

BIG GEOSPATIAL DATA ANALYTICS FOR SUSTAINABLE NATURAL RESOURCE MANAGEMENT

Ismayilova G.M.¹, Karimova K.S.^{1*}, Babakhanov A.E.², Teymurov J.T.³

¹Mingachevir State University, Azerbaijan

21, D. Aliyeva Street, Mingachevir, AZ4500

²Institute of Geography, Ministry of Science and Education of the Republic of Azerbaijan

117, H. Javid Ave., Baku, AZ1073

³LANDAU School, Azerbaijan

77, Samad Vurgun Street, Baku, AZ1022

*Corresponding author: kamala.karimova@mdu.edu.az

Keywords: cloud computing,
decision support,
Earth observation,
machine learning,
spatial cross-validation

Summary. Big geospatial datasets offer opportunities for natural-resource monitoring, but spatial dependence, unstable indicator weights, model transferability, and uncertainty complicate their integration. This study evaluates a reproducible workflow combining pressure indexing, degradation-risk modelling, spatiotemporal validation, predictor interpretation, intervention analysis, and computational scaling. A synthetic benchmark of 2,400 grid cells observed over 24 months (57,600 records) was analysed using seven normalized indicators. Indicator weights were assessed through 100 spatial block-bootstrap samples, while logistic regression and gradient boosting were evaluated under four validation schemes. Permutation analysis and Monte Carlo simulations assessed predictor contributions and intervention uncertainty. The first principal component explained 60.7% of variance, the mean pressure index was 0.502, and Moran's I was 0.954. Logistic regression achieved AUC values of 0.804 under random validation and 0.785 under buffered spatial validation, performing comparably to gradient boosting. Hydro-ecological variables provided the strongest transferable predictive information. The integrated intervention reduced the priority-area index by 12.3% and the critical share by 29.0 percentage points. Computational scaling remained near-linear up to 2.30 million records. The workflow supports uncertainty-aware natural-resource assessment, although operational use requires local observations and independent field validation.

© 2026 Earth Science Division, Azerbaijan National Academy of Sciences. All rights reserved.

1. Introduction

Natural-resource monitoring increasingly integrates satellite observations, climate reanalysis, digital soil products, sensor networks, and administrative data. Their volume, heterogeneity, and spatial dependence make them central to geospatial big-data research, particularly for monitoring water, soil, forests, and agricultural systems (Ma et al., 2015; Koh et al., 2022). Remote sensing provides repeated observations, GIS links them to terrain and management units, and cloud platforms facilitate large-scale processing (Gorelick et al., 2017; Sharma et al., 2024).

However, computational capacity alone does not ensure reliable environmental prediction. Spatial dependence can make random train-test splits overly optimistic, while collinearity, unstable composite-index weights, model transferability, and scenario uncertainty can affect interpretation and decision support (Dorrmann et al., 2013; Roberts et al., 2017; Valavi et al., 2019; Ploton et al., 2020; Koldasbayeva et al., 2024; Stock, 2025).

This study addresses the lack of a compact workflow that jointly evaluates index-weight stability, spatial and temporal transfer, correlated predictor groups, intervention uncertainty, and computational scaling. A controlled synthetic benchmark is used because its data-generating process can be evaluated systematically (Morris et al., 2019). The analysis examines the stability of the pressure index under spatial resampling, the effect of validation design on predictive performance, the transferability of predictor groups, and the robustness of intervention results to variation in weights and treatment effects.

2. Materials and methods

2.1. Analytical design and benchmark construction

Figure 1 summarizes the workflow. In an empirical application, observations could be drawn from Sentinel-2, ERA5-Land and SoilGrids and then joined with terrain, infrastructure and field measurements (Drusch et al., 2012; Munoz-Sabater et al., 2021; Poggio et al., 2021). The present benchmark is synthetic because no verified dataset for a named study area was available.

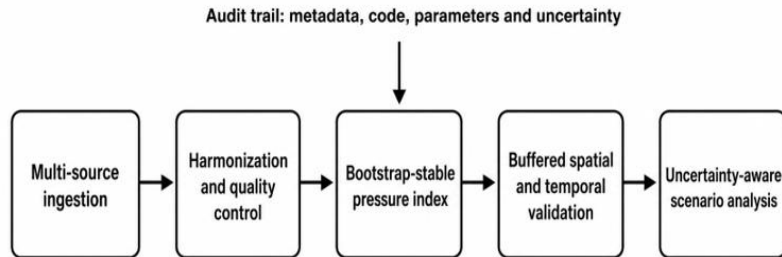


Fig. 1. Uncertainty-aware and reproducible analytical workflow

The benchmark contains 40 rows, 60 columns and 24 monthly observations, giving 57,600 cell-month records over a synthetic area of 6,000 square kilometres. The random seed was fixed at 42. Five spatially structured latent fields represent aridity, human pressure, terrain, water proximity and an unobserved effect. Gaussian-smoothed noise was rescaled to the 0-1 interval and transformed into the variables shown in Table 1.

Dry- and wet-season components were calculated first.

$$D_m = \frac{1 + \sin(2\pi(m \bmod 12)/12 - \pi/2)}{2}$$

$$W_m = 1 - D_m$$

Rainfall anomaly, road density, slope and distance to water were then generated.

$$R = \text{clip}[\varepsilon(0, 0.65) - 0.90A - 0.55D_m + 0.25W_m, -2.5, 1.5]$$

$$\text{Road density} = \text{clip}[0.05 + 0.90(0.75H + 0.25H_2), 0, 1]$$

$$\text{Slope} = \text{clip}(0.05 + 0.85S, 0, 1)$$

$$\text{Distance to water} = 1 - P$$

In the expressions, A, H, S and P denote aridity, human pressure, slope and water proximity; H_2 is a second smoothed human-pressure field, and $\varepsilon(0, \sigma)$ denotes zero-mean Gaussian error with standard deviation σ .

Table 1

Synthetic benchmark variables and generation rules

Variable	Range	Pressure direction	Rule before clipping
Vegetation index	0.08-0.85	Lower = pressure	$0.80 - 0.42A - 0.17H - 0.13S + 0.11W + 0.10R + \varepsilon(0, 0.035)$
Water index	-0.35-0.45	Lower = pressure	$0.34 - 0.46A - 0.10H + 0.10W + 0.09R + \varepsilon(0, 0.045)$
Soil moisture	0.04-0.58	Lower = pressure	$0.56 - 0.38A - 0.07H + 0.09W + 0.075R + \varepsilon(0, 0.028)$
Temperature anomaly	-5.9 to 9.9 °C	Higher = pressure	$-3.0 + 9.5A + 2.1H + 2.7(1 - W) - 0.8R + \varepsilon(0, 0.70)$
Drought occurrence	0 or 1	Occurrence = pressure	$s_D = -1.55 + 2.5A + 1.25(1 - W) - 1.05R$ $\text{Bernoulli}\{\text{logit}^{-1}(s_D)\}$
Land-use intensity	0.08-0.96	Higher = pressure	$0.06 + 0.86H + 0.025\sin(\text{month phase}) + \varepsilon(0, 0.02)$
Erosion susceptibility	0.07-0.87	Higher = pressure	$0.03 + 0.43S + 0.28(\text{land use}) + 0.20A + 0.11(1 - \text{vegetation}) + \varepsilon(0, 0.025)$

Source: Synthetic benchmark used in this study. A = aridity; H = human pressure; S = slope; W = wet-season component; R = rainfall anomaly.

2.2. Bootstrap-stable pressure index

All layers were matched by grid cell and month. For empirical data, the same stage would retain cloud masks, quality flags, measurement units, resampling rules and spatial reference information. The normalized difference vegetation index and normalized difference water index were calculated with Equations (1) and (2), where NIR denotes near-infrared reflectance (Tucker, 1979; McFeeters, 1996).

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (1)$$

$$NDWI = \frac{Green - NIR}{Green + NIR} \quad (2)$$

Each variable was converted to a pressure score q_{ij} between 0 and 1. Equation (3) was used for direct-pressure indicators and Equation (4) for indicators where higher original values represented better conditions. Here x_{ij} is observation i for indicator j , while a_j and b_j are fixed lower and upper limits.

$$q_{ij} = \frac{x_{ij} - a_j}{b_j - a_j} \quad (3)$$

$$q_{ij} = \frac{b_j - x_{ij}}{b_j - a_j} \quad (4)$$

Pressure scores were standardized before principal component analysis. Absolute first-component loadings were rescaled to sum to one. Weight stability was evaluated with 100 spatial block-bootstrap samples based on 10 by 10 cell blocks, retaining all months for selected blocks. Equation (5) defines the bootstrap weights, their median and the bootstrap-stable index. Critical probability was the share of bootstrap indices at or above 0.65. Equal weights and thresholds from 0.60 to 0.70 provided sensitivity checks.

$$w_j^{(b)} = \frac{|l_{j1}^{(b)}|}{\sum_k |l_{k1}^{(b)}|}, \quad \tilde{w}_j = \text{median}_b [w_j^{(b)}], \quad \text{BS-SRPI}_i = \sum_j \tilde{w}_j q_{ij} \quad (5)$$

2.3. Prediction, validation and decision analysis

A binary degradation event was generated for prediction. Equation (6) combines the seven normalized pressures, temperature-soil-moisture and land-use-erosion interactions, negative rainfall anomaly and an unobserved spatial field U . The outcome y_i follows a Bernoulli distribution with probability p_i ; prevalence in the generated data was 47.3%.

$$\begin{aligned} \eta_i = & -4.55 + 1.35q_V + 1.05q_W + 1.35q_{SM} + 0.95q_T + 0.65D + 1.15q_{LU} + 1.10q_E \\ & + 1.15q_T q_{SM} + 0.80I(q_{LU} > 0.58, q_E > 0.55) + 0.35\max(-R, 0) + 0.35U \\ p_i = & [1 + \exp(-\eta_i)]^{-1}, \quad y_i \sim \text{Bernoulli}(p_i) \end{aligned} \quad (6)$$

Two classifiers were compared. Logistic regression used L2 regularization, $C = 1.0$ and balanced class weights. Histogram-based gradient boosting used 120 iterations, learning rate 0.06, 31 terminal leaves, minimum leaf size 30 and L2 regularization 0.5. Predictors were the seven indicators, rainfall anomaly, slope, road density, distance to water and normalized coordinates.

Validation used a stratified 75:25 split, five buffered vertical spatial folds, five horizontal folds for logistic regression, a temporal holdout with months 1-18 for training and 19-24 for testing, and a combined holdout for later months in an unseen eastern block. A two-cell buffer separated spatial test blocks from training data. Performance was measured by area under the receiver operating characteristic curve, balanced accuracy, F1, Brier score and calibration coefficients (Gneiting & Raftery, 2007). Buffer-width sensitivity was checked at 0, 1, 2, 4 and 6 cells. A paired 2,000-resample bootstrap compared model AUCs on the spatiotemporal holdout; details are reported in Supplementary File S2.

$$\text{Balanced accuracy} = \frac{1}{2} \left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP} \right) \quad (7)$$

$$\text{Brier score} = \frac{1}{n} \sum_i (p_i - y_i)^2 \quad (8)$$

Global Moran's I was calculated from the 24-month mean index with first-order rook adjacency and 999 permutations. Grouped permutation importance was evaluated on the spatiotemporal holdout with 30 repetitions for hydro-ecological, land-and-erosion and spatial-context groups. Twenty individual permutations were retained as a secondary check.

Three intervention bundles targeted the highest-pressure 15% of cells. Two hundred Monte Carlo runs varied treatment effects by 20% and sampled a bootstrap weight vector. Results are medians with 95% intervals and are interpreted as sensitivity estimates, not causal forecasts. Computational scaling was tested from 57,600 to 2,304,000 records using seven float32 layers; each size was repeated three times. Python 3.13 and five logical processors were used. Full code is supplied in Supplementary File S1.

$$I = \frac{n}{W} \left(\frac{\sum_i \sum_j w_{ij} z_i z_j}{\sum_i z_i^2} \right) \quad (9)$$

3. Results

The first principal component explained 60.7% of variance. Median stable weights were 0.190 for vegetation, 0.185 for water, 0.188 for soil moisture and 0.184 for temperature. Land use had the least stable weight at 0.031, with a 95% interval of 0.003-0.069. The mean pressure index was 0.502. Equal- and stable-weight indices were strongly related ($r = 0.982$), but class agreement was lower at 86.3%. Only 2.4% of records had an uncertain critical classification.

Low, moderate, high and critical shares were 18.6%, 30.2%, 32.7% and 18.4%. Figure 2 shows the spatial gradient. Moran's I was 0.954 ($p = 0.001$), indicating strong clustering in the synthetic grid. When the critical threshold changed from 0.60 to 0.70, the critical share fell from 28.0% to 11.2%.

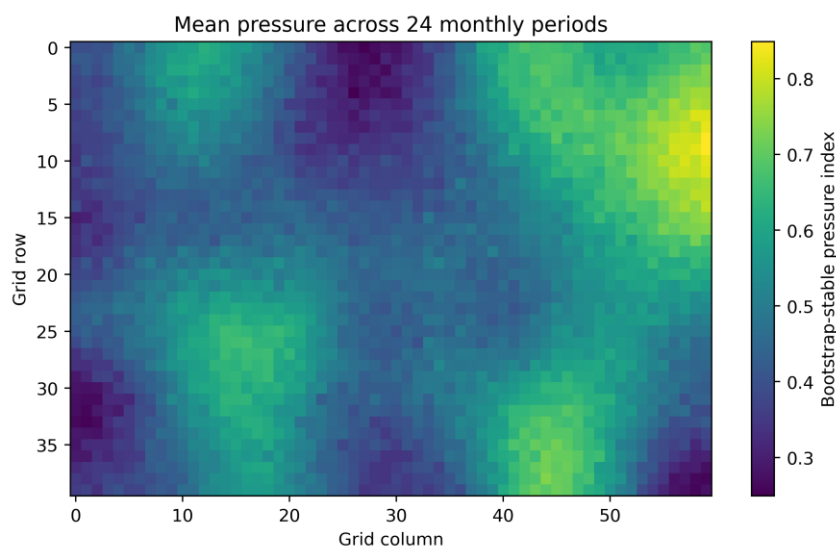


Fig. 2. Mean bootstrap-stable pressure index across the synthetic grid

Note: The grid is synthetic and not georeferenced. A scale bar, north arrow and coordinate reference system are not applicable.

Model discrimination was nearly identical across algorithms (Table 2). Logistic regression reached an AUC of 0.804 with random validation and 0.785 with buffered spatial validation; gradient boosting reached 0.802 and 0.784. Temporal AUC values were 0.802 and 0.799, while spatiotemporal values were 0.809 and 0.808. For logistic regression, mean spatial AUC stayed within 0.78531-0.78540 across buffer widths from 0 to 6 cells. The paired spatiotemporal AUC difference was 0.0017 (95% CI -0.0014 to 0.0048), supporting practical equivalence of the two models.

Table 2

Discrimination across validation designs

Model	Random AUC	Buffered spatial AUC	Temporal AUC	Spatiotemporal AUC
Logistic regression	0.804	0.785 +/- 0.016	0.802	0.809
Gradient boosting	0.802	0.784 +/- 0.017	0.799	0.808

Source: Calculations based on the synthetic benchmark. AUC = area under the receiver operating characteristic curve. Buffered spatial values are reported as mean +/- standard deviation.

Buffered spatial classification and calibration are reported in Table 3. Both models had a mean calibration slope of 0.987 and a Brier score of 0.180. Logistic regression was retained for interpretation because its predictive performance was essentially the same as gradient boosting. Its spatiotemporal calibration intercept was -0.346.

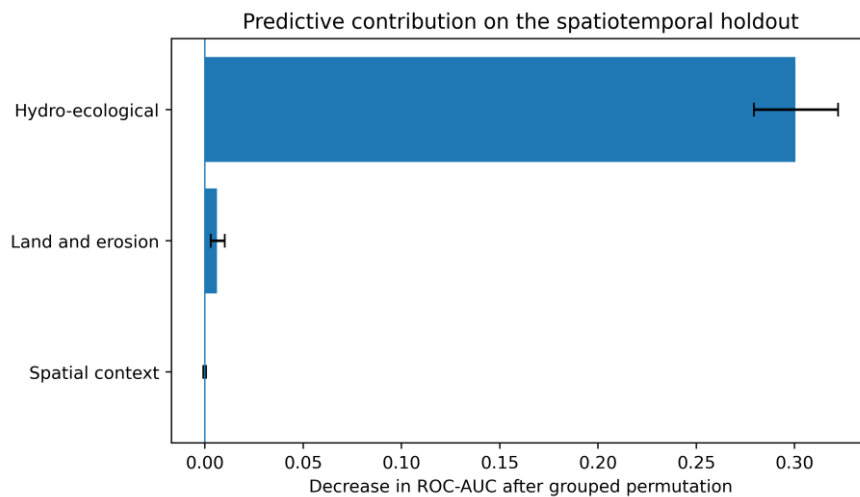
Table 3

Classification and calibration under buffered spatial validation

Model	Balanced accuracy	F1 score	Brier score	Calibration slope
Logistic regression	0.690 +/- 0.022	0.680 +/- 0.129	0.180 +/- 0.008	0.987 +/- 0.031
Gradient boosting	0.690 +/- 0.023	0.676 +/- 0.131	0.180 +/- 0.008	0.987 +/- 0.023

Source: Calculations based on the synthetic benchmark. Values are mean +/- standard deviation across five buffered folds.

Grouped permutation results are shown in Figure 3. Permuting the hydro-ecological group reduced AUC by 0.300, with a 95% interval of 0.279-0.322. The decrease for land and erosion was 0.006, while spatial context was unstable. Individual checks ranked soil moisture, vegetation and drought highest. These values describe predictive contribution rather than causality.

**Fig. 3.** Grouped permutation importance on the spatiotemporal holdout

Source: Calculations based on the synthetic benchmark. Error bars show 95% empirical intervals from 30 permutations.

Table 4 summarizes the intervention scenarios. The integrated option reduced the priority-area index by a median of 12.3% (95% interval 10.2-15.0%) and the critical share by 29.0 percentage points (23.4-35.5). Hydro-ecological and land-and-erosion actions reduced the index by 7.6% and 3.8%, respectively. The direction of change remained stable across uncertainty runs.

Table 4

Monte Carlo intervention effects in priority cells

Intervention bundle	Median index change (95% interval)	Median critical-share change (95% interval)
Hydro-ecological	-7.6% (-9.6 to -5.4)	-17.0 pp (-22.5 to -11.9)
Land and erosion	-3.8% (-6.0 to -1.5)	-8.0 pp (-13.4 to -3.4)
Integrated	-12.3% (-15.0 to -10.2)	-29.0 pp (-35.5 to -23.4)

Source: Calculations based on the synthetic benchmark. Negative values show lower pressure; pp = percentage points.

At 2,304,000 records, the in-memory calculation had a scaling exponent of 1.12 and estimated working memory of 138.2 MB (Figure 4). This result describes the normalization and index kernel only; file transfer, masking and archive latency were outside the timing test.

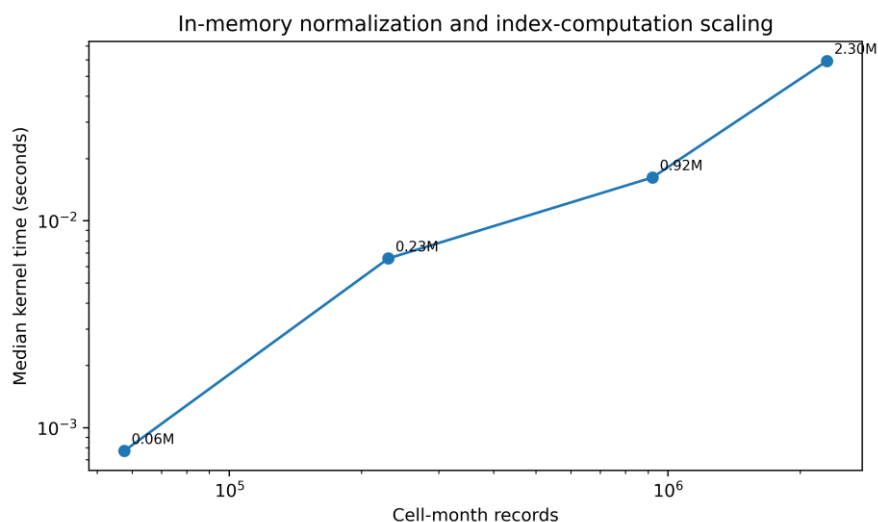


Fig. 4. Scaling of the in-memory normalization and index kernel

Source: Calculations based on the synthetic benchmark. Each point represents the median of three runs.

4. Discussion

The results show that uncertainty in a composite index enters through the indicator weights as well as through the final score. The block bootstrap revealed that land-use weight varied more than the main hydro-ecological weights. Broad spatial ordering may therefore remain similar while records near a management threshold change class. For decision support, a probability of critical status is more informative than a single fixed boundary (Koldasbayeva et al., 2024).

Validation design produced a second key finding. Random AUC was about 0.02 higher than buffered spatial AUC, which is consistent with evidence that random cross-validation can be optimistic when structured dependence is ignored (Roberts et al., 2017; Ploton et al., 2020). In this benchmark, changing the buffer from 0 to 6 cells had little effect on mean AUC, but the result still shows that validation geometry should match the prediction setting (Stock, 2025; Wang et al., 2025).

Temporal and spatiotemporal tests add a calibration perspective. Discrimination remained close to the spatial result, but calibration shifted, indicating that locations may still be ranked usefully while predicted probabilities require local recalibration. This supports evaluating generalization, autocorrelation and uncertainty together rather than as isolated technical checks (Koldasbayeva et al., 2024).

Logistic regression and gradient boosting produced almost identical transfer performance. Because the paired bootstrap interval for the spatiotemporal AUC difference included zero, the more complex model showed no meaningful discrimination advantage. Logistic regression was therefore retained because it is easier to audit while offering comparable accuracy and calibration.

The workflow also differs from studies focused mainly on data access or image processing. Google Earth Engine makes planetary-scale remote-sensing analysis technically feasible (Gorelick et al., 2017), but computational access alone does not solve validation and uncertainty problems. The present benchmark therefore concentrates on the analytical stage between data access and management use: index stability, model transfer and uncertainty-aware scenarios.

Table 5 positions the workflow relative to selected studies. Its novelty lies in combining uncertainty-aware indexing, spatial-temporal validation, grouped interpretation, intervention analysis and a reproducible scaling check within one framework.

This integration has practical value. Pressure maps can identify areas needing closer monitoring, bootstrap probabilities can show where prioritization is uncertain, spatial validation can estimate transfer beyond sampled locations, and grouped importance can guide continued monitoring. The integrated package produced the largest reduction, but its uncertainty interval is more informative than a single point estimate because local evidence may change expected treatment effects.

Empirical transfer should follow a clear sequence. Synthetic layers should be replaced by harmonized Sentinel-2, ERA5-Land, SoilGrids and local administrative data with quality and spatial-reference metadata. Field sampling should cover both high- and low-pressure zones, index weights and decision thresholds should be recalculated locally, and model outputs should be compared with independent field observations before the workflow supports investment or regulation (Sharma et al., 2024).

Table 5

Position of the present workflow relative to selected studies

Study	Main focus	Validation / uncertainty emphasis	Relation to this workflow
Roberts et al. (2017); Stock (2025)	Cross-validation design for structured spatial data	Spatial dependence and block-size choice	Supports blocked validation and sensitivity checks
Ploton et al. (2020); Wang et al. (2025)	Geospatial prediction evaluation	Spatial transfer and clustered samples	Supports caution when interpreting a single CV design
Koldasbayeva et al. (2024)	Environmental geospatial modelling	Autocorrelation, generalization and uncertainty	Motivates joint treatment of validation and uncertainty
Sharma et al. (2024)	Remote sensing and GIS for resource management	Importance of in-situ information	Supports field validation before operational use
Present benchmark	Integrated decision workflow	Weights, spatial-temporal transfer, grouped importance and scenarios	Combines these checks in one reproducible experiment

Source: Comparative synthesis based on the cited studies.

The limitations remain important. The degradation event is synthetic and generated from known predictors, so the benchmark demonstrates analytical behaviour rather than external ecological validity. The scenarios omit costs, delays, institutional constraints and ecological feedback, and the scaling test covers only the in-memory kernel. The next step is to test the same workflow in a named study area with observed outcomes and an independent field-validation set.

5. Conclusion

The analysis shows that big geospatial data can be organized into a reproducible workflow for pressure indexing, transfer testing, predictor interpretation and uncertainty-aware scenario comparison. Spatial resampling exposed weight instability that a single principal-component solution would hide. Random validation was slightly optimistic relative to buffered spatial validation, while logistic regression and gradient boosting produced nearly identical transfer performance. Hydro-ecological information dominated the grouped predictive signal, and the integrated intervention remained beneficial after uncertainty was propagated through both weights and treatment effects.

The main contribution is methodological: accuracy, spatial dependence, calibration, uncertainty and management sensitivity are evaluated together rather than in separate steps. The benchmark is suitable for auditing the workflow, not for direct regulatory use. The next stage is an empirical application in a defined study area using harmonized satellite-climate-soil data and independent field observations. Local thresholds and probability calibration should be estimated and tested before operational use.

REFERENCES

- Ahmadova, Kh.N., Jabarov, S.H., Aliyev, Y.I., Aliyeva, Sh.N., Trukhanov, A.V., Trukhanov, S.V., & Mirzayev, M.N. (2025). Design, production and investigation of structural singularities of layered metal-oxide nanostructures. *International Journal of Nanoscience*, 24, Article 2450015. <https://doi.org/10.1142/S0219581X24500157>
- Aslanova, K., & Güngöroğlu, C. (2026). Some ecological factors controlling the distribution of forested areas in the Zangilan and Gubadli regions. *Acta Botanica Caucasica*, 5(1), 84–92. <https://doi.org/10.30546/abc.2026.1176>
- Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carre, G., Garcia Marquez, J. R., Gruber, B., Lafourcade, B., Leita, P.J., Munkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schroder, B., Skidmore, A.K., Zurell, D., & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), 27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>
- Dritsas, E., & Trigka, M. (2025). Remote sensing and geospatial analysis in the big data era: A survey. *Remote Sensing*, 17(3), Article 550. <https://doi.org/10.3390/rs17030550>
- Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., & Bargellini, P. (2012). Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment*, 120, 25–36. <https://doi.org/10.1016/j.rse.2011.11.026>
- Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477), 359–378. <https://doi.org/10.1198/016214506000001437>
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>
- Jafarov, S. (2024). Internationalisation and cultural aspects of online learning, training, and research. *Pakistan Journal of Life and Social Sciences*, 22(2), 6437–6452. <https://doi.org/10.57239/PJLSS-2024-22.2.00486>

- Jafarov, S., Babashova, S., Dadashov, T., & Garibov, S. (2024). The influence of intonation on the interpretation of parenthetical clauses. *Pakistan Journal of Life and Social Sciences*, 22(2). <https://doi.org/10.57239/pjlss-2024-22.2.00810>
- Jolliffe, I.T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), Article 20150202. <https://doi.org/10.1098/rsta.2015.0202>
- Koh, K., Hyder, A., Karale, Y., & Kamel Boulos, M.N. (2022). Big geospatial data or geospatial big data? A systematic narrative review on spatial data infrastructures. *Remote Sensing*, 14(13), Article 2996. <https://doi.org/10.3390/rs14132996>
- Koldasbayeva, D., Tregubova, P., Gasanov, M., Zaytsev, A., Petrovskaya, A., & Burnaev, E. (2024). Challenges in data-driven geospatial modeling for environmental research and practice. *Nature Communications*, 15, Article 10700. <https://doi.org/10.1038/s41467-024-55240-8>
- Ma, Y., Wu, H., Wang, L., Huang, B., Ranjan, R., Zomaya, A., & Jie, W. (2015). Remote sensing big data computing: Challenges and opportunities. *Future Generation Computer Systems*, 51, 47–60. <https://doi.org/10.1016/j.future.2014.10.029>
- McFeeters, S. K. (1996). The use of the normalized difference water index in the delineation of open water features. *International Journal of Remote Sensing*, 17(7), 1425–1432. <https://doi.org/10.1080/01431169608948714>
- Morris, T.P., White, I.R., & Crowther, M.J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, 38(11), 2074–2102. <https://doi.org/10.1002/sim.8086>
- Munoz-Sabater, J., Dutra, E., Agusti-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Bousetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D.G., Piles, M., Rodriguez-Fernandez, N.J., Zsoter, E., Buontempo, C., & Thepaut, J.-N. (2021). ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, 13, 4349–4383. <https://doi.org/10.5194/essd-13-4349-2021>
- Ploton, P., Mortier, F., Rejou-Mechain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N., Lya-pustin, A., Gourlet-Fleury, S., & Pelissier, R. (2020). Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nature Communications*, 11, Article 4540. <https://doi.org/10.1038/s41467-020-18321-y>
- Poggio, L., de Sousa, L.M., Batjes, N.H., Heuvelink, G.B.M., Kempen, B., Ribeiro, E., & Rossiter, D. (2021). SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty. *SOIL*, 7, 217–240. <https://doi.org/10.5194/soil-7-217-2021>
- Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Aroita, G., Hauenstein, S., Lahoz-Monfort, J.J., Schroder, B., Thuiller, W., Warton, D.I., Wintle, B.A., Hartig, F., & Dormann, C.F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8), 913–929. <https://doi.org/10.1111/ecog.02881>
- Sharma, S., Beslity, J.O., Rustad, L., Shelby, L.J., Manos, P.T., Khanal, P., Reinmann, A.B., & Khanal, C. (2024). Remote sensing and GIS in natural resource management: Comparing tools and emphasizing the importance of in-situ data. *Remote Sensing*, 16(22), Article 4161. <https://doi.org/10.3390/rs16224161>
- Stock, A. (2025). Choosing blocks for spatial cross-validation: Lessons from a marine remote sensing case study. *Frontiers in Remote Sensing*, 6, Article 1531097. <https://doi.org/10.3389/frsen.2025.1531097>
- Tucker, C.J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2), 127–150. [https://doi.org/10.1016/0034-4257\(79\)90013-0](https://doi.org/10.1016/0034-4257(79)90013-0)
- Valavi, R., Elith, J., Lahoz-Monfort, J.J., & Guillera-Aroita, G. (2019). blockCV: An R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods in Ecology and Evolution*, 10(2), 225–232. <https://doi.org/10.1111/2041-210X.13107>
- Wang, Y., Khodadadzadeh, M., & Zurita-Milla, R. (2025). A dissimilarity-adaptive cross-validation method for evaluating geospatial machine learning predictions with clustered samples. *Ecological Informatics*, 90, Article 103287. <https://doi.org/10.1016/j.ecoinf.2025.103287>

АНАЛИЗ БОЛЬШИХ ГЕОПРОСТРАНСТВЕННЫХ ДАННЫХ ДЛЯ УСТОЙЧИВОГО УПРАВЛЕНИЯ ПРИРОДНЫМИ РЕСУРСАМИ

Исмаилова Г.М.¹, Керимова К.С.^{1*}, Бабаханов А.Э.², Теймуров Дж. Т.³

¹Мингячевирский государственный университет, Азербайджан
AZ4500, Мингячевир, ул. Д. Алиевой, 21

²Институт географии Министерства науки и образования Азербайджанской Республики, Азербайджан
AZ1073, Баку, пр. Г. Джавида, 117

³Школа LANDAU, Азербайджан
AZ1022, Баку, ул. Самеда Вургуня, 77

*Автор, отвечающий за переписку: kamala.karimova@mdu.edu.az

Резюме. Большие массивы геопространственных данных открывают новые возможности для мониторинга природных ресурсов, однако их интеграцию осложняют пространственная зависимость, нестабильность весов показателей, переносимость моделей и неопределённость. В исследовании оценивается воспроизводимая методика, объединяющая построение индекса нагрузки, моделирование риска деградации, пространственно-временную валидацию, интерпретацию предикторов, анализ мер воздействия и оценку вычислительной масштабируемости. Синтетический набор данных, включающий наблюдения за 2400 ячейками сетки в течение 24 месяцев (57 600 записей), проанализирован с использованием семи нормированных показателей. Веса показателей оценивались по 100 выборкам пространственного блочного бутстрэпа, а логистическая регрессия и градиентный бустинг — по четырём схемам валидации. Перестановочный анализ и моделирование методом Монте-Карло использовались для оценки вклада предикторов и неопределённости мер воздействия. Первая главная компонента объясняла 60,7% дисперсии, средний индекс нагрузки составлял 0,502, а индекс Морана I — 0,954. Значения AUC логистической регрессии достигли 0,804 при случайной валидации и 0,785 при пространственной валидации с буферными зонами; её результаты были сопоставимы с результатами градиентного бустинга. Гидроэкологические переменные обеспечивали наибольшую переносимую прогностическую информативность. Комплекс мер снизил индекс приоритетных территорий на 12,3%, а долю

территорий в критическом состоянии — на 29,0 процентного пункта. Рост вычислительных затрат оставался близким к линейному при объёме данных до 2,30 млн записей. Методика позволяет оценивать природные ресурсы с учётом неопределённости, однако её практическое применение требует местных наблюдений и независимой полевой валидации.

Ключевые слова: облачные вычисления, поддержка принятия решений, наблюдение Земли, машинное обучение, пространственная перекрёстная валидация

TƏBİİ RESURLARIN DAYANIQLI İDARƏ EDİLMƏSİ ÜÇÜN BÖYÜK GEOMƏKAN MƏLUMATLARININ TƏHLİLİ

İsmayilova G. M.¹, Kərimova K. S.^{1*}, Babaxanov A. E.², Teymurov C. T.³

¹Mingəçevir Dövlət Universiteti, Azərbaycan

AZ4500, Mingəçevir, D. Əliyeva küç. 21

²Azərbaycan Respublikası Elm və Təhsil Nazirliyi, Coğrafiya İnstitutu, Azərbaycan

AZ1073, Bakı, H. Cavid pr. 117

³LANDAU Məktəbi, Azərbaycan

AZ1022, Bakı, Səməd Vurğun küç. 77

*Məsul müəllif: kamala.karimova@mdu.edu.az

Xülasə. Böyük həcmli geoməkan məlumatları təbii ehtiyatların monitorinqi üçün yeni imkanlar yaradır, lakin məkan asılılığı, göstəricilərin çəki əmsallarının qeyri-sabitliyi, modellərin digər ərazilərə tətbiq edilə bilməsi və qeyri-müəyyənlik onların integrasiyasını çətinləşdirir. Tədqiqatda təzyiq indeksinin hesablanması, degradasiya riskinin modelləşdirilməsini, məkan-zaman validasiyasını, proqnozlaşdırıcı dəyişənlərin şərhini, müdaxilə tədbirlərinin təhlilini və hesablama miqyaslanmasının qiymətləndirilməsini birləşdirən təkrarlana bilən metodika qiymətləndirilir. 24 ay ərzində müşahidə olunan 2400 şəbəkə xanasını əhatə edən sintetik məlumat dəsti (57 600 qeyd) yeddi normallaşdırılmış göstərici əsasında təhlil edilmişdir. Göstəricilərin çəki əmsalları 100 məkan bloklu butstrep seçməsi, loqistik regressiya və qradiyent gücləndirmə modelləri isə dörd validasiya sxemi üzrə qiymətləndirilmişdir. Proqnozlaşdırıcı dəyişənlərin töhfəsini və müdaxilə tədbirlərinin qeyri-müəyyənliyini müəyyənləşdirmək üçün permutasiya təhlili və Monte-Karlo simulyasiyaları tətbiq edilmişdir. Birinci əsas komponent dispersiyanın 60,7%-ni izah etmiş, orta təzyiq indeksi 0,502, Moranın I indeksi isə 0,954 olmuşdur. Loqistik regressiyanın AUC göstəricisi təsadüfi validasiyada 0,804, bufer zonaları tətbiq edilən məkan validasiyasında isə 0,785 təşkil etmiş və nəticələr qradiyent gücləndirmə modeli ilə müqayisə olunan olmuşdur. Hidroekoloji dəyişənlər digər ərazilərə tətbiq zamanı ən yüksək proqnostik informativliyi təmin etmişdir. Kompleks müdaxilə tədbirləri prioritet ərazilərin indeksini 12,3%, kritik vəziyyətdə olan ərazilərin payını isə 29,0 faiz bəndi azaltmışdır. Məlumatların həcmi 2,30 milyon qeydə çatdıqda hesablama xərclərinin artımı xəttiyyə yaxın qalmışdır. Metodika təbii ehtiyatların qeyri-müəyyənlik nəzərə alınmaqla qiymətləndirilməsini dəstəkləyir, lakin praktik tətbiq üçün yerli müşahidə məlumatları və müstəqil çöl validasiyası tələb olunur.

Açar sözlər: bulud hesablamaları, qərarların qəbulunun dəstəklənməsi, Yer səthinin müşahidəsi, maşın öyrənməsi, məkan üzrə çarpaz validasiya