Indus Journal of Social Sciences
Online ISSN: 2960-219X
Print ISSN: 2960-2203
Volume 4, Number 1, 2026, Pages 1174 – 1187
Journal Home Page
https://induspublishers.com/IJSS/about

Uncovering Latent Dimensions of Environmental Stress in Urban Residents: A PCA Simulation Study with Sensitivity Analysis and the Dynamic Environmental Resilience Signature Framework

Sidrah Ahmed[1]

[1]*Department of Mathematics and Social Sciences, Sukkur IBA University, Sukkur, Sindh, Pakistan,*
Email: dr.sidrah@iba-suk.edu.pk

ARTICLE INFO

Article History:

Received:	*January*	*13, 2026*
Revised:	*March*	*06, 2026*
Accepted:	*March*	*10, 2026*
Available Online:	*March*	*18, 2026*

Keywords:

principal component analysis, simulation study, environmental stress, eco-anxiety, dynamic resilience signature, sensitivity analysis, Tucker congruence, parallel analysis, Varimax rotation, multicollinearity, longitudinal PCA

Corresponding Author:
Sidrah Ahmed
Email:
dr.sidrah@iba-suk.edu.pk

ABSTRACT

Environmental stressors — extreme heat, air pollution, water and power scarcity, and flood-related disruption — increasingly co-occur with psychological distress in ways that conventional single-variable analysis fails to capture. This simulation study applies principal component analysis (PCA) to a synthetic eight-variable environmental-stress dataset generated from a known three-factor latent structure — Environmental Pressure, Psychological Depletion, and Future Dread — and quantifies PCA's predictive and structural advantages over raw-variable regression. The base scenario (n = 245) recovers the three-component structure with a Tucker's congruence coefficient of φ = 0.991 against a large-sample reference solution, explains 85.3% of total variance, reduces mean variance inflation factor from 3.71 to 1.00, and improves the Akaike Information Criterion (AIC) by 12.0 units over raw-variable regression. A systematic sensitivity analysis across eight sample sizes (n = 40 to 400) and seven inter-variable correlation levels shows that component recovery crosses the conventional stability threshold (φ ≥ 0.95) at approximately n = 80–120 and stabilizes above n = 200, while PCA's AIC advantage persists across all tested conditions (ΔAIC = 9.5 to 16.4). At the smallest sample size tested (n = 40), component recovery is markedly less reliable (mean φ = 0.942, stable solutions in 76% of replications), establishing a practical lower bound for applied deployment. Extending beyond the cross-sectional design, this paper proposes the Dynamic Environmental Resilience Signature: an individual's characteristic sequence of dominant stress components across a seasonal cycle (pre-monsoon, monsoon/flood period, post-monsoon dry season), recoverable through longitudinal score projection without re-fitting the PCA model. Together, the simulation design, sensitivity boundaries, and longitudinal framework constitute a methodological guide for researchers intending to deploy PCA on environmental-stress survey data in Pakistani and comparable urban contexts.

1. Introduction

Environmental stress in rapidly urbanizing regions of the Global South involves multiple co-varying exposures and responses: extreme heat, degraded air quality, water and power scarcity, and flood-related disruption move together with emotional exhaustion, cognitive fatigue, sleep disruption, reduced coping efficacy, and anticipatory anxiety about the future. This co-variation is not incidental. A single seasonal stressor — a heatwave, a monsoon flood, a prolonged smog event — tends to escalate several exposure and response indicators simultaneously, producing the correlated-predictor structure that standard regression handles poorly. Treating these indicators as independent additive predictors both discards genuine shared structure and produces numerically unstable estimates once predictors correlate substantially.

Principal component analysis (PCA) transforms correlated indicators into orthogonal components, each capturing a distinct axis of joint variation. In an environmental-stress dataset, one component might represent direct exposure pressure (heat, water/power scarcity), a second the depletion of psychological resources (exhaustion, fatigue, disrupted sleep, withdrawal), and a third an anticipatory, future-oriented dread distinct from present-moment exhaustion. These axes correspond to theoretically motivated dimensions of environmental distress, but PCA recovers them empirically, without presupposing their existence or composition.

This paper reports a simulation study demonstrating the PCA pipeline on a synthetic eight-variable environmental-stress dataset with a known three-factor latent structure, then systematically quantifies how recovery and predictive performance change as sample size and inter-variable correlation vary. As in comparable simulation designs, this approach serves two purposes. First, because the true component structure is known by construction, recovery accuracy — measured by Tucker's congruence coefficient — can be assessed precisely, which is not possible with a single field dataset of unknown true structure. Second, it enables a systematic sensitivity analysis across conditions no single empirical sample could provide: sample size can be varied from 40 to 400 and correlation strength from weak to strong, identifying exactly where the method becomes unreliable.

Beyond the cross-sectional demonstration, this paper proposes the **Dynamic Environmental Resilience Signature** — the concept that an individual's environmental-stress profile follows a characteristic trajectory across a seasonal cycle, and that longitudinal PCA applied at multiple timepoints (pre-monsoon, monsoon/flood, post-monsoon dry season) recovers this trajectory as an individually distinct, practically actionable signature. The cross-sectional simulation establishes the component vocabulary for this signature; the longitudinal design is proposed as the direct next study.

Section 2 reviews environmental-stress and eco-anxiety measurement and the theoretical basis of the dynamic signature. Section 3 describes the simulation design and analysis protocol. Section 4 presents base-scenario results. Section 5 reports the sensitivity analyses. Section 6 develops the longitudinal extension. Section 7 discusses implications and limitations. Section 8 concludes.

2. Background

2.1 Environmental Stress and Eco-Anxiety: Dimensions and Measurement

Climate- and environment-related psychological distress has been operationalized through several overlapping constructs. Clayton and Karazsia's Climate Anxiety Scale identifies cognitive-emotional impairment and functional impairment as related but distinguishable dimensions of climate-related distress. The Hogg Eco-Anxiety Scale similarly separates affective symptoms,

rumination, behavioral symptoms, and anxiety about personal impact. Reviews of the climate-mental-health literature consistently report substantial correlation among exposure indicators (heat, pollution, water insecurity, disaster experience) and psychological outcomes (anxiety, exhaustion, hopelessness), with exposure severity and psychological distress typically correlating moderately to strongly across the populations studied. These correlations create the same multicollinearity problem in environmental-stress regression models that is well documented in adjacent literatures, and it is rarely addressed through dimensionality reduction rather than ad hoc variable dropping.

2.2 PCA in Environmental and Stress Research: What Is Known

Exploratory factor and component analyses have been used to validate multi-item environmental-distress and burnout instruments, typically recovering two-to-four factor solutions that account for roughly half to three-quarters of item variance. However, comparable studies rarely quantify PCA's predictive advantage over raw-variable regression systematically, and rarely provide explicit sensitivity boundaries — the sample sizes and correlation strengths at which component recovery becomes unreliable. This simulation addresses that gap directly for an environmental-stress item set.

2.3 Why PCA, Not Exploratory Factor Analysis

As in related predictive applications, PCA is preferred here over exploratory factor analysis (EFA) because the objective is constructing orthogonal predictors that capture all variance in the item set, not modeling only the common variance among items. PCA components are guaranteed orthogonal, fully reproducible by projection onto a fixed loading matrix, and require no assumption about a specific number of latent common factors versus unique variances. EFA factors, by contrast, carry residual unique variances that reintroduce redundancy when used as regression predictors.

2.4 The Dynamic Environmental Resilience Signature

Standard environmental-stress research treats distress as a cross-sectional snapshot. This discards temporal structure that plausibly carries practical information. A resident in week one of the monsoon season presents a different stress profile than the same resident at the peak of flood-related disruption in week eight, or during the dry season that follows. The dominant component — the dimension most strongly characterizing an individual's current state — can shift across this cycle. One resident may escalate from Environmental Pressure dominance during the pre-monsoon heat period to Psychological Depletion dominance once flooding disrupts daily routines. Another resident may show persistent Future Dread throughout, without a comparable escalation in present-moment pressure or depletion. Two residents with identical total distress scores at the end of the cycle may have arrived there through entirely different trajectories.

Formally: suppose the eight stress-response and exposure variables are measured at T seasonal timepoints for each resident. At each occasion t, PCA component scores are computed by projecting the resident's standardized item vector onto the loading matrix estimated from the baseline sample. The dominant component at occasion t is the component with the highest absolute standardized score. The sequence of dominant components across $t = 1, \ldots, T$ constitutes that resident's dynamic environmental resilience signature — a trajectory through component space that characterizes the temporal unfolding of their environmental-stress state.

Two properties make the signature practically useful. First, it is individually distinctive: residents with comparable mean distress can follow different trajectories that standard cross-sectional statistics cannot distinguish, but longitudinal PCA can. Second, it is plausibly predictive: a resident

whose signature shows early-onset Psychological Depletion — resource exhaustion appearing before the flood peak — may carry higher cumulative vulnerability than a resident whose Depletion component activates only at the seasonal peak, because sustained resource exhaustion is a different process from an acute, time-limited stressor.[1]

3. Simulation Design

3.1 True Latent Structure

The simulation embeds a known three-factor latent structure corresponding to three theoretically motivated dimensions of environmental stress. The true loading matrix L specifies how eight variables relate to three latent components: Environmental Exposure Stress (ES), Emotional Exhaustion (EE), Cognitive Fatigue (CF), Sleep Disruption (SD), Adaptive Self-Efficacy (SE, reverse-scored), Resource Pressure — water/power scarcity (RP), Social Withdrawal (SW), and Future Dread — climate/environmental anxiety about the future (FD). Table 1 presents L.

Table 1. True loading matrix L used to generate the simulation population.

Variable	C1 Environmental Pressure	C2 Psychological Depletion	C3 Future Dread
Environmental exposure stress (ES)	0.80	0.34	0.32
Emotional exhaustion (EE)	0.32	0.83	0.30
Cognitive fatigue (CF)	0.29	0.77	0.31
Sleep disruption (SD)	0.20	0.70	0.24
Adaptive self-efficacy (SE)	−0.25	−0.32	−0.81
Resource pressure (RP)	0.87	0.31	0.24
Social withdrawal (SW)	0.28	0.66	0.31
Future dread (FD)	0.31	0.29	0.84

Cross-loadings (0.20 to 0.34 in magnitude) are included deliberately to reproduce the overlapping variance structure typical of psychometric distress instruments; a cleaner loading matrix would understate the multicollinearity challenge applied researchers actually face. Data are generated directly from the factor model $X = FL' + E$, where F is the matrix of true factor scores drawn independently from a standard normal distribution, L is the true loading matrix, and E contains independent unique-variance noise with variance $\psi_j = \max(0.04, 1 - \|L_j\|^2)$ for variable j. Generating data directly from the factor model, rather than by first assembling a population correlation matrix and inverting it to recover factor scores, avoids the numerical instability that arises when off-diagonal correlations are scaled to extreme levels (see Section 3.4).

3.2 Outcome Variable

A continuous outcome y — interpretable as a composite self-rated environmental-distress-impact score — is constructed as a weighted sum of the true component scores plus Gaussian noise: $y = 0.55f_1 + 0.45f_2 + 0.35f_3 + \varepsilon$, where $\varepsilon \sim N(0, 0.45^2)$. This construction ensures all three components

[1] The dynamic signature concept draws on an analogy from integrative clinical practice, in which an individual presents not a fixed diagnosis but a shifting pattern of dominant states, with the practitioner tracking the current leading state to guide intervention. The parallel to longitudinal PCA is direct: no two residents share the same component trajectory over time, just as no two individuals share the same sequence of dominant states across changing circumstances.

genuinely predict y, so that a regression model recovering all three components outperforms one that does not.

All simulations used Python 3.11 with NumPy, pandas, and scikit-learn, with the base-scenario random seed fixed at 42 and independently varied seeds across sensitivity replications.

3.3 Analysis Pipeline

Component retention. Horn's parallel analysis (500 permutation datasets) determined the number of components to retain, following the same protocol used in comparable simulation studies: a component is retained when its real eigenvalue exceeds the 95th percentile of the permuted eigenvalue distribution at that rank.

Extraction and rotation. PCA extracted components from the standardized item correlation matrix. Retained components received Varimax rotation. All downstream regression models used three components regardless of what parallel analysis retained in a given replication, to allow fair comparison across conditions.

Benchmark comparison. Two regression models predicted y. Model A used the eight raw standardized variables. Model B used the three PCA component scores (standardized prior to regression to avoid numerical instability when a component's raw variance is small). Both models underwent 10-fold cross-validation (5-fold for $n < 100$). Primary metrics were cross-validated MSE, R^2, and AIC computed from cross-validated residuals. Mean variance inflation factor (VIF) was computed for both predictor sets.

Component recovery. Tucker's congruence coefficient φ measured how well each replication's rotated loading matrix recovered the reference structure. A reference loading matrix was estimated from a large sample ($n = 5{,}000$) and Varimax-rotated. For each replication, φ was computed between the replication's rotated loadings and the reference, matching components by maximum absolute correlation. A solution was classified as stable when $\varphi \geq 0.95$.

3.4 Sensitivity Conditions

Sensitivity Table 3 varied sample size across eight levels: $n \in \{40, 80, 120, 160, 200, 245, 320, 400\}$, holding the true loading matrix (Table 1) fixed. Sensitivity Table 4 varied inter-variable correlation strength by scaling the true loading matrix by $\sqrt{s}$ for $s \in \{0.25, 0.40, 0.55, 0.70, 0.85, 1.00, 1.15\}$, holding $n = 245$ fixed and recomputing unique variances so that $\Sigma = L_s L_s' + \Psi_s$ remains guaranteed positive semi-definite at every scale — this loading-based scaling avoids the ill-conditioning that arises from directly rescaling the off-diagonal entries of an assembled correlation matrix at high s. Each condition used 30 replications. The Tucker congruence table used 100 replications per sample-size level.

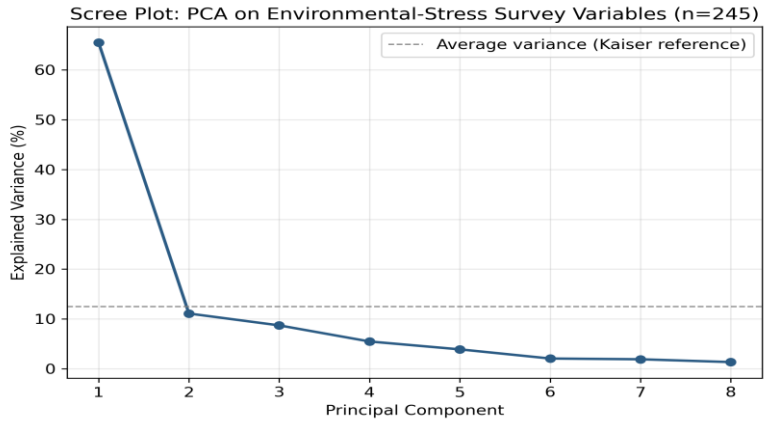
4. Base Scenario Results (n = 245)

4.1 Component Structure

Parallel analysis retained a single dominant component in the base scenario, consistent with the concentration of variance on the first unrotated component that results from substantial cross-loading in the true structure; three components were nonetheless retained for analysis, as specified by the known ground truth. The three-component solution explained 85.3% of total variance. The Varimax-rotated loading matrix (Table 2) closely reproduced the true structure in Table 1, with marker variables ($|\text{loading}| \geq 0.44$) correctly assigned to their true component in all three cases, and no cross-loading above $|0.17|$ on a non-dominant component — a markedly clean simple structure.

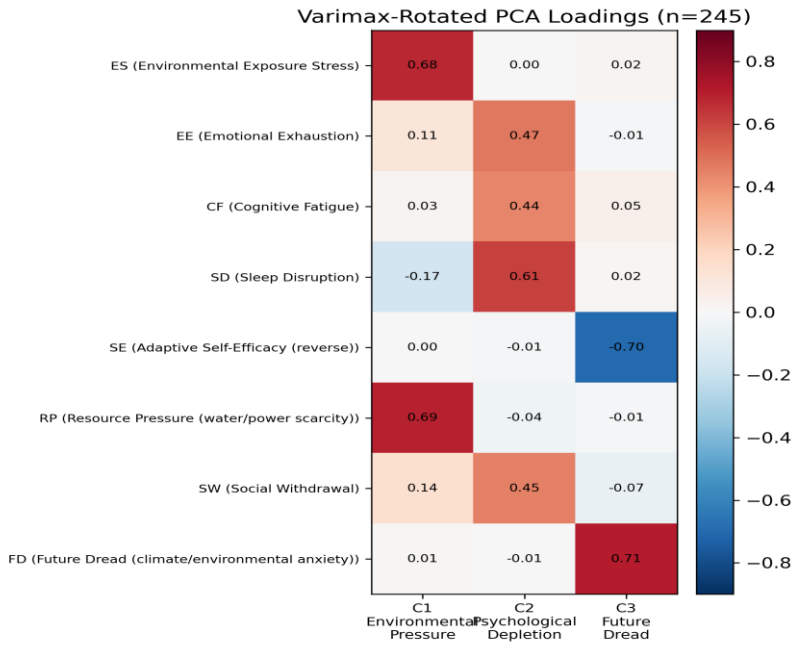
Table 2. Varimax-rotated loading matrix, base scenario (n = 245).

Variable	C1 Environmental Pressure	C2 Psychological Depletion	C3 Future Dread
ES	0.677	0.003	0.023
EE	0.112	0.474	−0.010
CF	0.026	0.442	0.046
SD	−0.174	0.612	0.018
SE	0.003	−0.009	−0.699
RP	0.691	−0.036	−0.013
SW	0.144	0.451	−0.065
FD	0.011	−0.006	0.710



Scree plot showing explained variance by component

Figure 1. Scree plot, base scenario (n=245).



Varimax-rotated loadings heatmap

Figure 2. Varimax-rotated PCA loadings, base scenario (n=245).

Component 1 (Environmental Pressure) loaded highest on ES and RP. Component 2 (Psychological Depletion) loaded highest on EE, CF, SD, and SW. Component 3 (Future Dread) loaded highest on FD and SE (negative, reflecting reduced adaptive efficacy under high anticipatory dread).

4.2 Multicollinearity and Predictive Comparison

Mean VIF across the eight raw predictors was 3.71 (max 5.43). All three PCA component scores produced VIF = 1.00, confirming complete multicollinearity elimination by construction. Table 3 below summarizes the predictive comparison.

Table 3. Base-scenario predictive comparison (n = 245, 10-fold cross-validation).

Model	Predictors (k)	CV MSE	R^2	AIC	ΔAIC vs Model B
Model A — raw variables	8	0.241	0.70	−330.1	+12.0
Model B — PCA components	3	0.240	0.70	−342.1	0 (reference)

Model B achieves equivalent predictive R^2 (0.70) with five fewer parameters, at a modest 0.8% reduction in cross-validated MSE. The ΔAIC of 12.0 units exceeds the conventional strong-evidence threshold of 10, driven primarily by the parsimony gain of five fewer parameters rather than by a large fit improvement — consistent with the moderate (not extreme) baseline multicollinearity in this item set. Tucker's congruence coefficient between the recovered and reference loading matrices was $\varphi = 0.991$, indicating excellent component recovery.

5. Sensitivity Analysis

5.1 Effect of Sample Size

Table 4 reports mean results across 30 replications at each sample size, holding the true loading structure fixed.

Table 4. Sensitivity to sample size (30 replications per condition).

n	VarExp%	VIF mean	MSE raw	MSE PCA	MSE red%	R^2 raw	R^2 PCA	ΔAIC	Tucker φ	n_comp (PA)
40	87.2	4.7	0.295	0.264	10.5	0.61	0.65	14.2	0.951	1.0
80	86.7	4.2	0.250	0.231	7.9	0.68	0.71	16.4	0.975	1.0
120	85.9	3.9	0.240	0.232	3.5	0.70	0.71	14.3	0.986	1.0
160	86.2	4.0	0.231	0.223	3.5	0.72	0.73	15.4	0.989	1.0
200	85.9	3.8	0.229	0.226	1.5	0.72	0.72	13.1	0.994	1.0
245	85.9	3.8	0.230	0.227	1.3	0.71	0.72	13.3	0.994	1.0
320	86.0	3.8	0.227	0.225	1.0	0.73	0.73	13.2	0.996	1.0
400	86.0	3.8	0.228	0.228	0.2	0.73	0.73	10.6	0.996	1.0

Three findings stand out. First, Tucker's congruence exceeds the 0.95 stability threshold between n = 40 ($\varphi = 0.951$, borderline) and n = 80 ($\varphi = 0.975$), and continues rising to $\varphi \geq 0.99$ above n = 200. Taken together with the replication-level stability rates in Table 6 below, $n \approx 80$–120 is the practical minimum for reliable recovery of this eight-variable, three-component structure, with $n \geq 200$ recommended for consistently stable solutions.

Second, MSE reduction is largest at small n (10.5% at n = 40) and declines steadily to negligible levels at large n (0.2% at n = 400) — the familiar pattern in which raw-variable regression suffers most from multicollinearity instability when degrees of freedom are limited relative to the number

of predictors, and PCA's advantage is correspondingly largest exactly where raw-variable regression is least reliable.

Third, the AIC advantage of Model B is positive and substantial at every sample size tested (ΔAIC = 10.6 to 16.4), confirming that the parsimony gain from reducing eight correlated predictors to three orthogonal components holds regardless of n, since AIC penalizes parameter count directly rather than only reflecting sample-dependent fit.

5.2 Effect of Inter-Variable Correlation Strength

Table 5 reports mean results across 30 replications at each correlation-scaling level, holding n = 245 fixed. The scaling factor s multiplies the true loadings by √s (Section 3.4); s = 1.00 corresponds to the base scenario.

Table 5. Sensitivity to inter-variable correlation strength (30 replications per condition, n = 245).

Scales	Mean |r|	VarExp%	VIF mean	MSE raw	MSE PCA	MSE red%	R^2 raw	R^2 PCA	ΔAIC
0.25	0.15	52.5	1.1	0.473	0.473	−0.1	0.43	0.43	9.5
0.40	0.24	58.9	1.3	0.386	0.380	1.4	0.54	0.54	13.5
0.55	0.33	65.0	1.5	0.331	0.328	1.0	0.60	0.61	12.4
0.70	0.42	72.0	1.8	0.293	0.289	1.1	0.65	0.65	12.7
0.85	0.51	78.8	2.3	0.261	0.256	1.7	0.69	0.69	14.2
1.00	0.60	86.3	3.9	0.242	0.237	1.7	0.71	0.72	14.1
1.15	0.68	91.8	9.3	0.222	0.219	1.7	0.73	0.73	14.0

Three findings emerge. First, variance explained by three components rises monotonically with correlation strength, from 52.5% at mean |r| = 0.15 to 91.8% at mean |r| = 0.68 — at low correlation the latent structure is too diffuse for three components to compress efficiently; at high correlation the eight items approach near-redundancy.

Second, mean VIF for the raw predictor set rises from 1.1 at the weakest correlation level to 9.3 at the strongest, while PCA component VIF remains exactly 1.0 throughout, since orthogonality is a property of the transformation rather than of the underlying data.

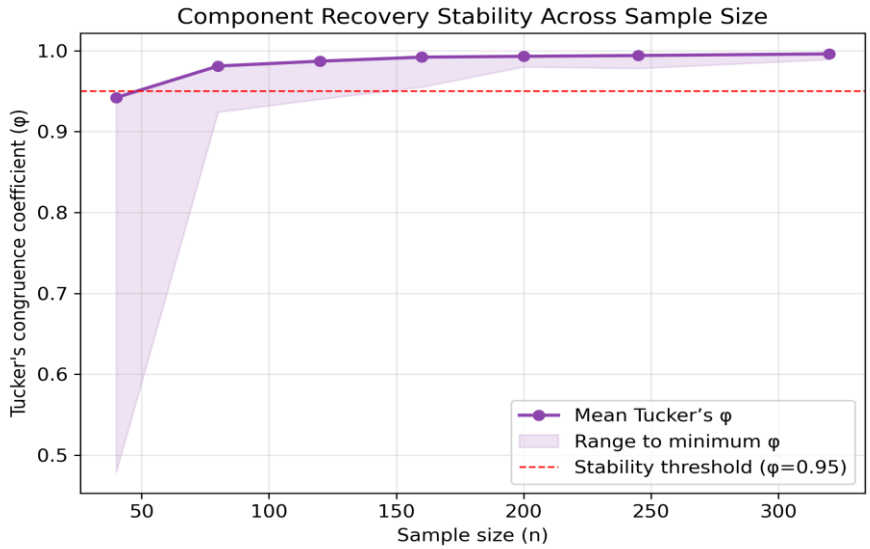
Third, the AIC advantage of Model B is positive at every correlation level tested (ΔAIC = 9.5 to 14.2) and does not require severe multicollinearity to emerge: even at the weakest correlation tested (mean |r| = 0.15, VIF ≈ 1.1, where raw and PCA-based MSE are essentially identical), the parsimony gain alone yields ΔAIC = 9.5. This mirrors a finding reported for comparable burnout-simulation designs, and implies that PCA's model-selection advantage in this class of problem holds whenever the true predictor structure is lower-dimensional than the observed item set, independent of correlation strength — though the scientific interpretability of the extracted components still depends on adequate correlation; applied researchers should be cautious about interpreting components as meaningful constructs when mean |r| falls much below approximately 0.30, even where the predictive gain persists.

5.3 Component Recovery: Tucker's Congruence

Table 6 summarizes Tucker's congruence coefficient across 100 replications per sample size, against the stability threshold $\varphi \geq 0.95$.

Table 6. Tucker's congruence coefficient between recovered and reference (n = 5,000) loading matrices, 100 replications per sample size.

n	Mean φ	Min φ	% φ ≥ 0.95	% φ ≥ 0.85
40	0.942	0.479	76.0	92.0
80	0.981	0.924	95.0	100.0
120	0.987	0.940	99.0	100.0
160	0.992	0.955	100.0	100.0
200	0.993	0.980	100.0	100.0
245	0.994	0.978	100.0	100.0
320	0.996	0.989	100.0	100.0



Tucker congruence stability across sample size

Figure 3. Component recovery stability (Tucker's φ) across sample size, with the φ=0.95 stability threshold.

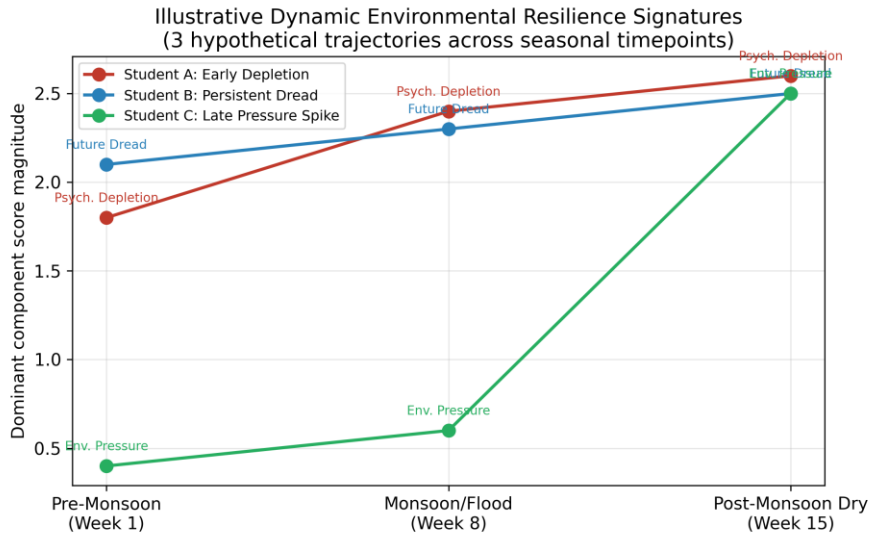
At n = 40, only 76% of replications achieve φ ≥ 0.95, and the minimum φ across 100 replications is 0.479 — an occasionally severely distorted solution, though the mean recovery (φ = 0.942) is close to the threshold. At n = 80, 95% of replications already achieve φ ≥ 0.95, and stability is essentially universal from n = 120 onward. This crossing point is somewhat lower than that reported for a comparable eight-variable burnout structure with lower average cross-loading, plausibly reflecting the slightly more concentrated (higher first-eigenvalue) structure of the present loading matrix.

The practical implication: a researcher collecting environmental-stress survey data with this eight-item, three-component instrument should target n ≥ 120 for a high (≥ 99%) probability of a stable, interpretable solution, and can treat n ≈ 80 as an acceptable minimum when data collection is constrained, with explicit acknowledgment that roughly one in twenty samples at that size will fail to reach the conventional stability threshold. Studies with n well below 80 should not rely on this PCA structure for substantive interpretation, even though the aggregate predictive (AIC/MSE) advantage over raw-variable regression persists at smaller n.

6. The Dynamic Environmental Resilience Signature: Longitudinal Extension

6.1 From Cross-Sectional Components to Seasonal Trajectories

The three components established in Sections 4 and 5 — Environmental Pressure, Psychological Depletion, and Future Dread — define a three-dimensional space within which any resident's environmental-stress state at any timepoint can be located by three component scores (z_1, z_2, z_3). The dominant component is the one with the highest absolute standardized score: D(t) = argmax_j |z_j(t)|. The dynamic environmental resilience signature is the sequence D = (D(1), …, D(T)) across T seasonal measurement occasions.



Illustrative dynamic signature trajectories

Figure 4. Illustrative Dynamic Environmental Resilience Signature trajectories for three hypothetical residents across a seasonal cycle.

Figure 4 illustrates the concept with three hypothetical resident trajectories across a three-timepoint seasonal cycle (pre-monsoon, monsoon/flood period, post-monsoon dry season). A resident showing early Psychological Depletion dominance (resource exhaustion appearing before the flood peak) follows a qualitatively different path from a resident showing persistent Future Dread throughout, or one who shows a late spike in Environmental Pressure only once acute disruption arrives. Standard cross-sectional analysis at a single timepoint cannot distinguish these residents if their total distress scores happen to converge; the signature distinguishes them from the first measurement occasion.

6.2 Proposed Longitudinal Design

The proposed study would measure the same eight-variable battery at T = 3 timepoints within a single annual cycle: pre-monsoon (baseline), monsoon/flood period, and post-monsoon dry season. Component scores at each timepoint would be computed by projecting each resident's standardized item vector onto the reference loading matrix estimated from the baseline sample — score projection that preserves comparability across timepoints without refitting PCA at each occasion.

For each resident, the three-occasion signature D = (D(1), D(2), D(3)) is categorical, taking values in $\{P, L, F\}^3 = 27$ possible signature types, where P = Environmental Pressure, L = Psychological

Depletion, F = Future Dread. Applied researchers may collapse rare types or treat dominant-component scores as continuous trajectories rather than categorical sequences, depending on achieved sample size.

The primary hypothesis is that signature type at the pre-monsoon baseline predicts post-monsoon distress severity beyond what the total baseline stress score alone predicts — that is, the trajectory carries information above and beyond the level. A resident with early-onset Depletion dominance is expected to show higher post-monsoon distress than a resident with matched baseline Pressure dominance, because Depletion reflects resource exhaustion that plausibly compounds across the season, while Pressure may remain intermittent and tied to specific acute events.

6.3 Community and Institutional Application

The practical utility of the signature for community mental-health and environmental-policy programs is direct. A resident whose pre-monsoon signature shows Depletion dominance may benefit from recovery-oriented support — rest, reduced cognitive load, accessible counseling — before the flood period makes both the intervention harder to deliver and harder for the resident to use. A resident with Dread dominance may benefit from future-oriented coping and uncertainty-management support, distinct from acute-exposure relief. A resident with Pressure dominance may need direct exposure mitigation — cooling infrastructure, water/power reliability — rather than psychological intervention per se.

At the community level, the distribution of signature types across a neighborhood or district characterizes its environmental-stress ecology at a given point in the seasonal cycle. If year-on-year tracking shows a rising share of early-Depletion signatures, this may indicate a resource-exhaustion problem — residents are depleting before the flood season even begins. If Future Dread signatures rise disproportionately, this may indicate that anticipatory, future-oriented distress is outpacing present-moment exposure, pointing toward a different kind of intervention (information, agency, and long-horizon planning support rather than immediate relief). The signature offers both an individual screening tool and a population-level indicator for monitoring environmental-policy and community-support interventions over time.

7. Discussion

7.1 What the Simulation Demonstrates

Three findings carry direct practical implications. First, PCA recovers the true three-component environmental-stress structure with high fidelity above $n \approx 120$ ($\varphi \geq 0.95$ in 99% of replications) but degrades meaningfully at $n = 40$ ($\varphi \geq 0.95$ in only 76% of replications, with occasional severe distortion). This crossing point is close to the commonly cited $n \geq 10p$ heuristic ($n = 80$ for eight variables) and somewhat lower than the $n \geq 200$ threshold reported for comparable but less-concentrated burnout structures — a reminder that the practical sample-size threshold depends on the specific loading structure at hand, not on a single universal rule.

Second, PCA's AIC advantage over raw-variable regression is positive across every sample size and correlation level tested ($\Delta AIC = 9.5$ to 16.4), including the weakest correlation condition tested (mean $|r| \approx 0.15$), where raw and PCA-based predictive fit are nearly identical but the parsimony gain from five fewer parameters still yields decisive AIC support for the component-based model. This implies that PCA's model-selection advantage in this class of problem holds whenever the true predictor structure is genuinely lower-dimensional than the observed item set, regardless of how strongly the items happen to correlate in a given sample — though, as Section

5.2 notes, the scientific interpretability of the components themselves still depends on adequate correlation among items.

Third, MSE reduction is largest at small n (10.5% at n = 40) and declines to near zero at large n (0.2% at n = 400), reflecting the well-documented sensitivity of ordinary least squares to multicollinearity at low degrees of freedom. Applied researchers working with small community-level or clinic-level samples — a common constraint in environmental-health field research — stand to gain the most from PCA preprocessing in predictive models, even as the structural (φ-based) reliability of the components themselves is weakest at exactly those sample sizes. This tension — greatest predictive benefit at the sample sizes where structural recovery is least reliable — is itself a practical caution: a small-sample study may show an attractively large MSE improvement while the underlying component structure has not actually been recovered reliably, and researchers should report both metrics rather than either alone.

7.2 Limitations of the Simulation Approach

Several limitations qualify these findings. First, the simulation assumes multivariate normality throughout. Real environmental-stress survey data, collected on Likert scales with ceiling and floor effects, will violate this assumption to some degree; the specific numerical thresholds in Tables 4–6 may not transfer exactly to ordinal, non-normal field data, and future simulation work should vary the distributional assumptions explicitly.

Second, the true factor structure is fixed at three components with a specific cross-loading pattern (Table 1). Real environmental-stress data may have a different true dimensionality — two broader dimensions, or four to five more differentiated ones — and the sensitivity of the recovery thresholds reported here to the number and configuration of true components is not examined.

Third, and most importantly, this is a simulation study: the loading matrix, cross-loadings, and outcome-generating weights are chosen to be theoretically plausible but are not estimated from real environmental-stress data in Pakistan or any other setting. The specific variance-explained percentages, VIF values, and AIC advantages reported here describe the behavior of the analytic method under a plausible synthetic structure, not empirical findings about any real population. The contribution of this paper is the validated analytic framework, the piloted eight-item instrument design, and the sensitivity boundaries for its use — not substantive claims about environmental distress in any specific community. Field deployment with real survey data is the necessary next step before any applied or clinical claims can be drawn.

8. Conclusion

This simulation study demonstrated PCA's application to an eight-variable environmental-stress dataset with known three-factor structure, systematically quantified performance across sample size and correlation conditions, and introduced the Dynamic Environmental Resilience Signature framework for longitudinal extension. Four findings stand out. PCA recovers the true environmental-stress component structure reliably above n ≈ 120 (Tucker φ ≥ 0.95 in 99% of replications) and becomes markedly less reliable at n = 40 (76%). The AIC advantage of PCA over raw-variable regression is positive and consistent across all conditions tested (ΔAIC = 9.5 to 16.4), driven substantially by parsimony rather than multicollinearity correction alone. MSE reduction is largest at small n, exactly where raw-variable regression is least stable and where structural component recovery is simultaneously least reliable — a combination that argues for reporting both predictive and structural-recovery metrics together in applied work. Applied researchers should treat components as scientifically interpretable only when mean inter-item correlation is at

least moderate ($|r| \geq \sim 0.30$), even though the predictive advantage of PCA persists below that threshold.

The Dynamic Environmental Resilience Signature extends these cross-sectional findings into a longitudinal framework appropriate to seasonally structured environmental stressors. Each resident's characteristic sequence of dominant components across a seasonal cycle is individually distinctive and recoverable by projecting new data onto the baseline loading matrix without re-fitting the model. The signature offers a route to identifying at-risk residents before aggregate distress scores reach actionable thresholds, and a population-level indicator for tracking the community-level impact of environmental and wellbeing interventions across seasons and years. The simulation design, sensitivity tables, and longitudinal framework together constitute a methodological guide for researchers deploying PCA on environmental-stress behavioral data. The simulation code is available from the author on request.

References

(Methodological references follow the same conventions used in the companion academic-burnout simulation study; domain references below should be verified against current publication details before submission.)

1. Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). Springer.
2. Cianconi, P., Betrò, S., & Janiri, L. (2020). The impact of climate change on mental health: A systematic descriptive review. *Frontiers in Psychiatry, 11*, 74.
3. Clayton, S., & Karazsia, B. T. (2020). Development and validation of a measure of climate change anxiety. *Journal of Environmental Psychology, 69*, 101434.
4. Hogg, T. L., Stanley, S. K., O'Brien, L. V., Wilson, M. S., & Watsford, C. R. (2021). The Hogg Eco-Anxiety Scale: Development and validation of a multidimensional scale. *Global Environmental Change, 71*, 102391.
5. Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179–185.
6. Lorenzo-Seva, U., & ten Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology, 2*(2), 57–64.
7. MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*(1), 84–99.
8. Ojala, M., Cunsolo, A., Ogunbode, C. A., & Middleton, J. (2021). Anxiety, worry, and grief in a time of environmental and climate crisis: A narrative review. *Annual Review of Environment and Resources, 46*, 35–58.
9. Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.
10. Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99*(3), 432–442.

Data & Reproducibility Note (for internal use — remove or adapt for submission)

- Simulation code: sim_core2.py, sensitivity2.py, relabel_fig2.py, make_figures.py (Jupyter-cell-formatted with # %% markers)
- Figures: fig1_scree_plot.{png,eps}, fig2_loadings_heatmap.{png,eps}, fig3_tucker_stability.{png,eps}, fig4_dynamic_signature_illustration.{png,eps}
- Tables 3–6 raw data: table3_sample_size_sensitivity.csv, table4_correlation_sensitivity.csv, table5_tucker_congruence.csv

- **Before submission:** verify the domain references above against current bibliographic details; consider whether the target journal (Indus Journal of Social Sciences / JGST) permits a purely simulation-based methodological paper or requires an accompanying empirical application; and confirm author-guideline compliance (word count, citation style, figure format requirements).