; represents soil-water availability.
RHU	Removed	1.45	0.47	VPD	VIF < 10; removed during ecological-representativeness screening to avoid information overlap with atmospheric-moisture indicators.
GI	Retained	1.27	0.27	TMN	VIF < 10; directly represents grazing pressure on pastoral grasslands.
NTL	Removed	1.10	0.26	TEM	VIF < 10; removed during ecological-representativeness screening because it is less process-specific than GI for the dominant anthropogenic disturbance in pastoral areas.

Note: VIF was evaluated iteratively. When any variable had VIF ≥ 10, the variable with the largest VIF was removed and VIFs were recalculated until all remaining variables had VIF < 10. "Final status" denotes the screening outcome for each candidate variable. The maximum |r| and corresponding variable were derived from the Pearson correlation matrix of the original 13 candidate variables and were used only as supplementary diagnostics, not as an automatic elimination rule.

(B) Pearson correlation matrix of the 13 candidate ecological drivers

Variable	SM	AET	PET	PRE	TEM	TMN	TMX	RHU	SR	VPD	AI	GI	NTL
SM	1.00	0.04	-0.14	0.73	0.17	0.32	-0.07	0.09	-0.70	-0.40	-0.53	0.07	0.04
AET	0.04	1.00	0.94	0.22	0.88	0.73	0.87	-0.38	-0.21	0.73	0.56	-0.13	0.22
PET	-0.14	0.94	1.00	0.08	0.90	0.74	0.88	-0.41	-0.09	0.83	0.69	-0.17	0.23
PRE	0.73	0.22	0.08	1.00	0.35	0.48	0.11	0.05	-0.81	-0.23	-0.57	0.06	0.14
TEM	0.17	0.88	0.90	0.35	1.00	0.93	0.84	-0.37	-0.42	0.57	0.51	-0.24	0.26
TMN	0.32	0.73	0.74	0.48	0.93	1.00	0.61	-0.37	-0.53	0.40	0.30	-0.27	0.22

Variable	SM	AET	PET	PRE	TEM	TMN	TMX	RHU	SR	VPD	AI	GI	NTL
TMX	-0.07	0.87	0.88	0.11	0.84	0.61	1.00	-0.29	-0.19	0.69	0.68	-0.09	0.26
RHU	0.09	-0.38	-0.41	0.05	-0.37	-0.37	-0.29	1.00	-0.03	-0.47	-0.33	-0.00	-0.04
SR	-0.70	-0.21	-0.09	-0.81	-0.42	-0.53	-0.19	-0.03	1.00	0.31	0.39	0.06	-0.15
VPD	-0.40	0.73	0.83	-0.23	0.57	0.40	0.69	-0.47	0.31	1.00	0.74	-0.03	0.15
AI	-0.53	0.56	0.69	-0.57	0.51	0.30	0.68	-0.33	0.39	0.74	1.00	-0.16	0.08
GI	0.07	-0.13	-0.17	0.06	-0.24	-0.27	-0.09	-0.00	0.06	-0.03	-0.16	1.00	-0.07
NTL	0.04	0.22	0.23	0.14	0.26	0.22	0.26	-0.04	-0.15	0.15	0.08	-0.07	1.00

Note: Values are Pearson correlation coefficients (r) and are reported to two decimal places.

(C) Distributional diagnostics of the 13 candidate ecological drivers

Variable	N	Skewness	Excess kurtosis	Normality test P value
SM	24,084	-0.658	0.324	<0.001
AET	24,084	-0.138	-0.041	<0.001
PET	24,084	0.613	1.002	<0.001
PRE	24,084	0.074	-0.115	<0.001
TEM	24,084	0.230	1.712	<0.001
TMN	24,084	-0.341	1.245	<0.001
TMX	24,084	0.609	1.258	<0.001
RHU	24,084	-0.097	-0.253	<0.001
SR	24,084	-0.539	-0.137	<0.001
VPD	24,084	0.462	0.553	<0.001
AI	24,084	0.986	2.249	<0.001
GI	24,084	1.556	5.039	<0.001
NTL	24,084	9.607	114.014	<0.001

Note: Distributional diagnostics were calculated from 24,084 grid-year observations covering 669 complete grids (1985–2020). Skewness and excess kurtosis describe each variable's marginal distribution. Nominal P values were obtained using the D'Agostino–Pearson omnibus normality test (all $P < 0.001$). Because spatial and temporal dependence may be present, the P values are interpreted descriptively.

(D) Spearman rank correlation matrix of the 13 candidate ecological drivers

Variable	SM	AET	PET	PRE	TEM	TMN	TMX	RHU	SR	VPD	AI	GI	NTL
SM	1.00	-0.14	-0.24	0.70	-0.04	0.12	-0.17	0.17	-0.65	-0.47	-0.59	0.07	0.15
AET	-0.14	1.00	0.98	0.13	0.89	0.66	0.88	-0.41	-0.22	0.77	0.60	-0.16	0.32

Variable	SM	AET	PET	PRE	TEM	TMN	TMX	RHU	SR	VPD	AI	GI	NTL
PET	-0.24	0.98	1.00	0.04	0.90	0.68	0.87	-0.44	-0.13	0.82	0.67	-0.18	0.28
PRE	0.70	0.13	0.04	1.00	0.26	0.38	0.08	-0.01	-0.84	-0.25	-0.60	0.08	0.24
TEM	-0.04	0.89	0.90	0.26	1.00	0.88	0.79	-0.44	-0.41	0.59	0.49	-0.28	0.32
TMN	0.12	0.66	0.68	0.38	0.88	1.00	0.50	-0.41	-0.51	0.35	0.24	-0.36	0.29
TMX	-0.17	0.88	0.87	0.08	0.79	0.50	1.00	-0.34	-0.18	0.71	0.66	-0.09	0.32
RHU	0.17	-0.41	-0.44	-0.01	-0.44	-0.41	-0.34	1.00	0.10	-0.47	-0.33	-0.03	-0.24
SR	-0.65	-0.22	-0.13	-0.84	-0.41	-0.51	-0.18	0.10	1.00	0.24	0.40	0.04	-0.29
VPD	-0.47	0.77	0.82	-0.25	0.59	0.35	0.71	-0.47	0.24	1.00	0.72	-0.01	0.19
AI	-0.59	0.60	0.67	-0.60	0.49	0.24	0.66	-0.33	0.40	0.72	1.00	-0.19	0.03
GI	0.07	-0.16	-0.18	0.08	-0.28	-0.36	-0.09	-0.03	0.04	-0.01	-0.19	1.00	-0.03
NTL	0.15	0.32	0.28	0.24	0.32	0.29	0.32	-0.24	-0.29	0.19	0.03	-0.03	1.00

Note: Values are Spearman rank correlation coefficients (ϱ), calculated from the same 24,084 grid-year observations as Table S1(C) and reported to two decimal places. Spearman correlations were used to assess the robustness of the Pearson correlation patterns and were not used as an automatic variable-elimination rule.

Table S2. Hyperparameter optimization settings and search spaces for the five candidate models.

Model	Hyperparameter	Search space	Sampling strategy	Trials per outer task
OLS	Not tuned	—	Baseline	0
SVR	C	0.01–30.0	Log-uniform	10
SVR	epsilon	0.001–0.15	Log-uniform	10
SVR	gamma (RBF)	0.0001–0.5	Log-uniform	10
RF	n_estimators	400–1400	Integer (step 100)	20
RF	max_depth	5–24	Integer	20
RF	min_samples_split	2–24	Integer	20
RF	min_samples_leaf	2–16	Integer	20
RF	max_features	sqrt/log2/0.6/0.8/1.0	Categorical	20
LightGBM	n_estimators	500–2400	Integer (step 100)	25
LightGBM	learning_rate	0.005–0.06	Log-uniform	25
LightGBM	max_depth	4–11	Integer	25

Model	Hyperparameter	Search space	Sampling strategy	Trials per outer task
LightGBM	num_leaves	15–127	Integer	25
LightGBM	min_child_samples	10–120	Integer	25
LightGBM	subsample	0.6–0.95	Uniform	25
LightGBM	colsample_bytree	0.55–0.95	Uniform	25
LightGBM	reg_lambda	0.1–50.0	Log-uniform	25
LightGBM	reg_alpha	1e-06–10.0	Log-uniform	25
XGBoost	learning_rate	0.005–0.06	Log-uniform	35
XGBoost	max_depth	3–9	Integer	35
XGBoost	min_child_weight	5–50	Integer	35
XGBoost	subsample	0.6–0.95	Uniform	35
XGBoost	colsample_bytree	0.55–0.95	Uniform	35
XGBoost	colsample_bynode	0.55–1.0	Uniform	35
XGBoost	gamma (min_split_loss)	1e-05–2.0	Log-uniform	35
XGBoost	reg_lambda	0.5–100.0	Log-uniform	35
XGBoost	reg_alpha	1e-06–10.0	Log-uniform	35
XGBoost	max_bin	128–512	Integer (step 64)	35
XGBoost	grow_policy	depthwise/lossguide	Categorical	35
XGBoost	max_leaves	16–256	Log-integer (lossguide only)	35
XGBoost	n_estimators cap	5000	Fixed upper bound	—
XGBoost	Early-stopping rounds	150	Fixed	—

Note: OLS was an untuned linear baseline. SVR, RF, LightGBM, and XGBoost were tuned with Optuna-TPE within each outer task to maximize the mean R^2 across the five inner validation folds. The numbers of trials per outer task were 10, 20, 25, and 35 for SVR, RF, LightGBM, and XGBoost, respectively. For XGBoost, the n_estimators upper bound was fixed at 5000 and early_stopping_rounds at 150. Early stopping used only inner validation folds; outer test folds were not used for tuning, early stopping, or model selection.

Table S3. Ranges of optimal hyperparameters obtained for the four tuned models across the 25 outer tasks.

Hyperparameter	SVR	RF	LightGBM	XGBoost
C	0.01238–26.94	—	—	—
epsilon	0.001101–0.1187	—	—	—
gamma (RBF)	0.002135–0.4557	—	—	—
gamma (min_split_loss)	—	—	—	1.058e-05–0.007108
n_estimators	—	400–1400	800–2400	—
final_n_estimators	—	—	—	250–2599
max_depth	—	15–24	8–11	8–9
min_samples_split	—	2–23	—	—
min_samples_leaf	—	2–9	—	—
max_features	—	0.8–1	—	—
learning_rate	—	—	0.005014–0.01594	0.005593–0.0435
num_leaves	—	—	40–123	—
min_child_samples	—	—	10–81	—
min_child_weight	—	—	—	5–38
subsample	—	—	0.6428–0.9055	0.6391–0.9476
colsample_bytree	—	—	0.5881–0.9432	0.6753–0.9492
colsample_bynode	—	—	—	0.593–0.9995
reg_lambda	—	—	0.1005–22.29	0.6354–40.27

Hyperparameter	SVR	RF	LightGBM	XGBoost
reg_alpha	—	—	1.676e-06–0.448	1.024e-06–0.2078
grow_policy	—	—	—	depthwise; lossguide
max_leaves	—	—	—	130–252
max_bin	—	—	—	128–512

Note: Values are the minimum–maximum ranges of the optimal hyperparameters across the 25 outer tasks; "—" indicates that a parameter was not applicable to the corresponding model. gamma (RBF) in SVR and gamma (min_split_loss) in XGBoost are distinct hyperparameters. For XGBoost, final_n_estimators denotes the final number of boosting rounds determined by early stopping on inner validation folds with n_estimators capped at 5000 and early_stopping_rounds = 150; outer test folds were not used for early stopping.

Table S4. Paired comparison of outer-test performance between RF and XGBoost.

(A) Paired outer-test R^2 results for RF and XGBoost

Outer task	RF test R^2	XGBoost test R^2	Difference (RF − XGBoost)
R1-F1	0.390	0.385	0.006
R1-F2	0.362	0.366	-0.004
R1-F3	0.381	0.372	0.009
R1-F4	0.378	0.360	0.018
R1-F5	0.322	0.297	0.025
R2-F1	0.368	0.358	0.009
R2-F2	0.350	0.342	0.008
R2-F3	0.346	0.340	0.005
R2-F4	0.356	0.343	0.013
R2-F5	0.377	0.372	0.005
R3-F1	0.366	0.356	0.010
R3-F2	0.379	0.360	0.019

Outer task	RF test R^2	XGBoost test R^2	Difference (RF − XGBoost)
R3-F3	0.355	0.335	0.020
R3-F4	0.341	0.324	0.017
R3-F5	0.351	0.345	0.006
R4-F1	0.375	0.361	0.014
R4-F2	0.362	0.341	0.021
R4-F3	0.346	0.332	0.014
R4-F4	0.351	0.340	0.011
R4-F5	0.372	0.360	0.013
R5-F1	0.373	0.374	-0.001
R5-F2	0.397	0.382	0.015
R5-F3	0.364	0.355	0.009
R5-F4	0.343	0.340	0.003
R5-F5	0.362	0.349	0.013

(B) Paired statistical tests of outer-test R^2 for RF and XGBoost

Test	Statistic	Two-sided P value	Role
Corrected repeated k-fold t-test	$t(24) = 2.992$	0.00633	Primary test
Wilcoxon signed-rank test	$W = 4.0$	4.17×10^{-7}	Supplementary test

Note: The comparison comprised 25 paired outer tasks (5 repeats × 5 folds). The mean difference in test R^2 (RF − XGBoost) was 0.011, with a standard deviation of 0.007.

Table S5. Consistency of OOF-SHAP feature-importance rankings for the RF model.

Comparison scope	Number of rankings	Kendall's W
All 25 outer models	25	0.875
Repeat 1 (5 folds)	5	0.857
Repeat 2 (5 folds)	5	0.946
Repeat 3 (5 folds)	5	0.831
Repeat 4 (5 folds)	5	0.886
Repeat 5 (5 folds)	5	0.891

Note: Rankings were based on OOF-SHAP importance for the seven ecological drivers. Kendall's W ranges from 0 (no agreement) to 1 (complete agreement).

Table S6. Full-model mean absolute OOF-SHAP importance and relative contributions of eight predictors, including Year.

Predictor	Predictor type	Mean absolute OOF-SHAP	Full-model share (%)
Year	Temporal covariate	0.03262924	48.68
TMX	Ecological driver	0.00695696	10.38
SR	Ecological driver	0.00559619	8.35
TMN	Ecological driver	0.00531992	7.94
SM	Ecological driver	0.00456558	6.81
PRE	Ecological driver	0.00441055	6.58
GI	Ecological driver	0.00430820	6.43
VPD	Ecological driver	0.00323769	4.83
Total	**All eight predictors**	**0.06702434**	**100.00**

Note: Mean absolute OOF-SHAP values were averaged over five repetitions for 24,084 grid-year observations. Full-model shares use all eight predictors as the denominator. Figure 6a and Table S9(B) instead use seven ecological drivers. SHAP shares indicate relative prediction importance, not explained variance or causal effects.

Table S7. Major ALE response features and cross-model consistency of key ecological drivers.

Driver	Response feature	Median position [Q1, Q3]	Consistency proportion
GI (SU/ha)	Zero crossing; negative to positive	0.19 [0.16, 0.21]	20/25 (80%)
GI (SU/ha)	Local maximum	0.39[0.39, 0.44]	25/25 (100%)
PRE (mm)	Zero crossing; negative to positive	722.77 [717.67, 725.13]	22/25 (88%)

Driver	Response feature	Median position [Q1, Q3]	Consistency proportion
PRE (mm)	Local minimum	686.35 [686.35, 686.35]	24/25 (96%)
SM (mm)	Zero crossing; negative to positive	110.11 [30.06, 114.81]	16/25 (64%)
SR (kWh/m²)	Zero crossing; negative to positive	3732.50 [3716.40, 3743.80]	17/25 (68%)
SR (kWh/m²)	Local minimum	3645.60 [3615.00, 3653.30]	15/25 (60%)
TMN (°C)	Zero crossing; positive to negative	-10.85 [-11.39, -10.45]	19/25 (76%)
TMN (°C)	Local maximum	-12.55 [-17.23, -11.79]	19/25 (76%)
TMX (°C)	Zero crossing; positive to negative	17.47 [17.29, 17.74]	25/25 (100%)
TMX (°C)	Local maximum	17.00 [16.47, 18.55]	16/25 (64%)

Note: Only ALE response features detected in at least 60% of the 25 outer RF models are reported. Medians and quartiles were calculated among models detecting the corresponding feature. Q1 and Q3 denote the 25th and 75th percentiles. Their intervals describe cross-model dispersion, not formal confidence intervals or total uncertainty; identical quartiles do not imply absence of uncertainty. Two decimal places are retained for reporting consistency, not to imply equivalent ecological precision. Units: SM and PRE, mm; SR, kWh/m²; TMN and TMX, °C; GI, SU/ha.

Table S8. Pairwise TreeSHAP interaction strengths and rankings among the seven ecological drivers.

Rank	Interaction pair	Mean absolute interaction strength (×10⁻³)	SD (×10⁻³)
1	TMX × SR	1.46	0.21
2	PRE × SR	1.26	0.19
3	TMN × SR	1.22	0.18
4	SM × PRE	1.12	0.16
5	SM × SR	1.10	0.20
6	TMX × GI	1.10	0.17
7	SM × TMX	1.06	0.21

Rank	Interaction pair	Mean absolute interaction strength (×10⁻³)	SD (×10⁻³)
8	TMN × TMX	1.00	0.13
9	PRE × TMX	0.99	0.12
10	TMX × VPD	0.98	0.20
11	PRE × TMN	0.96	0.13
12	SM × TMN	0.87	0.12
13	SM × GI	0.74	0.11
14	PRE × GI	0.68	0.09
15	SR × GI	0.68	0.09
16	SR × VPD	0.65	0.13
17	TMN × VPD	0.63	0.16
18	SM × VPD	0.63	0.10
19	TMN × GI	0.62	0.08
20	PRE × VPD	0.54	0.10
21	VPD × GI	0.50	0.12

Note: Interaction strengths were calculated on the corresponding outer test fold for each of the 25 RF models and ranked by the mean absolute TreeSHAP interaction strength across models. The standard deviation represents variation among the 25 outer models. Year was included in model training but was excluded from pairwise interaction rankings among the seven ecological drivers. Mean absolute interaction strengths and standard deviations are expressed as ×10⁻³ and reported to two decimal places.

Table S9. Sensitivity analysis of excluding Year.

(A) Model performance and SHAP-ranking stability after excluding Year

Metric	Main model	No-Year model	Difference (No-Year – main)
Test R^2	0.363 ± 0.017	0.111 ± 0.016	-0.251

Metric	Main model	No-Year model	Difference (No-Year − main)
Test RMSE	0.074 ± 0.002	0.087 ± 0.002	0.013
Test MAE	0.058 ± 0.001	0.069 ± 0.001	0.011
Test MSE ($\times 10^{-3}$)	5.443 ± 0.264	7.589 ± 0.299	2.146
Spearman ranking ϱ (7 drivers)	—	−0.286	—
Top-4 overlap	—	2/4 (50%)	—

(B) Relative SHAP contributions of ecological drivers before and after excluding Year

Driver	Main model (%)	Rank	No-Year model (%)	Rank	Difference (percentage points)
SM	13.27	4	9.25	7	-4.02
PRE	12.82	5	16.63	2	+3.81
TMN	15.47	3	15.89	3	+0.42
TMX	20.23	1	14.35	4	-5.88
SR	16.27	2	14.28	5	-1.99
VPD	9.41	7	19.12	1	+9.71
GI	12.53	6	10.48	6	-2.04

Note: Performance metrics are mean ± standard deviation across the 25 outer tasks. Spearman ϱ and Top-4 overlap were used descriptively to assess feature-ranking consistency; P values were not reported because the number of ranked variables was small. In the main model, SHAP contributions were normalized to 100% across the seven ecological drivers, excluding Year from the denominator despite its inclusion in model fitting. In the No-Year model, all seven available predictors were used in the denominator. The full eight-predictor contributions, including Year, are reported in Table S6. Main-model Top-4: TMX, SR, TMN, and SM; No-Year-model Top-4: VPD, PRE, TMN, and TMX.

Table S10. Predictive performance of the RF model under 100-km spatial block cross-validation.

Metric	Main analysis	100-km spatial block CV	Difference (block CV − main)
Test R^2	0.363 ± 0.017	0.241 ± 0.049	-0.122
Test RMSE	0.074 ± 0.002	0.080 ± 0.004	+0.007
Test MAE	0.058 ± 0.001	0.063 ± 0.003	+0.005
Test MSE ($\times 10^{-3}$)	5.443 ± 0.264	6.487 ± 0.568	+1.044

Note: Values are mean ± standard deviation across the 25 outer tasks. Positive changes in error metrics indicate increased prediction error under 100-km spatial block cross-validation.

Table S11. Sensitivity analysis of temperature-variable reparameterization.

(A) Predictive performance and temperature-group contribution after temperature-variable reparameterization

Metric	Main model (TMX + TMN)	Reparameterized model (Tmean + DTR)	Difference (reparameterized − main)
Test R^2	0.363 ± 0.017	0.375 ± 0.019	+0.012
Test RMSE	0.074 ± 0.002	0.073 ± 0.002	-0.001
Test MAE	0.058 ± 0.001	0.057 ± 0.001	-0.001
Test MSE ($\times 10^{-3}$)	5.443 ± 0.264	5.337 ± 0.267	-0.106
Temperature-group contribution (%)	35.69	34.76	−0.93 percentage points
Spearman ranking ϱ (5 non-temperature drivers)	—	0.900	—
Top-4 overlap	—	4/4 (100%)	—

(B) SHAP contributions of non-temperature drivers before and after temperature-variable reparameterization

Driver	Main model (%)	Rank	Reparameterized model (%)	Rank	Difference (percentage points)
SM	13.27	2	13.32	2	+0.05
PRE	12.82	3	12.67	4	-0.16
SR	16.27	1	15.55	1	-0.72
VPD	9.41	5	10.89	5	+1.48
GI	12.53	4	12.81	3	+0.28

Note: Performance metrics are mean ± standard deviation across the 25 outer tasks. The temperature variables were reparameterized from TMX + TMN to Tmean + DTR. Spearman ϱ and Top-4 overlap were used descriptively to assess the ranking consistency of non-temperature drivers; P values were not reported because the number of ranked variables was small. Main-model Top-4 non-temperature drivers: SR, SM, PRE, and GI; reparameterized-model Top-4: SR, SM, GI, and PRE.

Table S12. Temporal autocorrelation diagnostics and sensitivity of regional and pixelwise RI linear-trend inference.

(A) Regional linear-trend significance (1985–2020)

Method	Slope (RI yr^{-1})	Two-sided p-value
OLS	-0.002419	0.000221
HAC (lag 4)	-0.002419	0.000052
HAC (lag 6)	-0.002419	0.000014

(B) Pixelwise temporal autocorrelation diagnostics

Lag (years)	Mean raw RI ϱ	Mean detrended residual ϱ	Estimable area (%)
1	0.571	0.501	100.00
2	0.297	0.194	99.91
3	0.180	0.067	99.83
4	0.042	-0.087	99.76
5	-0.088	-0.229	99.70
6	-0.065	-0.191	99.63

(C) Sensitivity of pixelwise linear-slope inference to HAC and FDR adjustments

Specification	Detected area (km²)	Share of common valid area (%)
OLS, nominal p (matched)	58,968.35	41.46
OLS + BH-FDR	44,104.98	31.01
HAC (lag 4), nominal p	46,349.25	32.58
HAC (lag 4) + BH-FDR	28,905.77	20.32
HAC (lag 6) + BH-FDR	32,603.60	22.92
HAC (lag 4) + BY-FDR	10,475.33	7.36

Note: (A) Regional RI was evaluated for 36 annual observations. HAC estimates used a Bartlett kernel, a finite-sample adjustment, and two-sided t inference; regional significance was also examined across lags 1–6. (B) Means are area-weighted among pixels with an estimable coefficient at each lag. Detrended values refer to residuals from pixelwise linear fits; missing calendar years were not collapsed. (C) Linear-slope analyses used 193,479 pixels with at least 25 valid annual RI observations (141,040.22 km²) within the common valid extent of 195,117 pixels (142,243.73 km²); the remaining 1,638 pixels were not included in formal linear-slope tests. All displayed percentages use the common valid extent as the denominator. BH and BY adjustments were performed across the corresponding testable pixelwise p-values at $\alpha = 0.05$. These comparisons assess the sensitivity of linear-slope inference and do not confirm individual turning points under autocorrelation-adjusted, FDR-controlled testing.

Table S13. Sensitivity of RI turning-point detection to endpoint restrictions and exploratory surrogate-null assessment.

(A) Endpoint-restriction sensitivity and spatial agreement with the five-year baseline

Metric	3-year rule	5-year (original)	7-year rule
Eligible turning years	1988–2017	1990–2015	1992–2013
Total turning-point area (km²)	26,871.02	27,463.64	26,522.31
Earliest eligible year (turning-point area, km²)	1988 (1,287.18)	1990 (3,961.41)	1992 (6,239.97)
1990–1995 share of all turning-point area (%)	38.10	46.85	44.16
1992–1995 share within 1992–2013 (%)	35.96	36.05	44.16
Five-year baseline turning-point area retained (%)	94.16	100.00	89.54
Same turning year within shared turning-point area (%)	91.93	100.00	81.57
Turning year within ±2 years in shared area (%)	99.85	100.00	99.57

(B) Exploratory surrogate-null assessment under the five-year rule

Metric	Result
Candidate spatial tiles with complete 1985–2020 RI records	5,003
Selected spatial tiles / representative pixels	350 / 350
Valid surrogate realizations	199 / 199
Observed 1990–1995 turning-point share in sampled data (%)	50.34
Surrogate median share (%)	33.21
Surrogate 2.5th–97.5th percentile range (%)	4.05–67.61
Exploratory one-sided upper-tail probability	0.21

Note: (A) All rules used the same RI series and valid grassland extent (142,243.73 km²). The 3-, 5-, and 7-year rules specify the endpoint exclusion. Because eligible years differ, the 1990–1995 shares are descriptive rather than directly comparable. The common-period share uses the turning-point area in 1992–1995 divided by that in 1992–2013. Baseline retention denotes the five-year turning-point area also identified under the alternative rule; year agreement refers only to shared turning-point areas. The first reporting period (1990–1995) covers six years, whereas later periods cover five. (B) A total of 350 spatial tiles were sampled from 5,003 eligible 8 × 8-cell tiles with complete RI records; one area-weighted representative pixel was selected per tile. Tile-level complete-case areas supplied weights. Each of 199 surrogate realizations combined fitted linear trends, AR(1)-type residuals, and synchronized innovation resampling. Null percentiles are an empirical reference, not confidence bounds; the upper-tail probability is exploratory, not a formal whole-area significance test.

Table S14. Sensitivity analysis of grazing intensity (GI) discontinuity and period-specific responses around 2001.

(A) Grid-level GI temporal discontinuity diagnostics

Metric	Result
Mean signed GI change during 2000–2001 (SU/ha)	+0.0357
Mean absolute GI change during 2000–2001 (SU/ha)	0.0993
Mean absolute change in adjacent transitions (SU/ha)	0.0980
Excess absolute change relative to adjacent transitions (SU/ha)	+0.0014
95% 100-km spatial-block bootstrap interval for excess change (SU/ha)	-0.0034 to 0.0065
Rank of 2000–2001 mean absolute change (35 annual transitions)	5/35
Number of paired 20-km grid cells	669

(B) Period-specific RF predictive performance and GI ALE features

Metric	1985–2000	2001–2020
Grid-year observations	10,704	13,380
Number of 20-km grid cells	669	669
Repeated outer test tasks	25	25
Outer-test R^2 (mean ± SD)	0.208 ± 0.017	0.425 ± 0.025
Outer-test RMSE (mean ± SD)	0.078 ± 0.002	0.070 ± 0.003
Outer-test MAE (mean ± SD)	0.062 ± 0.001	0.054 ± 0.002
Models detecting a local GI ALE maximum	25/25 (100%)	3/25 (12%)
Median GI at detected local maximum (SU/ha)	0.765	1.325
Models with declining GI ALE in upper 25% of common GI range	6/25 (24%)	24/25 (96%)
Median GI ALE change over upper 25% of common GI range	+0.00070	-0.00477

Note: (A) Annual changes were evaluated for 669 paired 20-km grid cells. Adjacent transitions refer to 1999–2000 and 2001–2002. The 95% interval was obtained from 1,000 bootstrap resamples of 39 spatial blocks (100 km). (B) The RF model included SM, PRE, TMN, TMX, SR, VPD, GI, and Year in each period; hyperparameters were retuned within five repetitions of outer five-fold and inner five-fold grouped cross-validation by grid identifier. The shared GI range was 0.346–2.210 SU/ha (intersection of the two periods' 5th–95th percentile ranges), and ALE curves were shifted to zero at the shared reference value of 0.958 SU/ha. The high-GI diagnostic is the net ALE change over the upper 25% of this range. The later-period median peak is based on only three detections. ALE peak locations are descriptive model features, not optimal grazing levels or causal thresholds. The shaded envelopes in Figure S3 are empirical 2.5th–97.5th percentile intervals across 25 outer models, not formal confidence intervals. Period differences cannot be attributed solely to GI source-resolution changes.

Table S15. Robustness of grouped ecological-driver importance under alternative grouping schemes.

(A) Original grouping and predictor-group importance

Driver group	Predictors	Grouped OOF-SHAP share (%)	GPFI ΔR^2 (mean ± SD)	Group Shapley share (%)
Temperature–energy	TMX, TMN, SR	51.96	0.2268 ± 0.0197	48.18
Moisture	SM, PRE, VPD	35.51	0.1203 ± 0.0114	34.20
Grazing	GI	12.53	0.0335 ± 0.0073	17.62

(B) Sensitivity to alternative grouping schemes

Grouping scenario	Compared groups (first / second)	OOF-SHAP shares (%)	GPFI ΔR^2 (first / second)	Paired GPFI difference (mean ± SD)
Baseline	Temperature–energy / Moisture	51.96 / 35.51	0.2268 / 0.1203	+0.1065 ± 0.0167
VPD separated	Temperature–energy / Moisture supply	51.96 / 26.10	0.2268 / 0.1034	+0.1233 ± 0.0157
VPD to temperature–energy	Temperature–energy + VPD / Moisture supply	61.38 / 26.10	0.2487 / 0.1034	+0.1452 ± 0.0156
SR separated	Temperature only / Moisture	35.69 / 35.51	0.1606 / 0.1203	+0.0403 ± 0.0125

Note: (A) Grouped OOF-SHAP shares sum the mean absolute contributions of individual predictors and are normalized across the seven ecological predictors, excluding Year. GPFI is the decrease in held-out R^2 after jointly permuting complete, year-aligned grid trajectories for all predictors in a group while retaining Year and other predictors unchanged. The same 25 outer RF models were reused without retuning, with 20 permutations per fold. GPFI estimates are mean ± SD across the 25 outer tasks. The sampled interventional group Shapley analysis treated Year as a separate player and used up to 80 held-out grid observations and 32 training-background grid observations per outer fold (667 distinct test grids); group shares shown in (A) were normalized across ecological groups only. (B) Paired GPFI differences are the first group minus the second group, with both group values shown in the preceding columns. All 25 outer folds favored the first group under each scenario; these overlapping repeated-CV folds do not represent independent significance tests. Without Year, the temperature–energy and moisture grouped OOF-SHAP shares were 44.51% and 45.01%, respectively (Table S9). Group sizes vary among alternative definitions, and neither GPFI nor SHAP measures causal effects; grouped OOF-SHAP aggregation and group Shapley are not directly interchangeable.

Figure S1. Comparison of ALE responses for six ecological drivers between the main and No-Year RF models.

Solid lines represent the main model including Year, whereas dashed lines represent the sensitivity model excluding Year: (a) TMX (°C); (b) SR (kWh/m²); (c) TMN (°C); (d) SM (mm); (e) PRE (mm); and (f) GI (SU/ha).

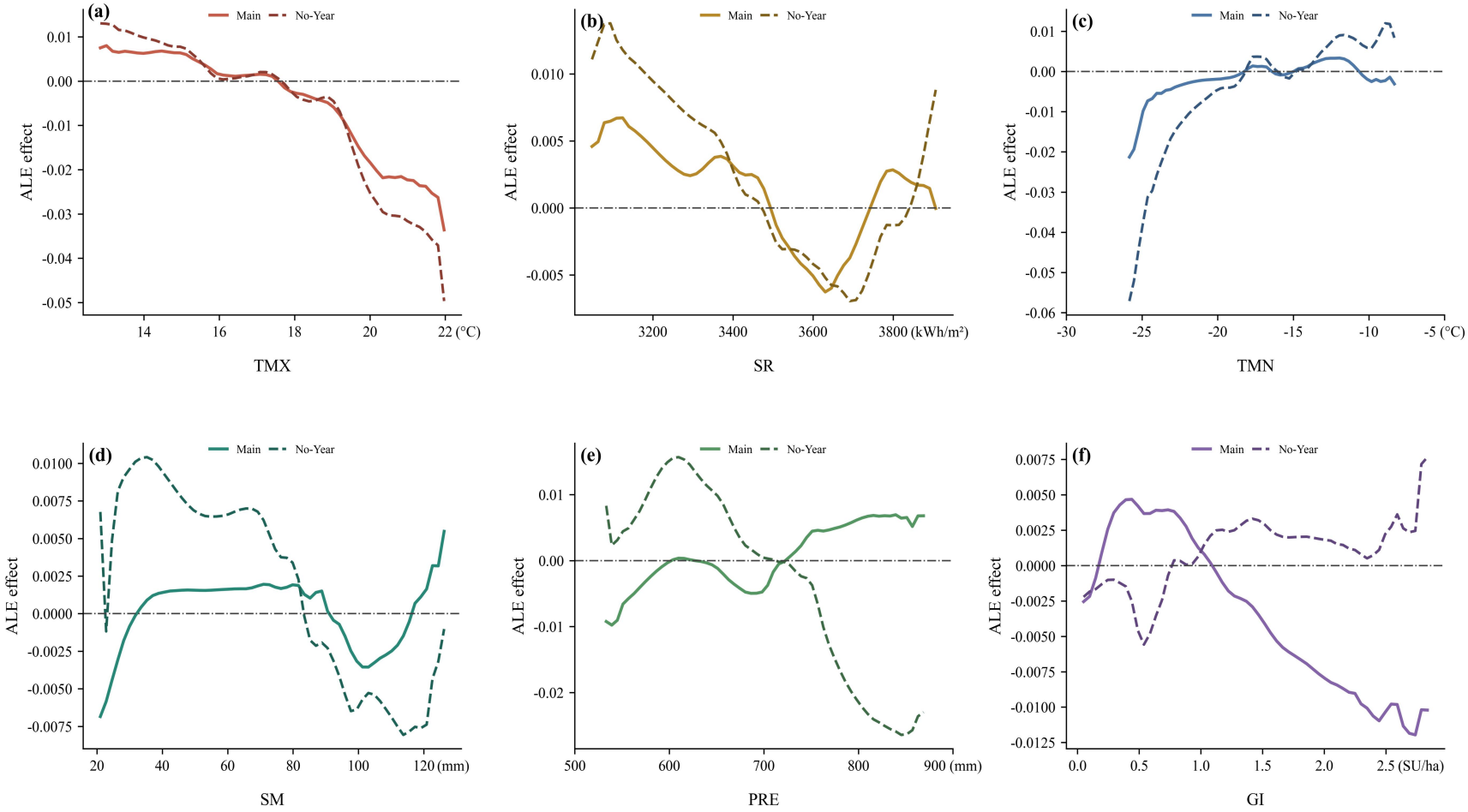


Figure S2. SHAP interaction effect between SR (kWh/m²) and SM (mm) in the RF model.

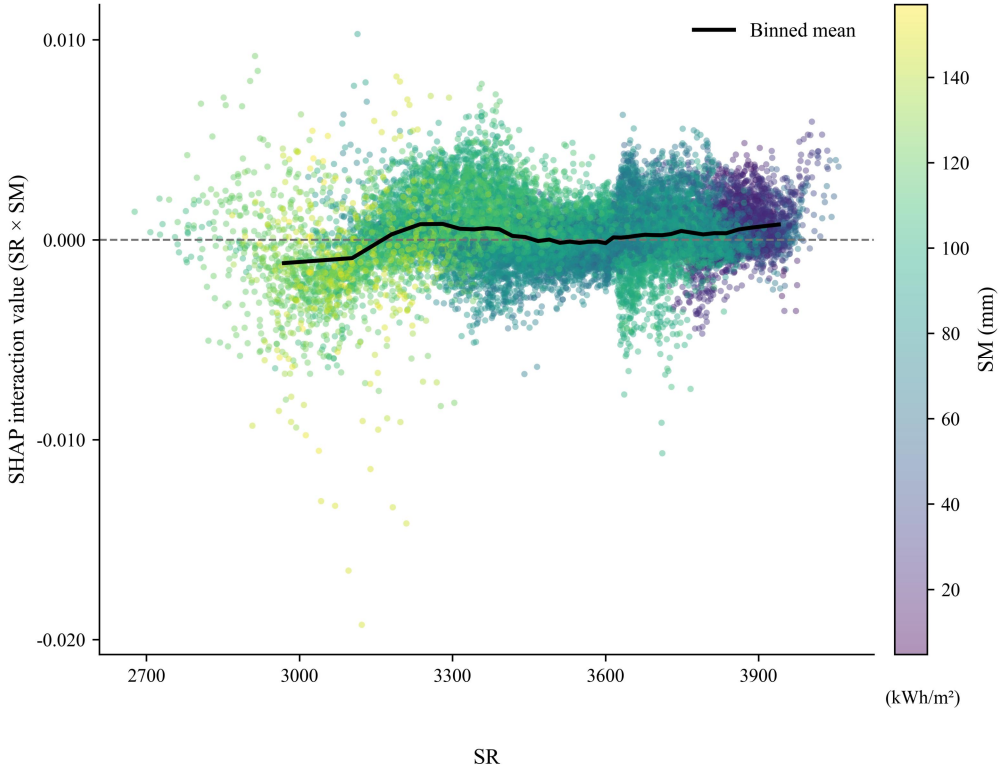


Figure S3. Period-stratified accumulated local effects (ALE) of grazing intensity (GI) on the predicted grassland resilience index (RI).

Solid curves show mean GI ALE responses across 25 outer RF models for 1985–2000 and 2001–2020. The comparison is restricted to the shared 5th–95th percentile GI range (0.346–2.210 SU/ha); each curve was centered at the common reference GI of 0.958 SU/ha (vertical dotted line). Shaded areas represent empirical 2.5th–97.5th percentile intervals across outer models rather than formal confidence intervals. The horizontal dashed line indicates zero relative ALE effect.

