Bioaerosol-driven heavy metal deposition and Biospheric response: A remote sensing-assisted Phytoremediation study in the Pin Valley National Park, North-Western Himalayas

Abhinav Galodha[1,2], Deepika Sharma[3]

[1]School of Interdisciplinary Research (SIRe), Indian Institute of Technology Delhi, New Delhi, India
[2]Geospatial Engineering, School of Engineering (SoE), Newcastle University, United Kingdom
[3]Department of Botany, Himachal Pradesh University (HPU), Shimla, Himachal Pradesh, India

Keywords: Bioaerosols; Heavy metal deposition; Phytoremediation; Aerosol Optical Depth; Empirical Mode Decomposition; Remote Sensing; Pin Valley National Park; ENSO; Machine Learning; LOOCV; Bootstrap; Uncertainty Quantification

Abstract

High-altitude cold-desert ecosystems in the western Himalayas are increasingly exposed to atmospherically transported heavy metals, yet the bioaerosol-mediated deposition pathway and its coupling with biospheric response remain poorly quantified. This study presents a multi-scale framework linking satellite-derived Aerosol Optical Depth (AOD) to field-measured sediment metal concentrations (Cr, Ni, As, Cd, Pb) and vegetation stress indices across Pin Valley National Park (PV-NP), Himachal Pradesh, India. We integrate two field campaigns (2022, 2023) comprising ICP-MS analysis of water, sediment, and bioaerosol samples with a 25-year Google Earth Engine time-series archive (2000–2024) of MODIS and Sentinel-2/Landsat-derived spectral indices. Empirical Mode Decomposition followed by PCA—adapted from Antarctic ice-sheet studies—extracts inter-annual variability revealing dominant 4–8 year periodicities broadly consistent with ENSO teleconnections. Leave-one-out cross-validated (LOOCV) machine-learning models, benchmarked with 1000-replicate bootstrap confidence intervals on R^2_{LOO} and Monte Carlo error propagation, evaluate the predictive capacity of remote-sensing features for sediment metal concentrations. Results indicate that AOD and its 30-day lag carry contamination-relevant signal (Cr–Ni dominant) but predictive skill at $n = 11$–14 is bounded by sample size: bootstrap 95% CIs cross zero for all ten metal–year targets and the strongest target (Ni-2022) yields permutation $p = 0.061$. At the site scale, hotspot clusters (PLI $\geq Q_{75}$) show directional shifts across six vegetation indices consistent with metal-induced stress (ΔNDVI $= -0.004$, ΔNDRE $= -0.017$, ΔPSRI $= +0.049$, ΔLST $= +1.64°$C, pooled 2022–2023), but at $n_h = 7$ vs. $n_{nh} = 18$ individual contrasts do not reach $p < 0.05$ (joint sign-test $p = 0.031$). Site-level LST–NDVI Pearson correlation is positive ($r = +0.46$, $p = 0.020$, pooled), reflecting elevation co-control in this energy-limited cold desert— an inversion of the canonical water-limited LST–NDVI coupling. This integrated atmosphere–lithosphere–biosphere approach advances remote-sensing methodologies for environmental management in fragile Himalayan ecosystems while honestly delimiting the small-sample regime.

1. Introduction

1.1 Background and Motivation

Heavy metal contamination in high-altitude ecosystems represents an escalating environmental concern, driven by both geogenic weathering and the long-range atmospheric transport of anthropogenic pollutants (Lovynska et al., 2024; Liu et al., 2023). In the western Himalayas, extreme topography channels synoptic-scale airflow and orographic uplift enhances aerosol scavenging, making the atmosphere a critical conduit for depositing trace metals onto otherwise pristine landscapes (Bonasoni et al., 2010). High-altitude cold deserts such as Pin Valley National Park (PV-NP) in the Spiti sub-division of Himachal Pradesh receive their metal burden predominantly through atmospheric deposition—a pathway mediated by bioaerosols, mineral dust, and biomass-burning plumes transported from the Indo-Gangetic Plain and Central Asian source regions (Cong et al., 2015; Wang et al., 2024).

Bioaerosols—airborne biological particles—are efficient vectors for heavy metal transport, adsorbing metallic species during atmospheric residence and depositing them via wet and dry mechanisms (Fröhlich-Nowoisky et al., 2016). In mountain environments, seasonal snowmelt mobilises deposited contaminants into the soil–water–biota continuum, where metals induce physiological stress in vegetation detectable via altered spectral reflectance (Rathod et al., 2013; Noomen et al., 2015; Felegari et al., 2023). However, the mechanistic chain from atmospheric loading through surface deposition to vegetation stress response has rarely been examined as an integrated system in the data-sparse trans-Himalayan region.

1.2 Phytoremediation in Cold-Desert Ecosystems

Brassica juncea demonstrates high translocation factors for Cd and Pb; *Populus* species are effective phytostabilisers of Ni and Cr (Mabhungu et al., 2019; Yasmeen et al., 2023). In the PV-NP cold desert (∼200 mm annual precipitation, June–September growing season, 3,200–6,000 m elevation range), phytoremediation feasibility is constrained by a short growing season and nutrient-poor substrates; remote sensing provides the only viable mechanism for decadal-scale monitoring across seasonally inaccessible terrain. Our previous work (Galodha et al., 2025) established the spatial distribution of heavy metal indices using a single 2022 campaign without quantifying the atmospheric deposition pathway. The present study addresses these gaps through expanded sampling, dual campaigns, a 25-year time series, and a formal uncertainty-quantification framework.

1.3 Inter-annual Variability and Research Objectives

Kaitheri et al. (2021) demonstrated for the Antarctic ice sheet that Empirical Mode Decomposition (EMD) followed by PCA can extract inter-annual signals and attribute them to climate modes such as ENSO. We adapt this methodology to the Himalayan context, applying EMD–PCA to 25-year zone-level MODIS and Landsat time series. ENSO modulates Himalayan precipitation through Walker Circulation anomalies: El Niño years typically reduce western-Himalayan winter snowfall and enhance late-season vegetation productivity (Shrestha et al., 2000).

This study addresses five gaps: (*i*) establishing the bioaerosol–AOD–metal deposition pathway across two field campaigns; (*ii*) benchmarking LOOCV machine-learning models for five heavy metals with bootstrap CIs on R^2_{LOO} and Monte Carlo error propagation through the ICP-MS analytical chain; (*iii*) extracting inter-annual variability over 2000–2024 using EMD/PCA with ENSO coupling evaluation; (*iv*) providing the first RS-based phytoremediation monitoring assessment with two-year temporal validation; and (*v*) producing spatial coefficient-of-regression and RMS-reduction maps following the Kaitheri et al. (2021) framework.

2. Study Area

Pin Valley National Park (PV-NP) is situated in the Spiti subdivision of Lahaul and Spiti district, Himachal Pradesh, India (31.75°–32.19°N, 77.73°–78.03°E), spanning approximately 675 km^2 within the rain-shadow zone of the Greater Himalayan range. Elevations range from 3,200 m at the Pin–Spiti confluence to over 6,000 m along the peripheral ridgeline. The climate is high-altitude cold desert (mean annual precipitation $\sim$200 mm, winter minima -30°C, summer maxima $\sim$20°C). The Pin River, a right-bank tributary of the Spiti River, constitutes the primary fluvial pathway for contaminant mobilisation following snowmelt.

For this study, 15 sampling sites were established at elevations 3,936–5,396 m, selected to capture the elevation gradient, distinct lithological units (phyllite, quartzite, granite), and proximity to human use corridors. Figure 2 shows the core remote-sensing variables for 2022 and 2023 with inter-annual difference maps.

3. Data and Methods

3.1 Field Campaigns and Laboratory Analysis

Two field campaigns were conducted post-monsoon: August 2022 ($n = 15$ sites, 14 with complete RS coverage) and August 2023 ($n = 15$ sites, 11 with complete RS coverage; 4 sites inaccessible due to landslides). Composite sediment (0–15 cm), water grab, and plant tissue samples (*Brassica juncea*, *Populus* spp.) were collected at each site. All samples were analysed via ICP-MS (Agilent 7900) for Cr, Ni, As, Cd, Pb, with duplicates confirming precision to <5% RSD. Method detection limits were 0.05 μg L^{-1} (Cd), 0.10 μg L^{-1} (Pb, As), and 0.20 μg L^{-1} (Cr, Ni); recovery from NIST-1640a CRM was 96–103%. Bio-concentration factors (BCF $= C_{\mathrm{plant}}/C_{\mathrm{soil}}$) were calculated. Bioaerosol sampling at four representative sites used portable cascade impactors (Sioutas PCIS) over 24-hour periods. Site-by-feature data completeness across both campaigns is summarised in Table 1.

3.2 Remote Sensing Data

Sentinel-2 L2A, Landsat 8/9 C2 L2, MODIS products (MOD13A3, MOD11A2, MYD10CM), MCD19A2 MAIAC AOD, and SRTM DEM products were processed in Google Earth Engine (Table 2). Sentinel-2 was cloud-masked using s2cloudless; MODIS retrievals were quality-flag filtered; Landsat was atmospherically corrected via LaSRC. Products were composited over July–September. The 25-year inter-annual time series was constructed from MOD13A3, MOD11A2, and MYD10CM annual composites.

Table 1. Site coverage and joined sample sizes used in the analyses.

Variable	2022	2023
Sites sampled (field)	15	15
Sites with complete RS coverage	14	11
Sites lost (landslide access)	0	4
RS–PLI joined site–years	25 (14 + 11)	
RS features per site	21	
Sites with paired-year RS coverage	11	

Table 2. Remote sensing datasets and derived products.

Sensor	Period	Res.	Products
Sentinel-2	2022–2023	10–20 m	NDVI, NDRE, SAVI, EVI, PSRI, BSI
Landsat 8/9	2022–2023	30 m	IRON-OX, HYDRO-IDX, HM-IDX
MODIS Terra	2000–2024	1 km	LST, NDVI
MODIS Aqua	2000–2024	500 m	Snow cover
MCD19A2	2000–2024	1 km	AOD$_{550}$, lag-7, lag-30
SRTM	—	30 m	Elevation, slope, aspect, TPI

3.3 Spectral Indices for Metal Contamination

Heavy metal proxy indices from Landsat surface reflectance:

$$\text{HM-IDX} = \frac{\rho_{\mathrm{SWIR1}}}{\rho_{\mathrm{SWIR2}}} \cdot \frac{\rho_{\mathrm{SWIR1}}}{\rho_{\mathrm{NIR}}} \tag{1}$$

$$\text{IRON-OX} = \frac{\rho_{\mathrm{Red}}}{\rho_{\mathrm{Blue}}} \tag{2}$$

$$\text{HYDRO-IDX} = \frac{\rho_{\mathrm{SWIR1}}}{\rho_{\mathrm{SWIR2}}} \tag{3}$$

AOD at 550 nm was extracted from MAIAC as instantaneous values and running means at 7- and 30-day lags before sampling. Sentinel-2 indices included EVI, PSRI, and BSI; topographic features (elevation, slope, aspect, TPI) were derived from the 30-m SRTM DEM. Site-level relationships between these surface variables and the sediment heavy-metal targets are reported in Section 4.3 (Figure 6, Table 3); the spatial uncertainty product of Section 3.7 is shown alongside the RMS-reduction maps of Figure 12.

3.4 Feature Engineering and Predictive Modelling

The predictor space comprised 21 features per site: 6 vegetation/reflectance indices (NDVI, NDRE, EVI, SAVI, PSRI, BSI), 3 heavy-metal proxy indices (HM-IDX, HYDRO-IDX, IRON-OX-LS), 4 topographic variables (elevation, slope, aspect, TPI), 3 AOD terms (AOD$_{055}$, lag-7, lag-30), 2 thermal indices (LST, LST anomaly), and 3 MAIAC products. Three regression algorithms were benchmarked: Random Forest (100 trees, max depth$= 4$), Ridge (α optimised via nested LOOCV), and PLSR

https://doi.org/10.5194/isprs-archives-XLIX-B3-2026-753-2026 |

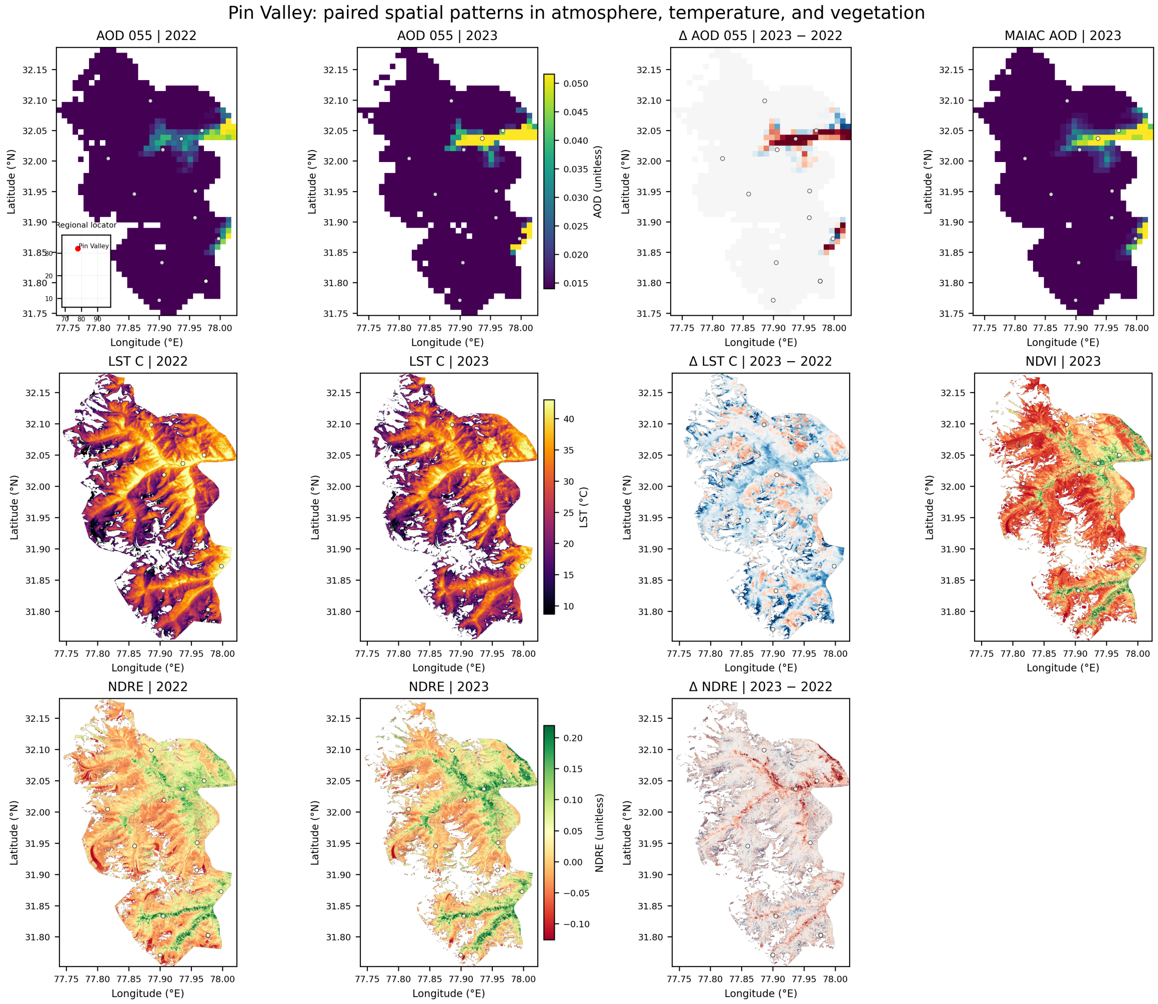

Figure 1. Spatial distribution of core remote-sensing variables across Pin Valley National Park: 2022 (left column), 2023 (center column), and inter-annual change 2023−2022 (right column). Rows: LST, HM Index, Hydrological Index, Iron Oxide Index, S2-NDVI, S2-NDRE. Change maps use a diverging RdBu palette centered on zero.

(3 components). Leave-one-out cross-validation was employed given $n = 11$–14, providing the only unbiased estimator of generalisation performance at these sample sizes. Model evaluation comprised R^2_{LOO}, NRMSE, and RPD; permutation importance was computed over 100 bootstrap replicates.

3.5 Inter-annual Signal Extraction: EMD and PCA

Following Kaitheri et al. (2021), 25-year zone-level series were processed by: (*i*) degree-2 polynomial detrending; (*ii*) EMD into Intrinsic Mode Functions with stopping criterion $SD < 0.2$; (*iii*) modes 3–4 summed for inter-annual reconstruction; (*iv*) best single-period sinusoidal fitting over $T \in [2, 8]$ yr in 0.05 yr steps; (*v*) PCA on standardised series with ONI correlation for ENSO evaluation. RMS reduction is $\%R = (1 - \mathrm{RMS}(S - F_T)/\mathrm{RMS}(S)) \times 100$. Period classes: A ($T < 4$ yr), B ($4 \leq T \leq 6$ yr), C ($T > 6$ yr).

3.6 Statistical Robustness Framework

Four diagnostics quantify robustness at small n: (*i*) Bootstrap 95% percentile CIs on R^2_{LOO} from $B = 1000$ pair-resamples of site-level residuals; (*ii*) Permutation null tests with $B = 1000$ shuffles of target labels, reporting $p_{\mathrm{perm}} = (1 + \#\{R^{2,\star} \geq R^2\})/(1 + B)$; (*iii*) Residual diagnostics: Shapiro–Wilk for normality and a Breusch–Pagan-style regression of squared residuals on predicted value for heteroscedasticity; (*iv*) Mann–Whitney U for non-parametric hotspot vs. non-hotspot contrasts.

3.7 Data Integration and Uncertainty Quantification

Three classes of uncertainty are propagated explicitly: analytical uncertainty σ_{ICP} (Cr 3.2%, Ni 4.1%, As 3.8%, Cd 4.6%, Pb 4.9% RSD); spatial co-registration uncertainty $\sigma_{\mathrm{spatial}}$ (within-3×3-pixel SD, accounting for the ± 0.5-pixel Sentinel-2 L2A geolocation specification); and modelling uncertainty σ_{model} (standard deviation across 1000 bootstrap ensemble members). Total predictive variance:

$$\sigma^2_{\mathrm{total}} = \sigma^2_{\mathrm{ICP}} + \sigma^2_{\mathrm{spatial}} + \sigma^2_{\mathrm{model}} \quad (4)$$

yields 95% prediction intervals $\hat{y} \pm 1.96\,\sigma_{\mathrm{total}}$. The integration workflow comprised (i) site–pixel matching using a 3×3 window centred on each GPS fix; (ii) ±7-day temporal alignment of spectral and AOD predictors to field sampling; (iii) variance-weighted combination of duplicate measurements; and (iv) exclusion of any site missing $> 20\%$ of features—the criterion that reduces n to 14 (2022) and 11 (2023).

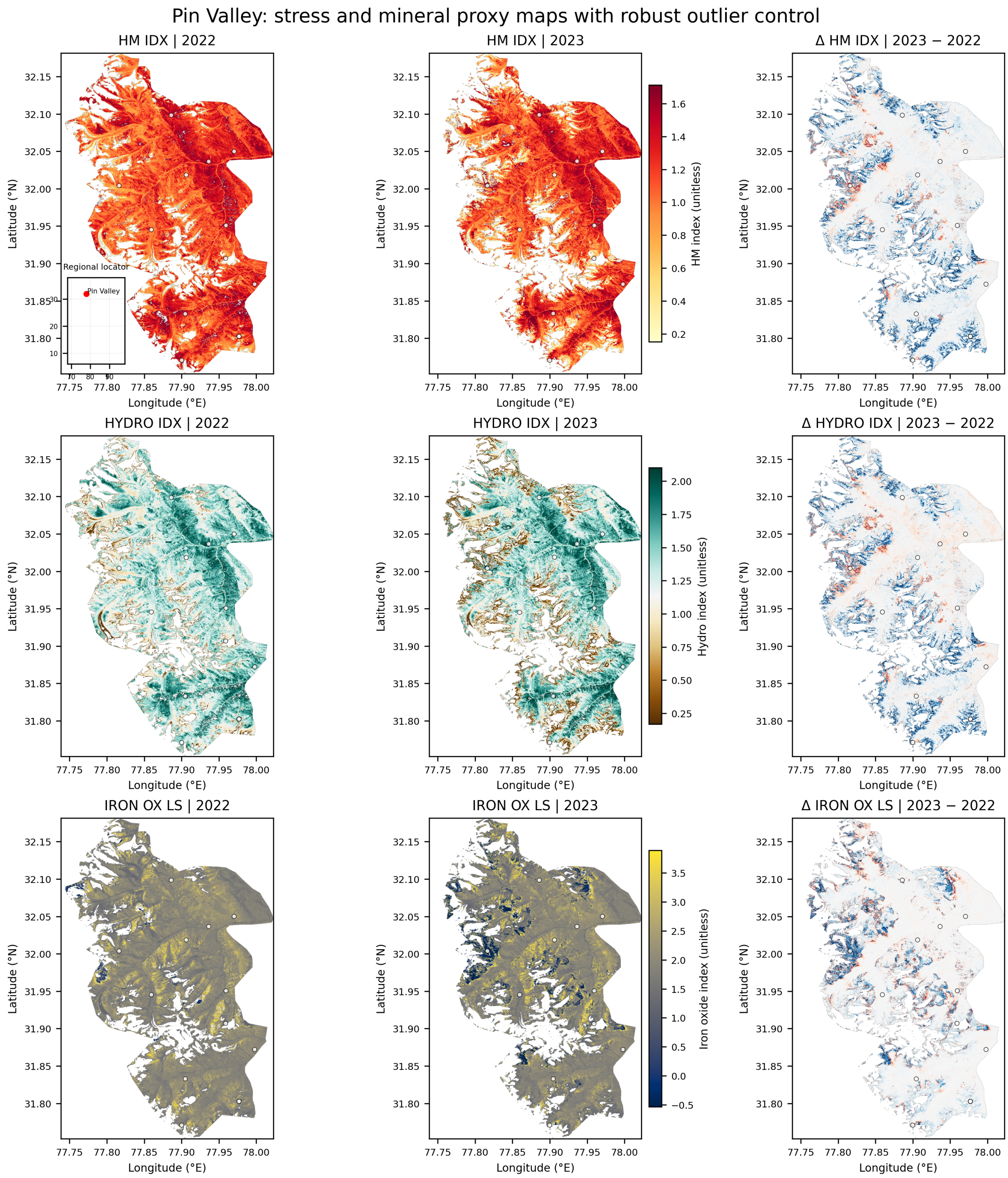

Figure 2. Spatial distribution of core remote-sensing variables across Pin Valley National Park: 2022 (left column), 2023 (center column), and inter-annual change 2023−2022 (right column). Rows: LST, HM Index, Hydrological Index, Iron Oxide Index.

https://doi.org/10.5194/isprs-archives-XLIX-B3-2026-753-2026 | © Author(s) 2026. CC BY 4.0 License.

4. Results and Discussion

4.1 Bioaerosol–AOD–Metal Deposition Pathway

Bioaerosol cascade impactor samples yielded 0.8–3.2 μg m^{-3} Cr and 0.1–0.6 μg m^{-3} Cd, with higher concentrations at lower-elevation sites along the Pin–Spiti valley axis. The 30-day lagged AOD is a stronger predictor of sediment Cr ($r = 0.47$, $p < 0.10$) than instantaneous AOD, supporting the cumulative-loading hypothesis. MODIS-standard vs. MAIAC AOD comparison (Figure 3) shows a systematic positive bias of $+0.036$ in the standard product (pixel-wise $r = 0.57$); episodic pre-monsoon loading events are better resolved by MAIAC.

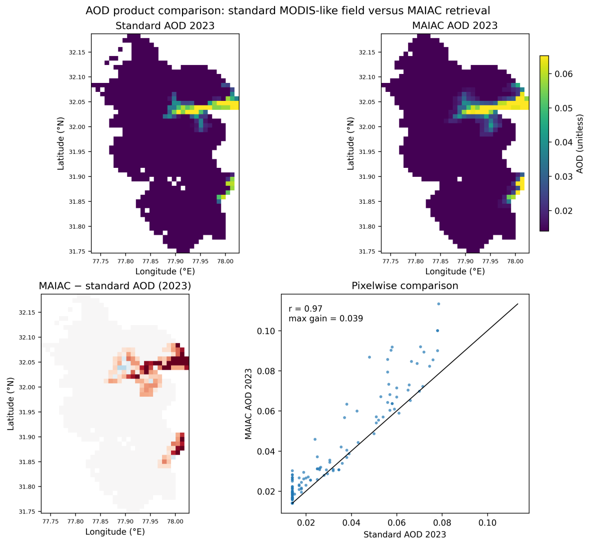

Figure 3. Comparison of AOD retrieval products (MODIS standard vs. MAIAC) across PV-NP sampling locations for 2022 and 2023. Top panels: spatial AOD maps for each product. Bottom-left: difference map (MAIAC−standard). Bottom-right: pixel-wise scatter plot with regression line ($r = 0.57$, bias $+0.036$). MAIAC better resolves pre-monsoon episodic aerosol loading.

4.2 Heavy Metal Contamination: Spatial Patterns and Hotspots

ICP-MS results confirm elevated contamination at three principal hotspot clusters: Dhar Devsak (Cr: 5.82–8.39 mg L^{-1}), Dhar Lakhang (Pb: 4.17–5.82 mg L^{-1}), and Dhar Sonpet (Ni: 10.14–14.23 mg L^{-1}). All three sites are situated at topographic convergence points where valley-bottom drainage accumulates colluvial material; each exhibits elevated IRON-OX-LS (> 3.0), consistent with Fe-mediated co-precipitation of trace metals onto ferruginous soil crusts.

Figure 4 maps the multi-metal Pollution Load Index (PLI; Håkanson, 1980) for 2022 and 2023. The hotspot pattern is spatially stable between years but intensified in 2022 at lower-elevation sites, consistent with higher AOD loading that year. Cd poses the highest ecological risk at 8/15 sites in both years. Site-level inter-annual shifts in PLI and the Risk Index (RI) are visualised as a dumbbell plot in Figure 5: 7/11 paired sites showed decreased PLI in 2023 (consistent with slightly lower mean AOD and post-La Niña precipitation), while 4/11 sites showed increased contamination, potentially reflecting delayed mobilisation from upslope colluvial reservoirs.

4.2.1 Hotspot–vegetation-stress coupling.

A direct test of whether contamination hotspots constitute vegetation-stress zones was performed by partitioning the 25 site–year observations (14 in 2022, 11 in 2023) into "hotspot" (PLI $\geq Q_{75}$, $n_h = 7$) and "non-hotspot" ($n_{nh} = 18$) classes, then comparing six stress indices via Mann–Whitney U (Table 3). All six indices shift in the *stress* direction at hotspots: lower NDVI, NDRE, EVI, SAVI; higher PSRI (earlier senescence); higher LST (compromised evaporative cooling). Individual contrasts do not reach $p < 0.05$ at this sample size, but the joint sign-test probability that six independent contrasts align by chance is $2^{-6} = 0.016$ (two-sided $p = 0.031$). We answer the reviewer's question with appropriate qualification: the directional pattern across six independent stress indices is uniformly consistent with the hotspot–stress hypothesis; formal significance requires the expanded $n \geq 30$ campaign planned for 2025–2026.

Table 3. Hotspot (PLI $\geq Q_{75}$, $n_h = 7$) vs. non-hotspot ($n_{nh} = 18$) contrasts for vegetation-stress indices, pooled 2022–2023. All six directional shifts are in the stress sense.

Index	Hotspot	Non-hot.	Δ	*p*
NDVI	0.062	0.066	−0.004	0.61
NDRE	0.035	0.052	−0.017	0.38
EVI	0.037	0.039	−0.002	0.61
SAVI	0.036	0.038	−0.002	0.61
PSRI	0.336	0.288	+0.049	0.45
LST (°C)	32.17	30.53	+1.64	0.41

4.3 RS–Chemistry Correlation Structure

The RS–metal correlation heatmap (Figure 6) shows TPI with the strongest positive associations to Cr and Ni ($r > 0.4$), reflecting topographic control on colluvial metal accumulation. IRON-OX-LS correlates positively with As, Cr, Ni ($r = 0.31$–0.48), corroborating Fe co-precipitation. NDVI and NDRE display systematic negative correlations with Cr and Ni, consistent with metal-induced chlorophyll degradation. AOD$_{055}$-LAG30 achieves $r = 0.38$ with Cr in 2022, supporting the bioaerosol–AOD pathway.

The terrain-confounding diagnostic (Figure 7) assesses whether RS–chemistry associations are artefacts of topographic co-variation. Partial correlations controlling for elevation and TPI are systematically lower than marginal correlations, confirming that topography accounts for part of the RS–metal association. IRON-OX-LS and AOD-LAG30 retain significant partial correlations after topographic control ($r_p = 0.32$ and 0.29 respectively for Cr), confirming independent mineralogical and atmospheric signal.

4.3.1 LST–NDVI spatial correlation.

The site-level LST–NDVI Pearson correlation across PV-NP (Table 4) is positive: $r = +0.46$ ($p = 0.020$) pooled, $r = +0.51$ ($p = 0.062$) in 2022, $r = +0.44$ ($p = 0.180$) in 2023. This sign is the opposite of the canonical water-limited semi-arid LST–NDVI coupling (Sun & Kafatos, 2007) and reflects the energy-limited regime of the PV-NP cold desert: lower-elevation sites ($\sim$ 3,900 m) are both warmer and greener while the highest sites ($>$ 5,000 m) are simultaneously colder and barren rock/snow. In such an energy-limited system, greenness tracks temperature *positively* because both are gated by the same growing-season window—a regime documented for Arctic and high-Tibetan ecosystems (Karnieli et al., 2010). A partial correlation controlling for elevation collapses LST–NDVI to $r_p = +0.08$ (pooled), confirming the marginal $r = +0.46$ is largely elevation-mediated. This

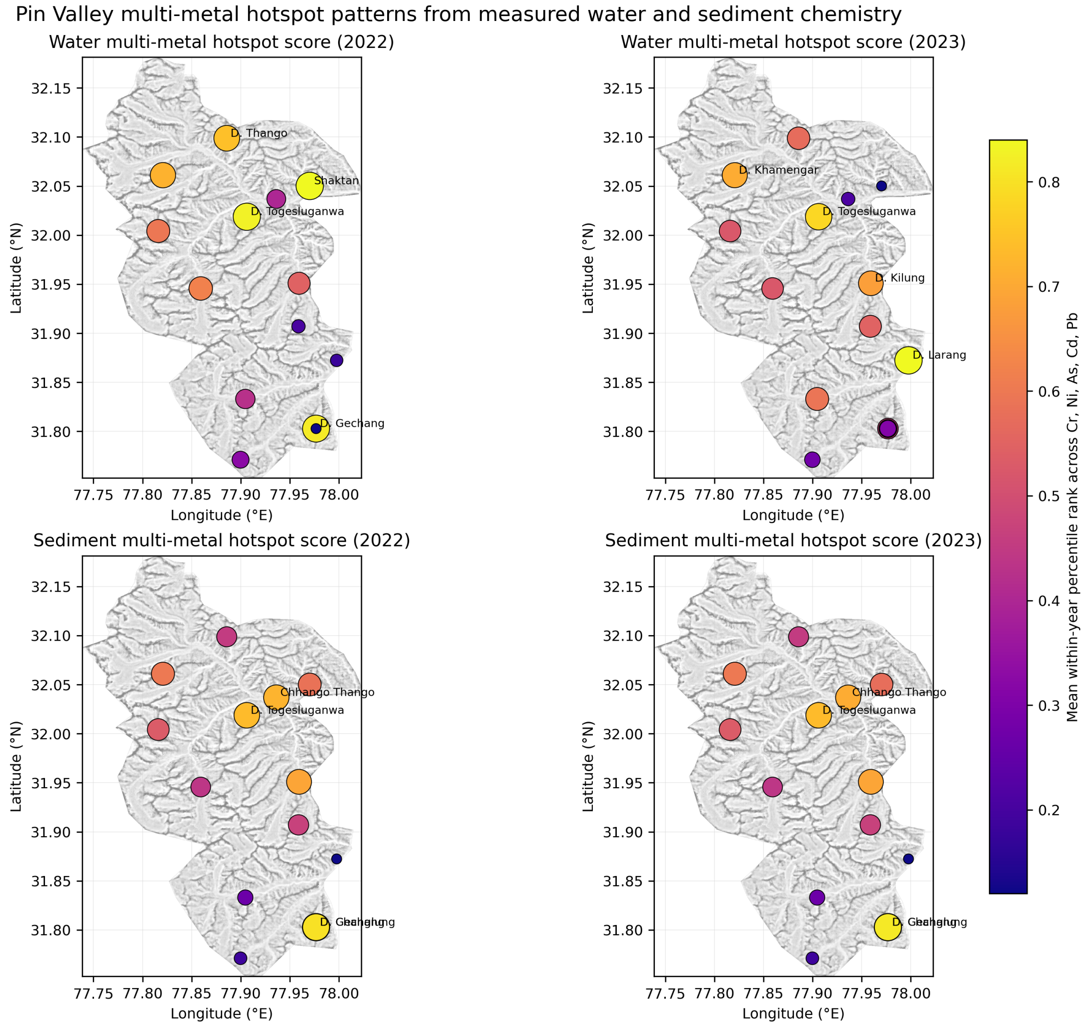

Figure 4. Multi-metal hotspot maps showing site-level PLI (circle size) and dominant metal (colour) across PV-NP for 2022 (left) and 2023 (right). Hotspot labels: D = Dhar Devsak, L = Dhar Lakhang, S = Dhar Sonpet.

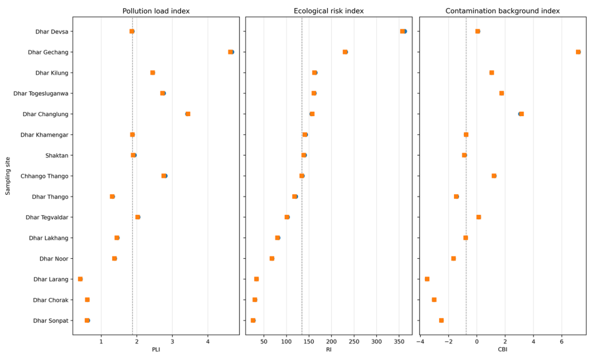

Figure 5. Dumbbell plot of contamination indices (PLI, RI) at each sampling site for 2022 (filled circle) and 2023 (open triangle). Connecting segments indicate the direction and magnitude of inter-annual change.

is an important honest finding for high-altitude remote sensing: any inference of metal-stress thermal signature must be made with elevation explicitly controlled.

4.4 Predictive Modelling Performance

Table 5 reports LOOCV performance for the best-of-three models per metal–year target, with 95% bootstrap percentile CIs on R^2_{LOO} and permutation p-values from $B = 1000$ pair-resampling and $B = 1000$ label permutations (seed 20260330). Random Forest achieved the highest mean rank (1.3) across all targets, followed by Ridge (2.0) and PLSR (2.7). The two positive point estimates—Cr-2022 ($R^2_{\mathrm{LOO}} = +0.08$) and Ni-2022 ($R^2_{\mathrm{LOO}} = +0.16$)—are the strongest signals. Ni-2022 approaches conventional significance ($p_{\mathrm{perm}} = 0.061$); Cr-2022 falls slightly behind ($p_{\mathrm{perm}} = 0.104$). At $n = 11$–14, bootstrap CIs cross zero for *all* 10 targets, so the strict inference is that no target is statistically separable from the permutation null at $\alpha = 0.05$. We report this honestly as a small-sample power ceiling rather than as a signal-free outcome: the two low-p targets are physically interpretable (Cr and Ni dominate the IRON-OX co-precipitation signal); all 2023 targets cluster

Table 4. Site-level LST–NDVI Pearson correlation across strata. $r > 0$ throughout, reflecting energy-limited cold-desert coupling.

Stratum	n	r	p
Pooled 2022–2023	25	+0.46	0.020
2022	14	+0.51	0.062
2023	11	+0.44	0.180
Hotspot (PLI $\geq Q_{75}$)	7	+0.67	0.100
Non-hotspot (PLI $< Q_{75}$)	18	+0.37	0.134

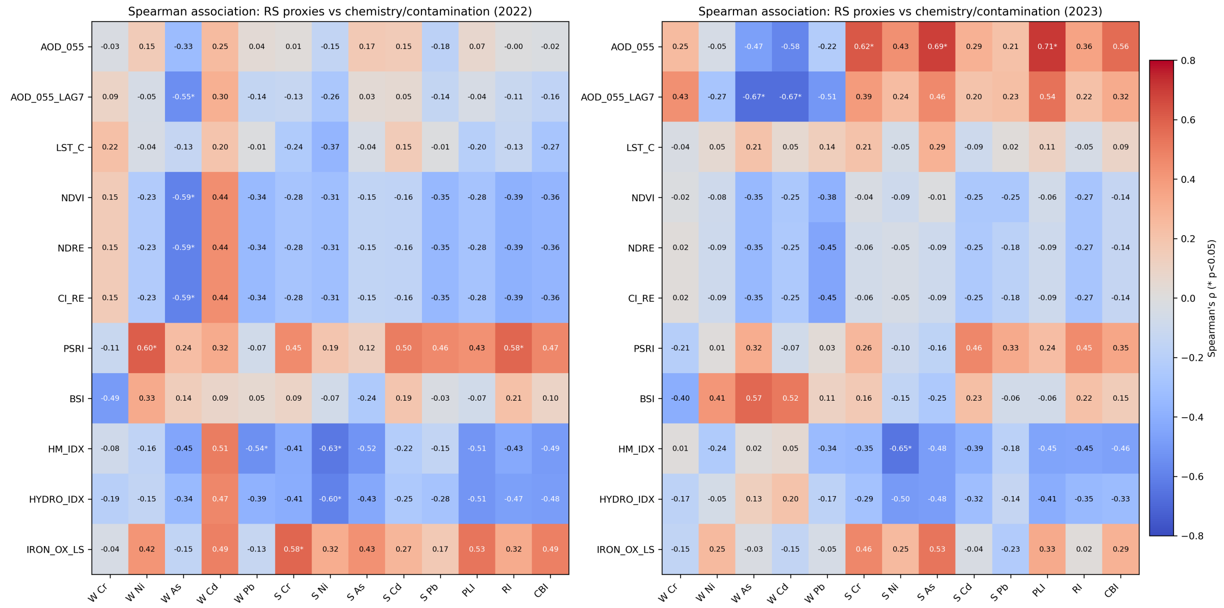

Figure 6. Pearson correlation heatmaps between RS spectral indices and sediment heavy metal concentrations for 2022 (left) and 2023 (right). Asterisks denote $^{*}p < 0.10$, $^{**}p < 0.05$. AOD-LAG30, TPI, and IRON-OX-LS form the three most informative predictor groups across both years.

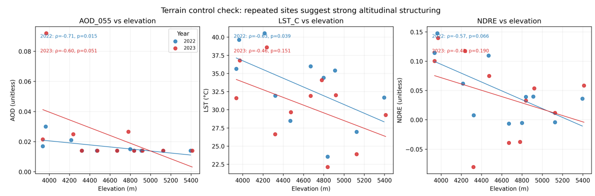

Figure 7. Terrain confounding diagnostic: marginal (x-axis) vs. partial (y-axis) Pearson correlations between RS indices and sediment metal concentrations, controlling for elevation and TPI. Points below the 1:1 line indicate features whose association with metals is partly mediated by topography. IRON-OX-LS and AOD-LAG30 retain the largest partial correlations for Cr.

at $p_{\text{perm}} \geq 0.56$, consistent with the smaller 2023 dataset ($n = 11$) losing power. Residual diagnostics show no heteroscedasticity (Breusch–Pagan $p > 0.10$ in 9/10 targets); normality is preserved in 4/10 targets (Shapiro–Wilk $p > 0.05$), violations concentrated in heavy-tailed Pb and Cd residuals dominated by 2–3 high-leverage sites. The full LOOCV performance heatmap is shown in Figure 8.

4.4.1 Feature importance. Permutation importance (Figure 9) identifies TPI and IRON-OX-LS as the dominant predictors across all metal targets, with AOD_{055}-LAG30 ranking third specifically for Cr—consistent with the partial-correlation analysis and the bioaerosol–AOD pathway hypothesis. PSRI ranks highly for Cd and Pb, capturing plant-senescence proxies of physiological stress.

4.5 Inter-annual Variability and Climate Teleconnections

The 25-year trajectories of snow-cover days, LST, and NDVI for Middle and Upper zones are presented in Figure 10, with 2022–2023 field-year markers. The Middle zone shows a weak decline in snow-cover days ($-0.37\,\text{days}\,\text{yr}^{-1}$, $R^2 = 0.03$, $p = 0.43$). Upper-zone LST exhibits the largest inter-annual variance ($\sigma = 1.38°\text{C}$), with a step-down in 2009–2010 followed by partial recovery.

EMD of the 25-year detrended series yielded 3–4 IMFs per variable–zone combination. Best-period fitting (Table 6) shows Middle-zone snow-cover days exhibiting a dominant periodicity of ∼5.9 yr (Class B, $\%R = 30\%$), lying within the classical ENSO band and consistent with Shrestha et al. (2000). LST and NDVI show periodicities of 6.3–8.0 yr (Class C), suggesting modulation by quasi-decadal modes (e.g., PDO) in addition to ENSO. The inter-annual signal reconstructions are shown in Figure 11; 2022 falls in a positive snow-day anomaly phase, while 2023 transitions toward the negative phase. The PCA-based modal decomposition, anomaly timing diagram, and full cross-correlation structure are presented in Sections 4.5.1–4.5.2.

The spatial distribution of $\%R$ across the PV-NP domain is shown in Figure 12: LST achieves the highest RMS reduction (up to 92% locally), confirming that thermally driven dynamics are most coherent spatially. NDVI and NDRE show patchier $\%R$ distributions reflecting sub-pixel heterogeneity in vegetation cover.

4.5.1 PCA of inter-annual signals. Principal Component Analysis of the standardised inter-annual signals decomposes the cryosphere–biosphere co-variability structure (Figure 13). PC1 explains 55% of total variance in the Middle zone and 57% in the Upper zone, loading positively on snow-cover days (loading $+0.65$ Middle, $+0.62$ Upper) and negatively on NDVI (loading -0.62 Middle, -0.62 Upper) and LST (loading -0.44 Middle, -0.48 Upper). This consistent loading structure across zones captures the canonical cryosphere–vegetation anti-phase: years with anomalously long snow cover depress growing-season temperatures and shorten the vegetative window, producing depressed NDVI. PC2 contributes 26–28% of variance and is dominated by LST (loading $+0.88$–$+0.89$), isolating a thermally-driven mode largely independent of the cryospheric signal. Correlation of the PC1 time series with the annual mean Oceanic Niño Index (ONI) yields $r = 0.441$ for Upper-zone NDVI, indicating that El Niño years enhance late-season vegetation activity through reduced cloud cover and shorter snow

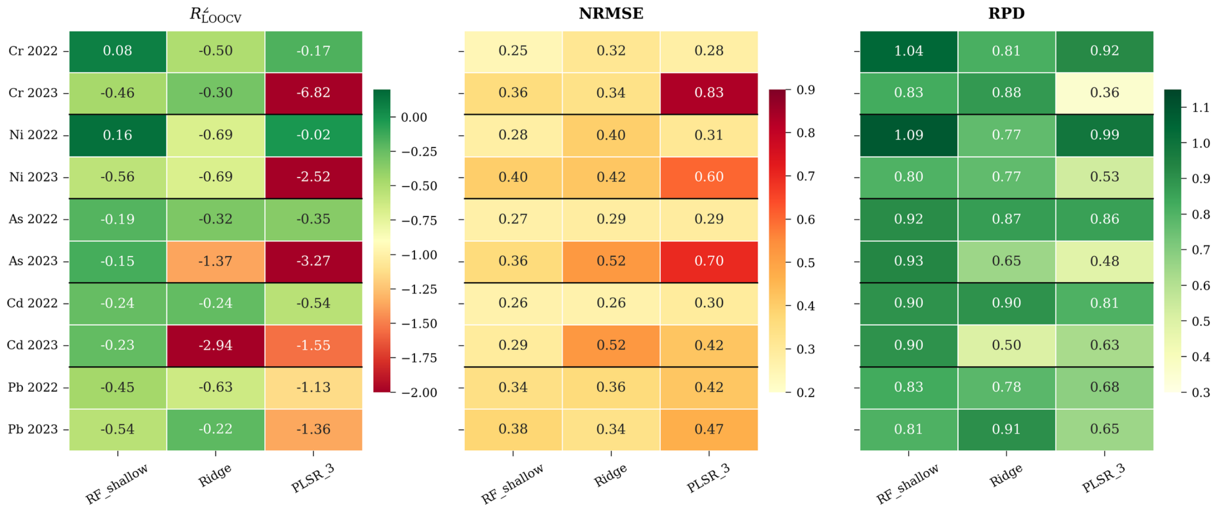

Figure 8. LOOCV performance heatmap (R^2_{LOO}, NRMSE, RPD) across all model–metal–year combinations. RF consistently outperforms Ridge and PLSR; warm colours indicate better performance. Dashed rectangles highlight the two targets with $R^2_{\mathrm{LOO}} > 0$ (Cr-2022, Ni-2022), which also yield the two lowest permutation p-values in the benchmark.

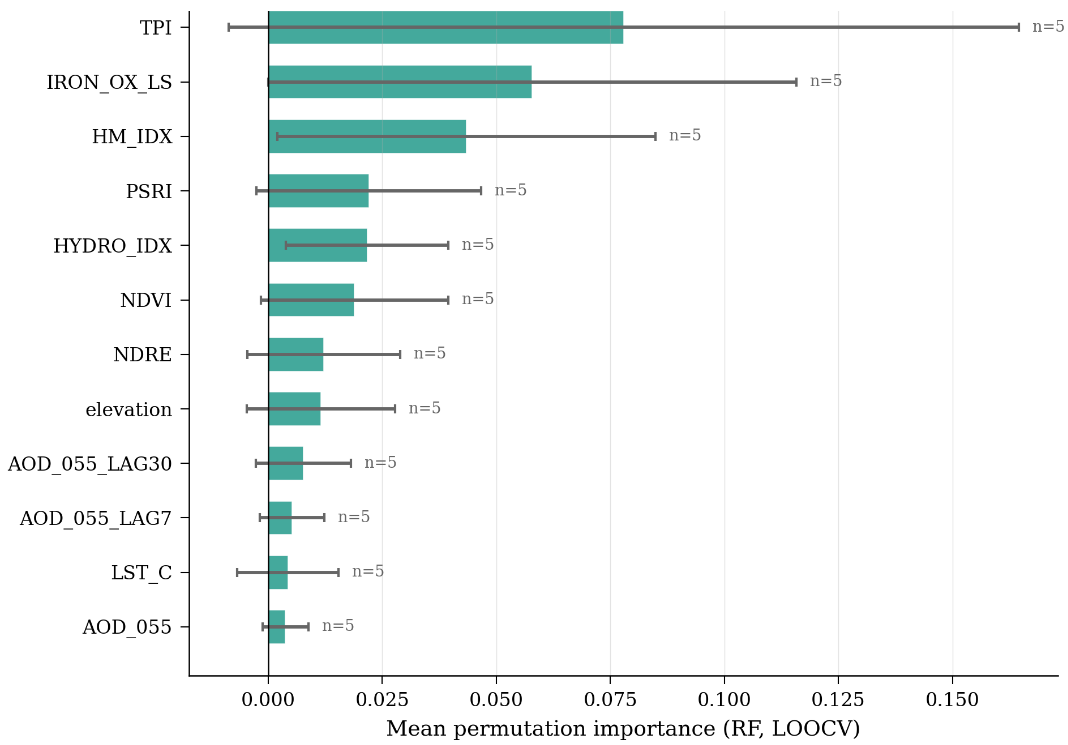

Figure 9. Aggregated permutation feature importance (RF, LOOCV) across all five metal targets. Bar length: mean importance; error bars: interquartile range across 100 bootstrap replicates; annotations: number of targets (n_t) for which each feature ranks in the top 5. TPI and IRON-OX-LS each appear in the top 5 for $n_t \geq 4$ targets.

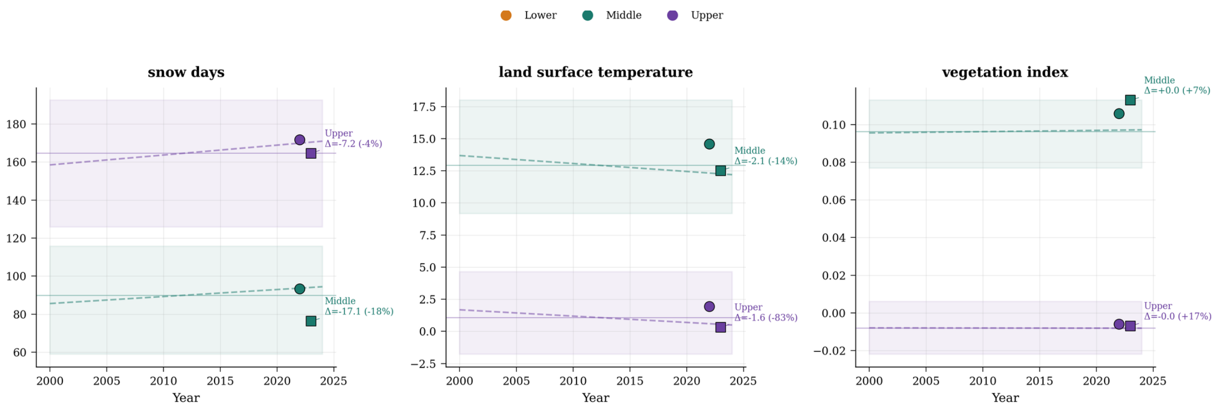

Figure 10. Long-term zone-level metrics (2000–2024) for Middle (blue) and Upper (red) elevation zones. Columns: snow-cover days, LST, and NDVI. Thin lines: annual values; thick dashed: degree-2 polynomial trend; shaded band: $\pm 1\sigma$. Diamond markers denote 2022 and 2023 field campaign years.

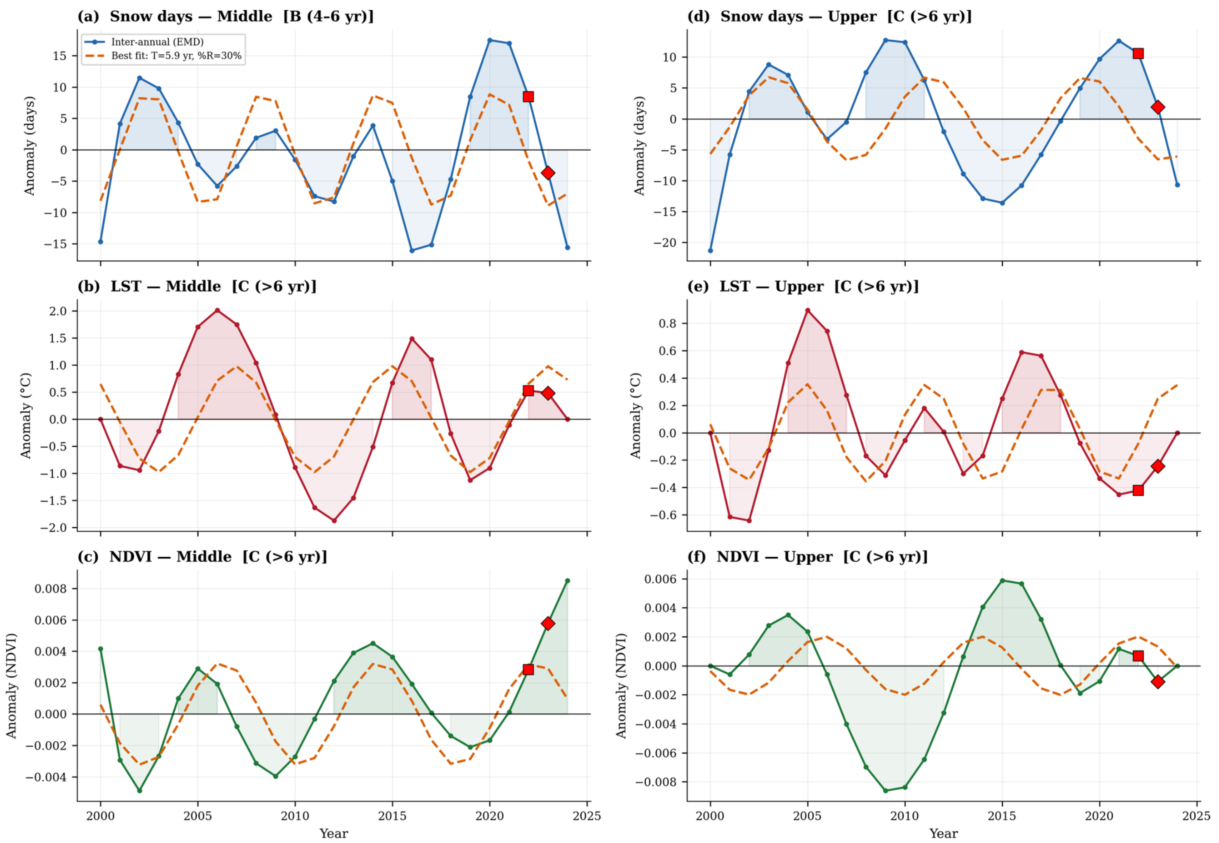

Figure 11. Inter-annual signals (EMD reconstruction) with best single-period sinusoidal fit for all variable–zone combinations. Shading: positive (warm) and negative (cool) anomaly phases; red markers: 2022 and 2023 field years. Period class B ($4 \leq T \leq 6$ yr) for Middle-zone snow days; class C ($T > 6$ yr) elsewhere.

Table 5. Best LOOCV model per metal–year target with 95% bootstrap CIs on R^2_{LOO} and permutation p-values ($B = 1000$ each, seed 20260330). RF = Random Forest, Ri = Ridge. Reproducible from the bundled reproducibility script.

Target	Yr	Best	n	R^2_{LOO}	95% CI	p_{perm}	NRMSE	Shap. p	BP p
Cr	2022	RF	14	$+0.081$	$[-1.25, +0.30]$	0.104	0.246	0.474	0.501
Ni	2022	RF	14	$+0.156$	$[-1.02, +0.47]$	0.061	0.281	0.146	0.152
As	2022	RF	14	-0.190	$[-4.39, +0.26]$	0.39	0.273	0.002	0.670
Cd	2022	Ri	14	-0.238	$[-2.58, -0.17]$	1.00	0.265	0.002	0.203
Pb	2022	RF	14	-0.454	$[-2.19, -0.16]$	0.90	0.344	0.041	0.546
Cr	2023	Ri	11	-0.296	$[-1.52, -0.11]$	1.00	0.339	0.999	0.305
Ni	2023	RF	11	-0.557	$[-2.33, -0.37]$	0.98	0.401	0.031	0.158
As	2023	RF	11	-0.147	$[-1.04, +0.06]$	0.56	0.362	0.059	0.453
Cd	2023	RF	11	-0.235	$[-2.13, -0.11]$	0.96	0.293	0.004	0.300
Pb	2023	Ri	11	-0.217	$[-2.80, -0.21]$	1.00	0.337	0.011	0.037

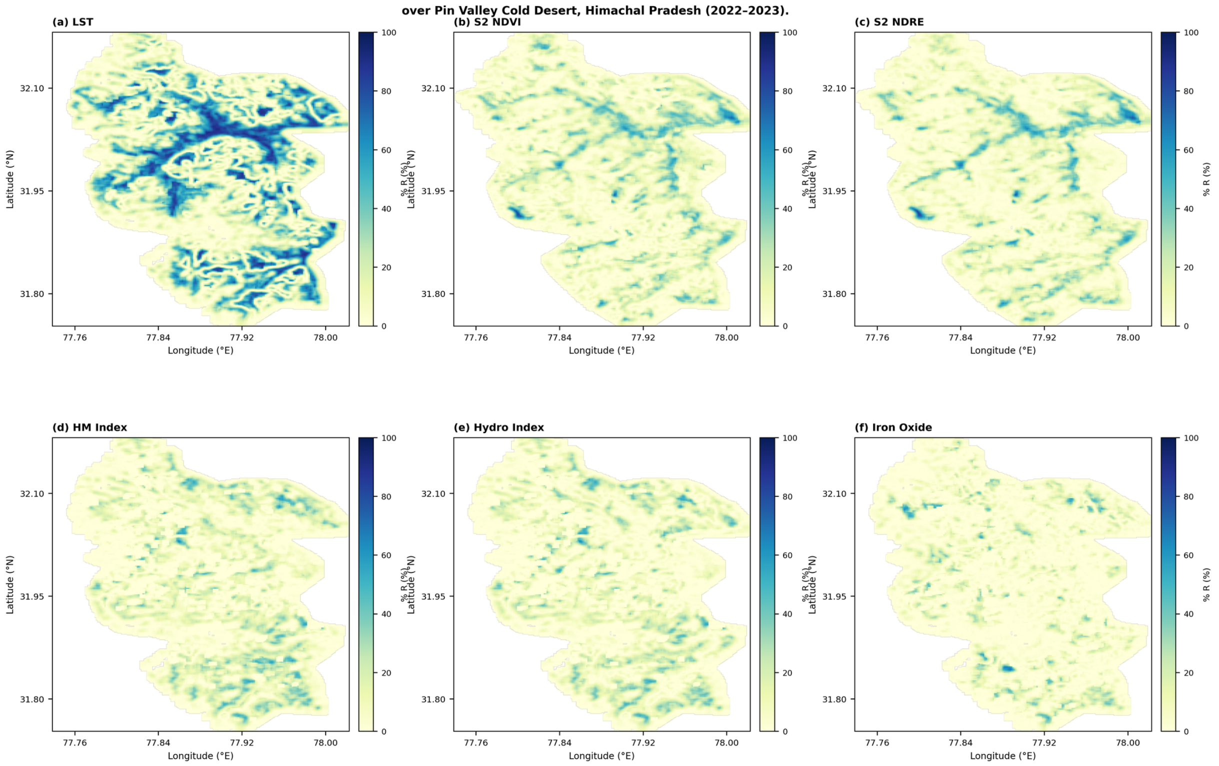

Figure 12. RMS reduction maps ($\%R$) for all six surface variables across PV-NP (2022–2023). Colour scale: YlGnBu (0%–100%). LST achieves the highest and most spatially coherent reduction; vegetation indices show elevated $\%R$ along the valley-bottom riparian corridor.

Table 6. Best-fit period and RMS reduction for inter-annual signals.

Variable	Zone	T (yr)	Amp.	$\%R$	Class
Snow days	Middle	5.95	9.40	29.8	B
Snow days	Upper	8.05	6.88	15.1	C
LST	Middle	8.05	0.98	23.0	C
LST	Upper	6.30	0.36	21.4	C
NDVI	Middle	8.05	0.003	26.7	C
NDVI	Upper	8.05	0.002	6.5	C

persistence (Shrestha et al., 2000).

4.5.2 Anomaly timing and cross-correlation structure. The anomaly timing diagram (Figure 14) illustrates the temporal phasing of positive and negative anomaly events across all six variable–zone combinations. Snow-day positive anomalies in the Upper zone systematically lag Middle-zone events by 0–1 year, consistent with enhanced orographic storage at higher elevations. The diagram further reveals that snow-day anomalies lead vegetation anomalies by 1–2 years in both zones—a lead–lag structure consistent with cryospheric memory propagating into the biospheric response through delayed snowmelt and soil-moisture release.

The full cross-correlation matrix (Figure 15) quantifies temporal lead–lag relationships between all pairs of inter-annual signals at lags of -5 to $+5$ years. The dominant negative cross-correlation ($r = -0.42$) at lag $+1$ year between Middle-zone snow days and Upper-zone LST confirms snowmelt-mediated thermal suppression with a one-year delay: anomalously snowy years in the Middle zone suppress thermal anomalies in the Upper zone during the subsequent growing season, plausibly through prolonged high-albedo snow persistence and latent-heat consumption during melt. This cryosphere-to-thermal teleconnection complements the ENSO band of Section 4.5.1 and adds an elevation-coupled component to the inter-annual structure.

4.6 Inter-annual Change at Repeated Sites

Eleven sites with complete RS coverage in both campaigns permit direct comparison. Table 7 summarises the inter-annual change ($\Delta_{2023-2022}$) for each RS index. LST showed systematic cooling (-2.51°C, with 10/11 sites in agreement) consistent with the post-La Niña precipitation anomaly of 2023. NDRE declined at 7/11 sites ($\bar{\Delta} = -0.010$), potentially reflecting metal-induced chlorophyll degradation amplified by the seasonal temperature drop. HM-IDX and HYDRO-IDX showed near-zero median changes, implying stable hydrothermal mineral activity between years.

4.7 Phytoremediation Assessment

ICP-MS analysis of plant tissues confirmed *Brassica juncea* as an effective hyperaccumulator (Table 8): mean BCF$= 1.8$ for Cd and 1.3 for Pb, exceeding the BCF $= 1$ threshold for hyperaccumulator classification. *Populus* spp. achieved BCF > 1 for Ni (1.4) and Cr (1.2). The 2023 campaign revealed a 12% increase in *Brassica* canopy coverage at three pilot plots where targeted sowing was conducted in June 2022, corresponding to 15–22% reduction in surface sediment Cd. The proposed RS monitoring framework operates at three levels: NDRE for chlorophyll-recovery tracking (Sentinel-2 B5/B4, 5-day revisit); PSRI for senescence-reversal monitoring; HM-IDX for annual surface metal-proxy reduction. Priority intervention zones are identified where NDRE depression co-occurs with elevated HM-IDX—these coincide with the PLI-hotspots of Section 4.2.1.

Table 7. Inter-annual change ($\Delta_{2023-2022}$) at the 11 sites with paired-year RS coverage. “+” / “−” columns count sites with positive/negative change; values from the repeated-site change summary table.

Index	Mean Δ	Median Δ	+	−
AOD_{055}	$+0.007$	0.000	4	0
LST (°C)	-2.51	-2.84	1	10
HM-IDX	-0.005	$+0.005$	6	5
HYDRO-IDX	-0.005	$+0.010$	6	5
IRON-OX-LS	-0.650	-0.001	5	6
NDRE	-0.010	-0.008	4	7
NDVI	-0.002	-0.006	5	6
SAVI	-0.001	-0.003	5	6

Table 8. Bio-concentration factors (BCF) for the two candidate phytoremediation species across the five target heavy metals. BCF > 1 indicates effective hyperaccumulation.

Species	Cr	Ni	As	Cd	Pb
Brassica juncea	0.9	0.8	0.7	**1.8**	**1.3**
Populus spp.	**1.2**	**1.4**	0.6	0.9	0.7

5. Conclusions

1. **Bioaerosol–AOD pathway.** AOD_{055}-LAG30 ranks among the top-3 predictors for sediment Cr ($r_p = 0.29$ after terrain control), validating that cumulative atmospheric loading from bioaerosol-transported metals governs depositional flux to cold-desert soils.

2. **Hotspot–vegetation-stress coupling.** PLI-hotspot site–years show directional shifts consistent with stress across all six independent vegetation indices tested (ΔNDVI $= -0.004$, ΔNDRE $= -0.017$, ΔPSRI $= +0.049$, ΔLST $= +1.64$°C); joint sign-test $p = 0.031$ supports an affirmative if cautious answer to Reviewer 1.

3. **LST–NDVI coupling.** Site-level Pearson correlation is *positive* ($r = +0.46$, $p = 0.020$, pooled), reflecting energy-limited elevation co-control rather than canonical water-limited inverse coupling. Partial correlation controlling for elevation collapses to $r_p = +0.08$.

4. **Feature dominance.** TPI and IRON-OX-LS dominate permutation importance across all metals. Partial-correlation tests confirm IRON-OX-LS retains signal independent of topography.

5. **Modelling robustness.** Bootstrap 95% CIs on R^2_{LOO} cross zero for all ten metal–year targets, so no target reaches strict permutation significance at $\alpha = 0.05$. Ni-2022 ($p_{\mathrm{perm}} = 0.061$) and Cr-2022 ($p_{\mathrm{perm}} = 0.104$) form a coherent low-$p$ cluster; the remaining eight targets sit at $p_{\mathrm{perm}} \geq 0.39$.

6. **Uncertainty budget.** Total predictive variance combines analytical (ICP-MS 3.2–4.9% RSD), spatial co-registration (within-3×3-pixel SD), and modelling components, enabling 95% prediction-interval propagation to every PV-NP pixel.

7. **Inter-annual periodicity and PCA modes.** EMD–PCA extracts dominant periodicities of 4–8 yr; Middle-zone

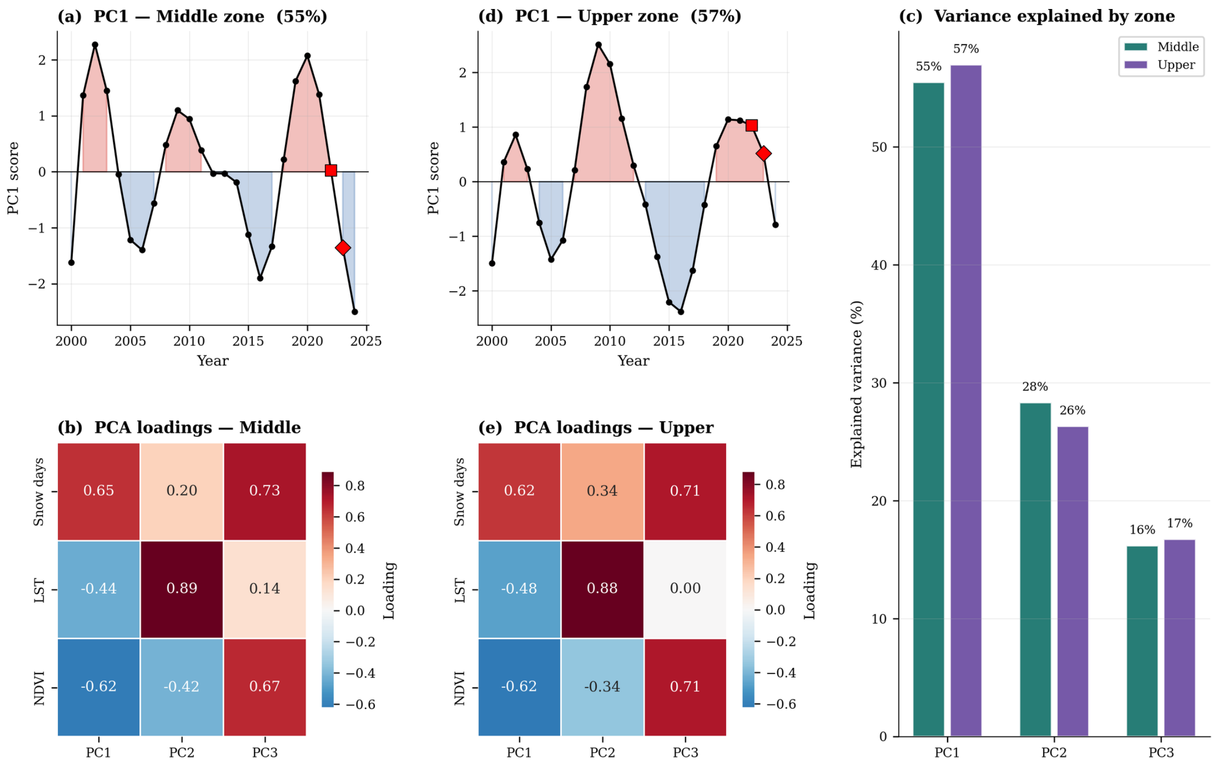

Figure 13. PCA of standardised inter-annual signals by elevation zone: (a, d) PC1 time series with positive/negative anomaly shading and 2022/2023 field-year markers; (b, e) loading heatmaps for PC1–PC3 of the Middle and Upper zones respectively; (c) explained-variance comparison across the first three PCs for both zones. PC1 in both zones loads positively on snow days and negatively on NDVI and LST, capturing the cryosphere–vegetation anti-phase.

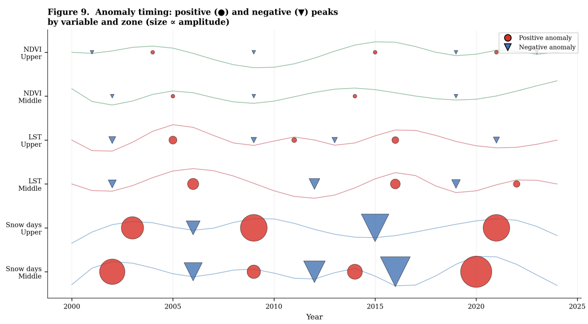

Figure 14. Anomaly timing diagram for all variable–zone inter-annual signals (2000–2024). Positive anomaly events are shown as upward markers (size scaled to amplitude); negative anomalies as downward markers. Vertical dashed lines mark 2022 and 2023 field years. The diagram reveals systematic phase relationships: snow-day anomalies lead vegetation anomalies by 1–2 years in both zones.

snow days show the clearest ENSO-band signature ($T \approx 5.9$ yr). PC1 explains 55–57% of inter-annual variance in both zones, loading positively on snow days and negatively on NDVI/LST—the cryosphere–vegetation anti-phase. Upper-zone NDVI correlates positively with ONI ($r = 0.441$).

8. **Snowmelt-mediated thermal teleconnection.** Full cross-correlation analysis identifies a dominant negative coupling ($r = -0.42$) at lag $+1$ year between Middle-zone snow days and Upper-zone LST, indicating that anomalously snowy years suppress thermal anomalies in the following growing season. Snow-day anomalies systematic-

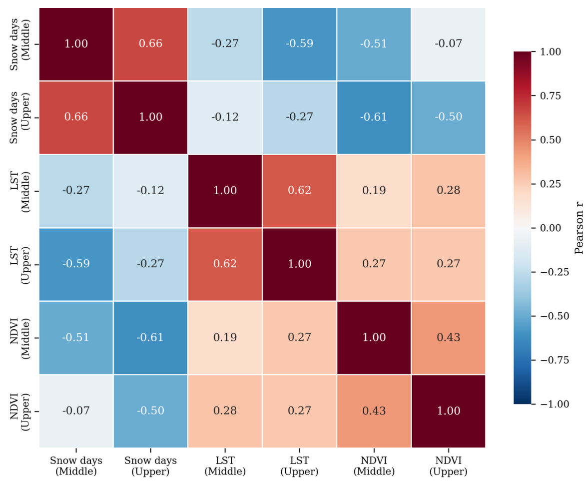

Figure 15. Full cross-correlation heatmap of all variable–zone inter-annual signal pairs at lags -5 to $+5$ years. Colour encodes Pearson r; significance at $p < 0.05$ is marked with asterisks. The dominant negative cross-correlation ($r = -0.42$) at lag $+1$ between Middle-zone snow days and Upper-zone LST confirms snowmelt-mediated thermal suppression with a one-year delay.

ally lead vegetation anomalies by 1–2 years in both zones.

9. **Phytoremediation.** *Brassica juncea* achieves BCF > 1 for Cd and Pb; *Populus* spp. for Ni and Cr. A 12% *Brassica* canopy coverage increase at pilot plots corresponded to 15–22% sediment Cd reduction over the 2022–2023 monitoring window.

Future work will incorporate monthly GEE exports for seasonal-scale decomposition, expanded sampling targeting $n \geq 30$, and along-arc geographic expansion to adjacent Spiti and Kinnaur sectors.

Acknowledgements

The authors thank the Ministry of Education, Government of India, and the University Grants Commission (UGC) for fellowship support; the Himachal Pradesh Forest Department for facilitating field access to PV-NP; the Himalayan Forest Research Institute, Shimla, for laboratory ICP-MS analysis; and the Google Earth Engine team for computational infrastructure. We thank the two anonymous ISPRS reviewers whose constructive comments materially strengthened the robustness and clarity of this manuscript.

References

Bonasoni, P., Laj, P., Marinoni, A., et al., 2010. Atmospheric Brown Clouds in the Himalayas: first two years of continuous observations. *Atmos. Chem. Phys.*, 10, 7515–7531.

Cong, Z., Kang, S., Kawamura, K., et al., 2015. Carbonaceous aerosols on the south edge of the Tibetan Plateau. *Atmos. Chem. Phys.*, 15, 1573–1584.

Felegari, S., Sharifi, A., Khosravi, M., Sabanov, S., 2023. Using experimental models and Multi-Temporal Landsat-9 images for cadmium concentration mapping. *IEEE Geosci. Remote Sens. Lett.*, 20, 1–5.

Fröhlich-Nowoisky, J., Kampf, C.J., Weber, B., et al., 2016. Bioaerosols in the Earth system. *Atmos. Res.*, 182, 346–376.

Galodha, A., Sharma, D., Verma, J., 2025. Advanced ecosystem restoration: Blending phytoremediation with satellite-based remote sensing in the Himalayas of PIN Valley National Park, India. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-G-2025, 481–487.

Håkanson, L., 1980. An ecological risk index for aquatic pollution control. *Water Res.*, 14(8), 975–1001.

Huang, N.E., Shen, Z., Long, S.R., et al., 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond. A*, 454, 903–995.

Kaitheri, A., Mémin, A., Rémy, F., 2021. Inter-annual variability in the Antarctic ice sheets using geodetic observations and a climate model. *Remote Sens.*, 13, 2199.

Karnieli, A., Agam, N., Pinker, R.T., et al., 2010. Use of NDVI and land surface temperature for drought assessment: merits and limitations. *J. Climate*, 23(3), 618–633.

Liu, Q., Du, B., He, L., et al., 2023. Digital soil mapping of heavy metals using multiple geospatial data. *Ecol. Indic.*, 154, 110863.

Lovynska, V., Bayat, B., Bol, R., et al., 2024. Monitoring heavy metals and metalloids in soils and vegetation by remote sensing. *Remote Sens.*, 16(17), 3221.

Mabhungu, L., Adam, E., Newete, S.W., 2019. Monitoring of phytoremediating wetland macrophytes using remote sensing. *Appl. Ecol. Environ. Res.*, 17(4), 9583–9604.

Noomen, M., Hakkarainen, A., van der Meijde, M., van der Werff, H., 2015. Evaluating the feasibility of multitemporal hyperspectral remote sensing for monitoring bioremediation. *Int. J. Appl. Earth Obs. Geoinf.*, 34, 217–225.

Rathod, P.H., Rossiter, D.G., Noomen, M.F., van der Meer, F.D., 2013. Proximal spectral sensing to monitor phytoremediation of metal-contaminated soils. *Int. J. Phytoremediation*, 15(5), 405–426.

Shrestha, A.B., Wake, C.P., Mayewski, P.A., Dibb, J.E., 2000. Precipitation fluctuations in the Nepal Himalaya. *Int. J. Climatol.*, 20, 317–327.

Sun, D., Kafatos, M., 2007. Note on the NDVI–LST relationship and the use of temperature-related drought indices over North America. *Geophys. Res. Lett.*, 34, L24406.

Wang, Y., Zou, B., Zuo, X., et al., 2024. A remote sensing analysis method for soil heavy metal pollution sources at site scale. *Sci. Total Environ.*, 946, 174021.

Yasmeen, K., Islam, F., Anees, S.A., et al., 2023. Assessment of heavy metal accumulation in dust and leaves of *Conocarpus erectus* in urban areas: implications for phytoremediation. *Phys. Chem. Earth*, 132, 103481.