Five Transition Archetypes of National PM2.5 Exposure: A Global Shape-Based Typology for Environmental Governance and Early Warning

Qifeng Li*, Shuo Wen
Guangdong Police College, Guangzhou, China
**Corresponding Author*

Abstract: PM2.5 is still the leading environmental health risk in the world. Yet national pollution trajectories have taken very different paths, and pooled trend analyses tend to hide this. We still know little about whether countries share a small set of transition pathways, or about which starting conditions go with each pathway. Here we build a data-driven typology of national PM2.5 transitions. We use population-weighted exposure series for 200 economies from 1990 to 2023 (World Development Indicators). We combine PELT changepoint detection under a joint mean–variance cost with k-Shape clustering of within-country standardized trajectories. Our analysis yields five interpretable archetypes. Sustained decline covers 89 economies, with median exposure falling from 25.6 to 14.3 μg/m³. Early-peak decline covers 31 economies, low-level plateau 21, and late-peak high-level 21. Sustained rise covers 38 economies, whose median exposure climbs from 23.7 to 29.5 μg/m³. Four economies in five (81.5%) show at least one decline changepoint, and the median first decline falls in 2000 (IQR 1996–2003). Nearly one in five economies kept worsening through 2023. A further one in ten remained at high exposure after peaking, so roughly three in ten either continue to rise or remain at high levels. Both groups concentrate in South Asia, the Middle East, and parts of Africa. We then ran a random-forest attribution analysis on 1990 baseline conditions (macro-F1 = 0.421 ± 0.076 under repeated stratified cross-validation). Demographic scale, population growth, forest cover, and renewable energy share show the strongest associations with archetype membership. No single income-related variable dominates; income per capita ranks eighth of eleven predictors. We believe the typology offers a reusable benchmark for judging national pollution transitions against peer trajectories. Its dashboards can also support social governance and public-security work as an early-warning screen. They flag economies where rising exposure may translate into public-health burdens, environment-related complaints, and pressure on administrative resources.

Keywords: PM2.5; Transition Trajectory; Time-Series Clustering; K-Shape; Changepoint Detection; SHAP; Environmental Transition; Sustainable Development

1. Introduction

PM2.5 is the largest environmental contributor to the global burden of disease. It is responsible for several million premature deaths each year [1–4]. The global mean hides pronounced cross-national heterogeneity. Exposure has declined steadily in North America and Western Europe since the 1990s. In South Asia and parts of the Middle East, by contrast, it has continued to rise. Nearly the entire global population lives in areas that exceed the most stringent air quality guidelines [5–8]. We need to know whether national trajectories organize into a small number of recurrent transition pathways. That question comes first if we are to interpret progress, benchmark laggards, and target cooperation.

Most work frames the problem as the environmental Kuznets curve (EKC), an inverted-U between income and pollution. Three decades of research show that this curve is, at best, a fragile average. Turning points vary by orders of magnitude across specifications, pollutants, and samples. The curve also shifts with trade openness, technological diffusion, and structural change [9–12]. This heterogeneity persists in recent meta-analyses [13]. Decomposition studies trace emission changes to

Copyright @ STEMM Institute Press

scale, composition, and technique effects, and the strength of these effects differs across development stages [10,14]. A single parametric curve fitted to pooled data therefore averages over qualitatively different experiences. What we lack is an empirical typology of the trajectories themselves. How many distinct pathways exist? Which economies follow each one? Which initial conditions go with each?

Typologies of transition pathways have paid off in socio-technical research, where classification came before mechanistic explanation [15]. Shape-based clustering now lets us make the same move with longitudinal data. It groups series by the shape of their evolution rather than by their level [16]. We apply this approach to population-weighted PM2.5 exposure series for 200 economies over 1990–2023, in four steps. First, we detect changepoints in each national trajectory to characterize the timing of pollution turning points. Second, we cluster within-country standardized trajectories with k-Shape, which is invariant to amplitude and temporal shifts, to derive transition archetypes. Third, we test robustness to alternative distances, smoothing, and cluster counts. Fourth, we relate archetype membership to 1990 baseline conditions with interpretable machine learning in a strictly associational design. The outcome is a five-archetype typology with per-country dashboards. It reframes the EKC question. Instead of asking where the turning point is, we ask which transition each economy is actually on. This reframing also carries direct governance value. An archetype-and-dashboard view can work as an early-warning screen for social governance and public security. It identifies the economies where adverse trajectories are most likely to create health burdens, environmental complaints, and administrative pressure.

2. Materials and Methods

2.1 Data

National PM2.5 exposure comes from the World Bank World Development Indicators (WDI), indicator EN.ATM.PM25.MC.M3. This indicator gives the population-weighted annual mean concentration of fine particulate matter ($\mu g/m^3$) [17], and it draws on satellite retrievals, chemical transport modelling, and ground measurements [7]. We used all 34 annual observations from 1990 to 2023. Over this window, 200 WDI economies had a complete series; these form our analysis sample. We excluded regional and income aggregates.

Eleven WDI indicators measured baseline structural conditions. They span economic scale and growth, economic structure, energy and land systems, and demography. The set contains GDP per capita (constant prices), GDP growth, industry and agriculture value added shares, trade openness, renewable energy share in final consumption, energy use per capita, urban population share, total population, population growth, and forest area share. We excluded carbon-dioxide indicators because CO2 and PM2.5 are partly co-emitted; including them would introduce circularity into the attribution. We also excluded primary energy intensity (EG.EGY.PRIM.PP.KD) because that series begins only in 2000. For each indicator, we took the first non-missing value within 1990 ± 2 years, or else the earliest value between 1993 and 2000, and we recorded the exact year used (see Data Availability). Missingness ranged from 0.0% (urbanization, population size and growth) to 15.0% (energy use per capita). We median-imputed the remaining gaps and report the imputation rate for each variable.

2.2 Trajectory Preprocessing

We z-standardized each country's 34-year series within country, which separates shape from level; we report levels per archetype. As a smoothing sensitivity check, we repeated the clustering on five-year moving averages (centred, minimum three periods), again standardized within country.

2.3 Changepoint Detection

We detected turning points in each national series with the pruned exact linear time (PELT) algorithm [18]. The cost was a joint mean–variance Gaussian, the minimum segment length was three years, and the penalty was a conservative modified-BIC of $4 \cdot \ln(n)$ ($n = 34$). Binary segmentation with identical cost and penalty served as a cross-check [19]. We also assessed penalty sensitivity at $2 \cdot \ln(n)$ and $6 \cdot \ln(n)$. For each country, we recorded the number and years of changepoints, the slope of each segment (ordinary least squares on year), the peak year and value, and the first changepoint after which the following segment slope turns negative (the "first decline point"). We ran the analyses with the ruptures package (v1.1.10) for Python [19].

 Copyright @ STEMM Institute Press

2.4 Shape-Based Clustering of Trajectories

We clustered the standardized trajectories with k-Shape [20]. This is a centroid-based algorithm, and its shape-based distance (SBD) comes from normalized cross-correlation. SBD is invariant to scaling and temporal shifts: countries with the same rise-then-decline shape cluster together even when their peaks differ in timing (Section 3.2 quantifies peak-year dispersion). We ran k-Shape with ten random initializations and up to 200 iterations (tslearn v0.9.0 [21]). The reference SBD implementation is asymmetric, so we computed the symmetric SBD matrix directly from Paparrizos and Gravano's definition [20] for the validation indices.

We selected the cluster count k on four criteria for k = 3–6. The first was average silhouette width on the SBD distance [22]. The second was bootstrap cluster-wise Jaccard stability over 50 resamples with Hungarian matching [23]. The third was agreement (adjusted Rand index, ARI [24]) with two alternatives: hierarchical average-linkage clustering on dynamic time warping (DTW) distances of the same series, and soft-DTW k-means [25] on log-transformed raw levels, which retains level differences. The fourth was interpretability and minimum cluster size. Agreement with the five-year smoothed clustering provided a further sensitivity check.

2.5 Baseline Conditions Associated with Archetype Membership

We modelled the k = 5 archetype label as a function of the eleven baseline indicators. Our question was which 1990 conditions are associated with pathway membership. The design is associational and out-of-sample. We measured predictors at the start of the trajectory window (1988–1992 for 94.6% of non-missing country–indicator cells, the remainder within 1993–2000; Section 2.1). We evaluated a benchmark of seven classifiers spanning linear, margin-, and tree-based learners, plus a majority-class dummy, under identical conditions (Section 3.4). We tuned hyperparameters of all seven classifiers by Bayesian optimization (Optuna TPE sampler, 50 trials per model) maximizing stratified five-fold macro-F1. For attribution, we used the tuned random forest (200 trees, feature fraction 0.5 per split, minimum leaf size 4, maximum depth 24, class-balanced subsampling) [26], with a tuned multinomial logistic regression (L2 penalty, standardized features) as the linear benchmark. We estimated performance with repeated stratified five-fold cross-validation (10 repeats, 50 fits), reporting macro-F1 and balanced accuracy, and we summarized per-archetype patterns in a pooled out-of-fold confusion matrix.

For interpretation, we computed SHAP (SHapley Additive exPlanations) values on the full-sample fit [27,28] with the Python shap package (v0.49.1). We summarized global contributions by the mean absolute SHAP value averaged across classes. Class-specific beeswarm plots showed directions of association. We implemented the classifiers with scikit-learn [26].

2.6 Transition Milestones

We recorded the first year in which each country's exposure falls below each WHO interim target for annual PM2.5 (IT-1, 35 μg/m³; IT-2, 25 μg/m³; IT-3, 15 μg/m³; IT-4, 10 μg/m³), using the target values tabulated in [5]. We then report crossing shares per archetype.

3. Results

3.1 National Pollution Trajectories Are Punctuated by Decline Onsets around the Turn of the Century

Over 1990–2023, the median economy shows two changepoints (distribution: 0 in 2 economies, 1 in 44, 2 in 71, 3 in 42, 4 in 29, 5 in 12). PELT and binary segmentation agree on the count for 79.0% of economies, so the detected structure does not depend on the search algorithm. Penalty sensitivity, however, is substantial. A liberal penalty ($2 \cdot \ln(n)$) yields a median of five changepoints; a strict one ($6 \cdot \ln(n)$) yields a median of one. We therefore treat exact counts cautiously and put weight on the first decline point. Its median timing is stable across penalties (median 1996–2001 across the three; median pairwise country-level difference 0 years). Country-level timing does vary with the penalty: 39.5–56.1% of countries have identical years between specifications, and 64–66% agree within three years.

Four economies in five (163 of 200, 81.5%) showed at least one changepoint followed by a declining segment. First decline points cluster around the turn of the century: median 2000, interquartile range 1996–2003, full range 1995–2018 (Figure 1). The median peak year across all

Copyright @ STEMM Institute Press

economies is 2003, and the median change between 1990 and 2023 is −15.0%. In addition, 121 of 200 economies (60.5%) show a negative linear slope over 2014–2023.

3.2 Five Transition Archetypes Organize the Diversity of National Trajectories

With k = 5, shape-based clustering produced five coherent and interpretable transition archetypes (Figures 2–4; Table 1). We label them by qualitative shape.

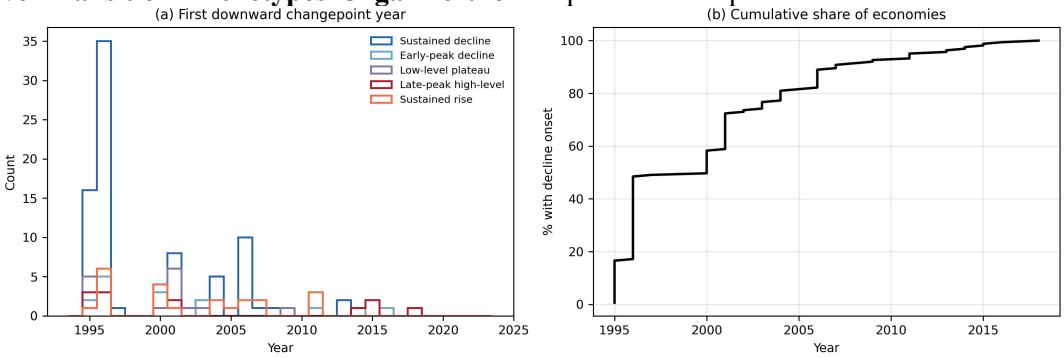

Figure 1. First Decline Changepoint Years, Histogram and Cumulative Share; (the 1995–1996 Concentration Reflects Left-Censoring of Early Transitions)

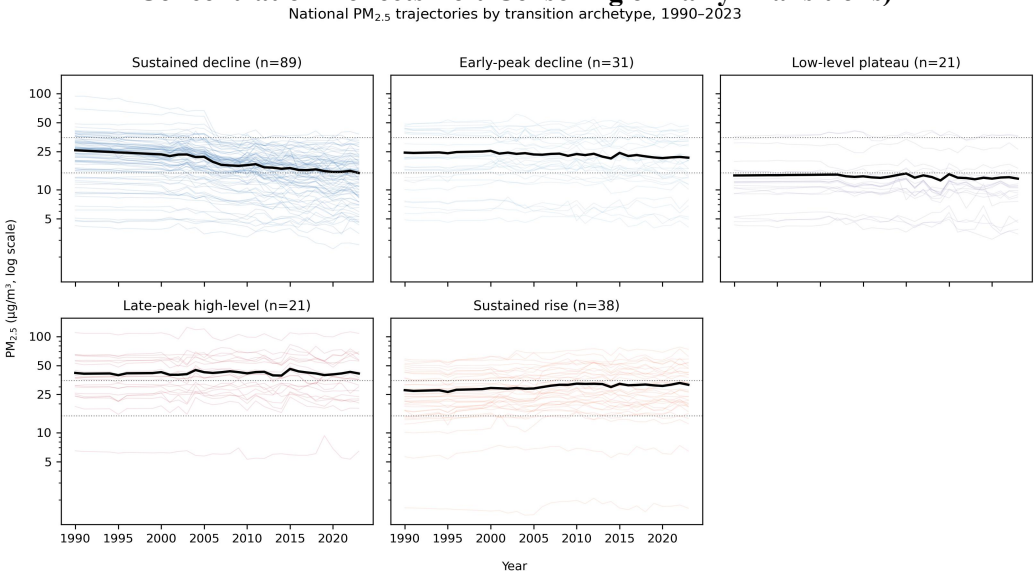

Figure 2. National PM2.5 Trajectories 1990–2023 by Archetype (Log Scale); (Thin Lines are Member Economies, the Bold Line is the Archetype Median)

- Sustained decline (n = 89, 44.5%): exposure falls throughout the window, from a median of 25.6 µg/m³ in 1990 to 14.3 µg/m³ in 2023 (median within-country change −39.2%). This archetype dominates Europe, North America, and parts of Latin America and East Asia. All members have crossed IT-1, 89.9% have crossed IT-2, and 30.3% have crossed IT-4.
- Early-peak decline (n = 31, 15.5%): exposure rises to a peak around 2002, then declines slowly. Median exposure changes only from 21.0 to 20.0 µg/m³ (median within-country change −8.9%). The group is heterogeneous in level and includes China, whose peak occurs later (2011) than the archetype median. We read this as a consequence of the shift invariance of the shape distance (Section 4).
- Low-level plateau (n = 21, 10.5%): trajectories stay flat at low concentrations (11.9 to 10.6 µg/m³). The group consists largely of small island states and low-density economies. It has the highest IT-4 crossing share of any archetype, at 66.7%.
- Late-peak high-level (n = 21, 10.5%): exposure rises until around 2014 and then declines slowly from the highest levels of any archetype (37.0 to 35.7 µg/m³; median within-country change −2.8%). Only 4.8% of members have crossed IT-3. The group concentrates in South Asia and Sub-Saharan Africa and includes Bangladesh and Nigeria.
- Sustained rise (n = 38, 19.0%): exposure increases over the full window, from 23.7 to 29.5 µg/m³ (median within-country change +10.4%), with a positive median slope over 2014–2023. Members include India, Ethiopia,

 Copyright @ STEMM Institute Press

Saudi Arabia, and Iran. Only 7.9% have crossed IT-4.

The archetypes map onto coherent macro-regions (Figure 3). Sustained decline accounts for 82.7% of Europe and Central Asia and all of North America. It accounts for only 4.2% of Sub-Saharan Africa, where early-peak decline (31.2%), late-peak high-level (27.1%), and sustained rise (25.0%) dominate. In the Middle East, North Africa, Afghanistan and Pakistan region, 60.9% of economies follow the sustained-rise path. The income gradient is just as sharp. Among high-income economies, 68.6% are sustained decliners. Among low-income economies, 72.0% sit in the late-peak high-level (40.0%) or sustained-rise (32.0%) archetypes. Each archetype comprises 21–89 economies, so none is residual or a singleton.

Two properties qualify these labels. SBD is invariant to temporal shifts, so a label describes shape rather than calendar timing. Within early-peak decline, for example, national peak years span more than a decade: China peaks in 2011 yet clusters with economies that peak around 2002. The early-window onset of sustained decline is also a left-censoring artefact (Section 4).

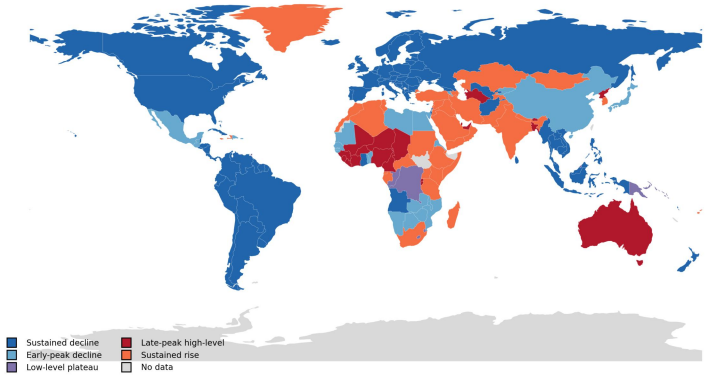

Figure 3. World Map of Transition-Archetype Membership (166 of 200 Economies Colored; Small Island States without Natural Earth Matches Shown in Grey)

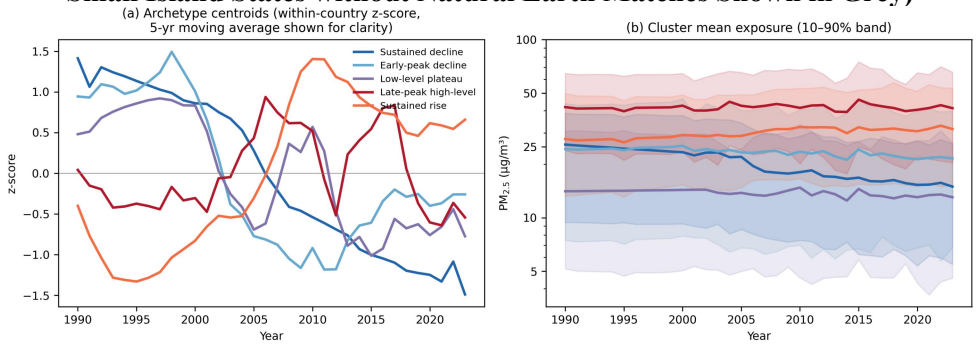

Figure 4. Archetype Centroid Trajectories (within-Country z-Standardized, 5-Year Moving Average) with 10–90 Percentile Bands in the Original Level Space

Table 1. Transition Archetypes: Size, Exposure Levels and Change, Timing, Recent Slope, and WHO Interim-Target Crossing Shares

Archetype	n (%)	1990	2023	Δ (%)	Peak year	First declining segment onset	Slope 2014–23	IT-1	IT-2	IT-3	IT-4
Sustained decline	89 (44.5)	25.6	14.3	−39.2	1990	1996	−0.127	100.0	89.9	59.6	30.3
Early-peak decline	31 (15.5)	21.0	20.0	−8.9	2002	2001	−0.044	90.3	64.5	41.9	19.4
Low-level plateau	21 (10.5)	11.9	10.6	−11.2	2010	2000–2001	−0.002	100.0	90.5	81.0	66.7
Late-peak high-level	21 (10.5)	37.0	35.7	−2.8	2014	2000	−0.123	61.9	28.6	4.8	4.8
Sustained rise	38 (19.0)	23.7	29.5	+10.4	2013	2000–2001	+0.102	78.9	55.3	23.7	7.9

Note: Levels are archetype medians of population-weighted PM2.5 (µg/m³); Δ is the median within-country percentage change between 1990 and 2023; slope is the median 2014–2023 linear trend (µg/m³/yr); IT columns give the share of members that have ever crossed each interim target. First declining segment onset is the median year of the first changepoint followed by a negative-slope segment of any length; in rising archetypes it typically marks a transient early decline rather than a sustained transition (22 of 38 sustained-rise members show such a segment while 33 of 38 (87%) rose over the full window), and for the flat low-level plateau the peak year has limited meaning.

Copyright @ STEMM Institute Press

3.3 The Typology Is Robust to Distance, Smoothing, and Level Information

We selected the five-archetype solution against k = 3–6 on pre-specified criteria (Table 2). Silhouette width on the SBD distance declines gradually: 0.244, 0.234, 0.231, and 0.221 for k = 3–6. Bootstrap Jaccard stability averages 0.666, 0.596, 0.548, and 0.449. We chose k = 5 as the largest k that retains acceptable stability with a minimum size above 20. It adds an interpretable distinction, late-peak high-level versus sustained rise, that disappears at smaller k. Stability at k = 5 is moderate (mean Jaccard 0.548, minimum 0.354), yet borderline assignments are rare: only 1 of 200 economies (0.5%) sits within a normalized SBD-distance margin of 0.02 between the assigned and nearest rival centroid. Stability is also uneven. The two large archetypes are most stable (bootstrap Jaccard 0.832 for sustained decline, 0.747 for sustained rise). The three smaller ones are less stable (0.354–0.417).

Three cross-checks support reading the typology as shape information. Hierarchical clustering on DTW distances of the same series recovers a similar partition (ARI 0.58–0.61 across k), so the archetypes do not depend on k-Shape specifically. Clustering log-transformed raw levels with soft-DTW yields essentially orthogonal partitions (ARI 0.08 at k = 5). Membership is almost entirely independent of exposure level, and we see this as the added value of the design. Clustering five-year moving averages recovers broadly similar assignments (ARI 0.61 at k = 5), so the archetypes reflect medium-term shapes rather than interannual fluctuations.

Table 2. Cluster Quality and Robustness for k = 3–6

Method	k	Silhouette	Inertia	Jaccard mean	Jaccard min	Min size	ARI vs k-Shape
k-Shape (z)	3	0.244	0.09	0.666	0.388	28	—
k-Shape (z)	4	0.234	0.08	0.596	0.368	20	—
k-Shape (z)	5	0.231	0.07	0.548	0.354	21	—
k-Shape (z)	6	0.221	0.06	0.449	0.243	17	—
DTW hierarchical (z)	3	0.326	—	—	—	—	0.592
DTW hierarchical (z)	4	0.321	—	—	—	—	0.598
DTW hierarchical (z)	5	0.309	—	—	—	—	0.605
DTW hierarchical (z)	6	0.307	—	—	—	—	0.583
soft-DTW (log levels)	3	—	—	—	—	—	0.082
soft-DTW (log levels)	4	—	—	—	—	—	0.075
soft-DTW (log levels)	5	—	—	—	—	—	0.077
soft-DTW (log levels)	6	—	—	—	—	—	0.095
k-Shape (5-yr MA)	3	—	—	—	—	—	0.709
k-Shape (5-yr MA)	4	—	—	—	—	—	0.599
k-Shape (5-yr MA)	5	—	—	—	—	—	0.612
k-Shape (5-yr MA)	6	—	—	—	—	—	0.671

Note: Silhouette and inertia are reported for k-Shape on within-country z-standardized trajectories (SBD distance) and for the DTW-hierarchical alternative; Jaccard stability from 50 bootstrap resamples; ARI measures agreement of each specification with the k-Shape solution at the same k.

3.4 Demographic and Land–Energy Baselines, More than Income, Are Associated with Archetype Membership

From 1990 baseline conditions, a tuned random-forest classifier predicted archetype membership with a cross-validated macro-F1 of 0.421 ± 0.076 and balanced accuracy of 0.441 ± 0.082. That is far above a most-frequent-class dummy baseline (macro-F1 0.123). It is also above the tuned linear benchmark (multinomial logit macro-F1 0.362 ± 0.078; gain 0.059; Wilcoxon signed-rank $p = 5.1 \times 10^{-6}$ across the 50 CV folds). We compared seven classifiers, tuned by Bayesian search (Section 2.5), under the same protocol. Tree-bagging methods performed best (random forest 0.421, extra trees 0.416), followed by RBF-SVM (0.398) and XGBoost (0.397). LightGBM (0.378), histogram gradient boosting (0.366), and the linear model (0.362) performed worse, and no classifier exceeded 0.43 (Table 3). Tuning improved the random forest by only 0.005 over its default configuration, and tuning-phase cross-validation estimates (up to 0.487) proved optimistic relative to this independent evaluation. We take both as signs that the performance ceiling reflects the limited signal in 1990 baseline conditions rather than learner choice. In the pooled confusion matrix, early-peak decline was most often

 Copyright @ STEMM Institute Press

confused with sustained decline and sustained rise, which fits its intermediate shape.

Table 3. Tuned Attribution Classifier Benchmark under Repeated Stratified Five-Fold cross-Validation (10 repeats, 50 fits; mean ± SD)

Model	Accuracy	Balanced acc.	Macro-F1	Weighted-F1	Cohen's κ	MCC
Random forest	0.504 ± 0.068	0.441 ± 0.082	0.421 ± 0.076	0.505 ± 0.065	0.324 ± 0.091	0.329 ± 0.093
Extra trees	0.506 ± 0.084	0.439 ± 0.099	0.416 ± 0.094	0.500 ± 0.082	0.321 ± 0.112	0.326 ± 0.114
SVM (RBF kernel)	0.470 ± 0.070	0.432 ± 0.092	0.398 ± 0.078	0.483 ± 0.066	0.302 ± 0.091	0.311 ± 0.093
XGBoost	0.525 ± 0.058	0.396 ± 0.076	0.397 ± 0.084	0.494 ± 0.059	0.301 ± 0.083	0.312 ± 0.086
LightGBM	0.454 ± 0.068	0.390 ± 0.084	0.378 ± 0.077	0.456 ± 0.066	0.254 ± 0.093	0.257 ± 0.095
Histogram gradient boosting	0.488 ± 0.059	0.373 ± 0.071	0.366 ± 0.078	0.461 ± 0.058	0.260 ± 0.081	0.266 ± 0.082
Multinomial logistic regression	0.498 ± 0.061	0.374 ± 0.077	0.362 ± 0.078	0.467 ± 0.060	0.273 ± 0.084	0.282 ± 0.088
Dummy (most-frequent class)	0.445 ± 0.000	0.200 ± 0.000	0.123 ± 0.002	0.274 ± 0.000	0.000 ± 0.000	0.000 ± 0.000

Note: Hyperparameters tuned by Bayesian optimization (50 trials per model).

SHAP attribution (Figure 5) ranks the baseline predictors consistently across classes. The four largest contributions come from total population (mean |SHAP| = 0.058), population growth (0.034), forest area share (0.032), and renewable energy share (0.025). Energy use per capita (0.020), urbanization (0.019), and trade openness (0.018) follow, and only then GDP per capita (0.017, eighth of eleven). Two checks qualify this ranking. Refitting the cross-validation with imputation inside training folds leaves performance unchanged (macro-F1 0.422 ± 0.072), so imputation leakage is negligible. Income per capita, urbanization, and energy use per capita are strongly correlated (r up to 0.81), and their contributions sum to a development bloc of 0.055. That is comparable to total population (0.058). We therefore read the headline ranking as showing that no single income-related variable dominates pathway membership, not that development level is irrelevant. Class-specific directions are interpretable. Rapid population growth at baseline is associated with a lower likelihood of sustained-decline membership (Spearman ρ between feature value and class SHAP value = −0.94). Higher GDP per capita is associated with a higher likelihood of that archetype (ρ = +0.90). A large baseline population is associated with sustained-rise membership (ρ = +0.86), and greater forest cover with a lower likelihood of that archetype (ρ = −0.79). The renewable energy share shows class-dependent signs: it associates positively with late-peak high-level membership (ρ = +0.27) and negatively with sustained-decline membership (ρ = −0.37). The baseline heatmap (Figure 6) shows this structure descriptively. Sustained-decline economies began the window with the highest incomes and urbanization and the lowest renewable shares (median 13.4%, consistent with modern-fuel transitions already underway by 1990). Sustained-rise economies began with the largest populations and the lowest forest cover. Late-peak high-level economies began with the largest agricultural sectors, the lowest urbanization, the fastest population growth, and the highest renewable shares (median 84.0%, dominated by traditional biomass).

4. Discussion

We reframe the question. Instead of asking where the turning point of the pollution curve lies, we ask which transition pathway each economy is on. Across 200 economies over 1990–2023, our answer is five recurrent archetypes. Most economies (81.5%) show a detectable decline onset, with median timing around 2000. Roughly three in ten, concentrated in South Asia, the Middle East, and parts of Africa, remained on rising (19.0%) or late-peak high-level (10.5%) paths through 2023. In our analysis, 1990 demographic and land–energy conditions are more strongly associated with pathway membership than income per capita.

Our typology complements the EKC literature

rather than contradicting it. Pooled parametric estimates recover a fragile inverted-U [9,11,12]. Our results suggest that this average hides at least five distinct paths. The majority rise-then-decline pattern coexists with a large minority for which no decline has begun. This heterogeneity fits several strands of work: decomposition accounts [10,14], critiques of the EKC as a mixture of dissimilar national stories [11,12], and pathway typologies in socio-technical transitions research [15].

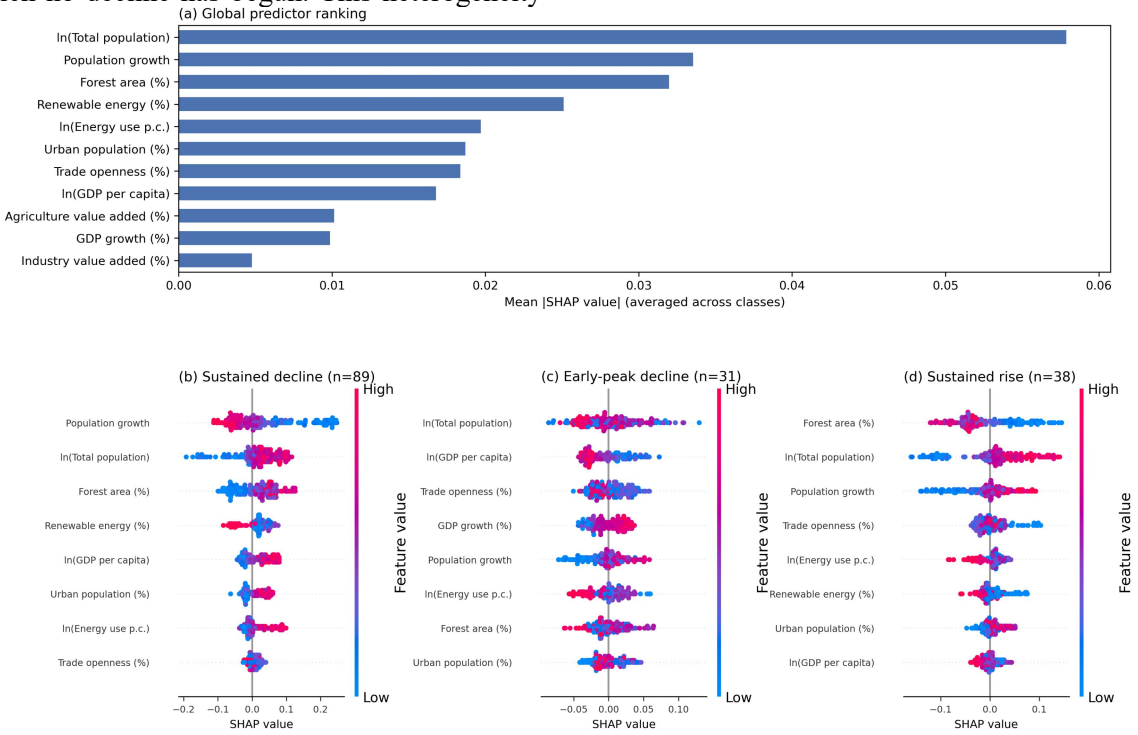

Figure 5. SHAP Attribution of Archetype Membership: Global Mean |SHAP| Ranking and Class-Specific Beeswarm Plots for Sustained Decline, Early-Peak Decline, and Sustained Rise

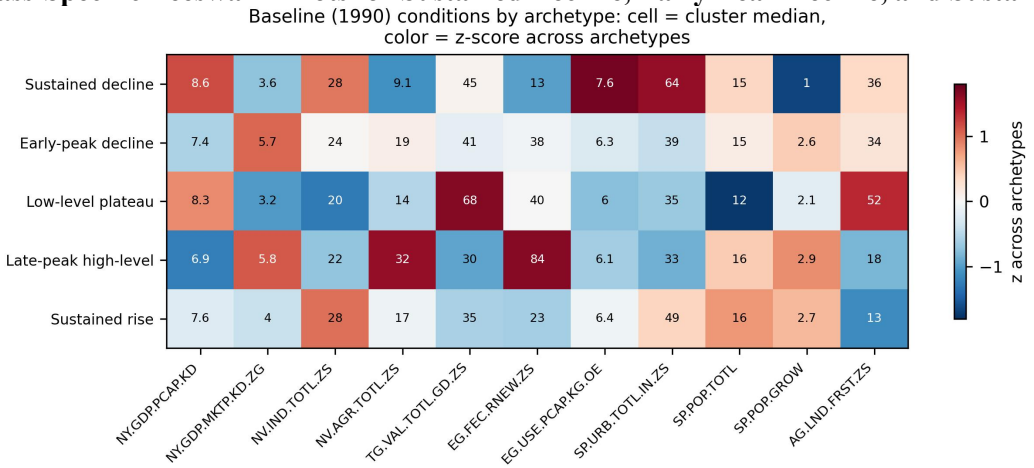

Figure 6. Median 1990 Baseline Conditions by Archetype, z-Scored across Archetypes

The archetypes also sharpen satellite-based assessments of decline in high-income regions and rise in South Asia and the Middle East [6,7]. They do so by assigning every economy an explicit pathway with milestones. Case evidence aligns. China's early-peak decline shape, a late peak around 2011 followed by steep reductions, matches analyses attributing post-2013 improvements to clean-air policies [29]. India's sustained-rise placement matches the continued deterioration in national assessments [30].

Several boundaries qualify these claims. First, the window begins in 1990, so trajectories already in decline are left-censored. The sustained-decline archetype therefore describes within-window behaviour, not the true onset, which for many OECD economies predates the satellite era. Second, the shape distance is deliberately shift-invariant. Labels therefore carry no calendar information, and economies a

 Copyright @ STEMM Institute Press

decade apart can share a label, as China and earlier-peaking economies do. Timing analyses should read archetype and peak year jointly. Third, clustering quality is moderate (silhouette 0.231, mean Jaccard 0.548 at k = 5). We therefore read the archetypes as centroids of a continuous spectrum. Fourth, the attribution analysis is associational by design and reports predictive associations only. The modest macro-F1 (0.421, the best of a tuned seven-classifier benchmark) is consistent with a substantial role for within-window policy and shocks that 1990 baseline conditions cannot capture. Fifth, exposure is population-weighted and model-based, not emissions. The typology describes where people live with pollution, not where pollution is produced, and trade-embodied transfers lie outside the design. Finally, small island states absent from the 110 m boundaries (34 of 200 economies) are unshaded in Figure 3 but enter all analyses. These boundaries point to next steps. We plan to model transition age rather than calendar year, which would separate "when" from "how". We also plan to link membership to within-window policy events and to couple the typology with health-burden data, so that pathways translate into avoided or accumulated mortality.

5. Conclusions

Using shape-based clustering and changepoint detection on 34 years of population-weighted PM2.5 exposure for 200 economies, we show that national pollution transitions organize into five recurrent archetypes: sustained decline (44.5% of economies), early-peak decline, low-level plateau, late-peak high-level, and sustained rise (19.0%). For the four in five economies that show decline onsets, these onsets cluster around the turn of the century. Roughly three in ten, concentrated in South Asia, the Middle East, and parts of Africa, remain on rising or late-peak high-level paths through 2023. Pathway membership is associated most strongly with demographic scale and growth, forest cover, and renewable energy share. No single income-related variable dominates. We believe income-centric narratives of environmental transition are therefore incomplete. Read jointly with peak-year timing (the labels are shift-invariant), the typology and its country dashboards offer a reusable benchmark for assessing progress against peer trajectories. The same dashboards can support social governance and public-security work as an early-warning screen. Economies on sustained-rise or late-peak high-level paths are where rising exposure is most likely to translate into public-health burdens, environment-related complaints, and pressure on administrative and public-order resources. These economies warrant priority attention.

Acknowledgment

This paper was funded by the "Golden Eagle Scholar" Team Project of Guangdong Police College (2026JY01), the "Young Eagle Program" Project of Guangdong Police College (2026QY12), the "Young Eagle Program" Project of Guangdong Police College (2025FY22), and 2024 Guangdong Province Undergraduate Teaching Quality and Teaching Reform Project (No. 879).

References

[1] Lelieveld, J.; Evans, J.S.; Fnais, M.; Giannadaki, D.; Pozzer, A. The contribution of outdoor air pollution sources to premature mortality on a global scale. Nature 2015, 525, 367–371.

[2] GBD 2019 Risk Factors Collaborators. Global burden of 87 risk factors in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020, 396, 1223–1249.

[3] GBD 2021 Risk Factors Collaborators. Global burden and strength of evidence for 88 risk factors in 204 countries and 811 subnational locations, 1990–2023: A systematic analysis for the Global Burden of Disease Study 2021. Lancet 2024, 403, 2162–2203.

[4] Lelieveld, J.; Haines, A.; Burnett, R.; Tonne, C.; Klingmueller, K.; Munzel, T.; et al. Air pollution deaths attributable to fossil fuels: Observational and modelling study. BMJ 2023, 383, e077784.

[5] Health Effects Institute. State of Global Air 2024; Health Effects Institute: Boston, MA, USA, 2024.

[6] Hammer, M.S.; van Donkelaar, A.; Li, C.; Lyapustin, A.; Sayer, A.M.; Hsu, N.C.; et al. Global estimates and long-term trends of fine particulate matter concentrations (1998–2018). Environ. Sci. Technol. 2020, 54, 7879–7890.

[7] van Donkelaar, A.; Martin, R.V.; Brauer, M.; Hsu, N.C.; Kahn, R.A.; Levy, R.C.; et

Copyright @ STEMM Institute Press

al. Global estimates of fine particulate matter using a combined geophysical-statistical method with information from satellites, models, and monitors. Environ. Sci. Technol. 2016, 50, 3762–3772.
[8] Fuller, R.; Landrigan, P.J.; Balakrishnan, K.; Bathan, G.; Bose-O'Reilly, S.; Brauer, M.; et al. Pollution and health: A progress update. Lancet Planet. Health 2022, 6, e535–e547.
[9] Dasgupta, S.; Laplante, B.; Wang, H.; Wheeler, D. Confronting the environmental Kuznets curve. J. Econ. Perspect. 2002, 16, 147–168.
[10] Copeland, B.R.; Taylor, M.S. Trade, growth, and the environment. J. Econ. Lit. 2004, 42, 7–71.
[11] Cherniwchan, J. Economic growth, industrialization, and the environment. Resour. Energy Econ. 2012, 34, 442–467.
[12] Millimet, D.L.; List, J.A.; Stengos, T. The environmental Kuznets curve: Real progress or misspecified models? Rev. Econ. Stat. 2003, 85, 1038–1047.
[13] Sarkodie, S.A.; Strezov, V. A review on environmental Kuznets curve hypothesis using bibliometric and meta-analysis. Sci. Total Environ. 2019, 649, 128–145.
[14] Levinson, A. Technology, international trade, and pollution from US manufacturing. Am. Econ. Rev. 2009, 99, 2177–2192.
[15] Geels, F.W.; Schot, J. Typology of sociotechnical transition pathways. Res. Policy 2007, 36, 399–417.
[16] Aghabozorgi, S.; Shirkhorshidi, A.S.; Wah, T.Y. Time-series clustering—A decade review. Inf. Syst. 2015, 53, 16–38.
[17] World Bank. World Development Indicators; World Bank: Washington, DC, USA, 2025.
[18] Killick, R.; Fearnhead, P.; Eckley, I.A. Optimal detection of changepoints with a linear computational cost. J. Am. Stat. Assoc. 2012, 107, 1590–1598.
[19] Truong, C.; Oudre, L.; Vayatis, N. Selective review of offline change point detection methods. Signal Process. 2020, 167, 107299.
[20] Paparrizos, J.; Gravano, L. k-Shape: Efficient and accurate clustering of time series. In Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, Melbourne, Australia, 31 May–4 June 2015; pp. 1855–1870.
[21] Tavenard, R.; Faouzi, J.; Vandewiele, G.; Divo, F.; Androz, G.; Holtz, C.; et al. Tslearn, a machine learning toolkit for time series data. J. Mach. Learn. Res. 2020, 21, 1–6.
[22] Rousseeuw, P.J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 1987, 20, 53–65.
[23] Hennig, C. Cluster-wise assessment of cluster stability. Comput. Stat. Data Anal. 2007, 52, 258–271.
[24] Hubert, L.; Arabie, P. Comparing partitions. J. Classif. 1985, 2, 193–218.
[25] Cuturi, M.; Blondel, M. Soft-DTW: A differentiable loss function for time-series. In Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; pp. 894–903.
[26] Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; et al. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 2011, 12, 2825–2830.
[27] Shapley, L.S. A value for n-person games. In Contributions to the Theory of Games, Volume II; Kuhn, H.W., Tucker, A.W., Eds.; Princeton University Press: Princeton, NJ, USA, 1953; pp. 307–317.
[28] Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems 30; Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates: Red Hook, NY, USA, 2017; pp. 4765–4774.
[29] Zhang, Q.; Zheng, Y.; Tong, D.; Shao, M.; Wang, S.; Zhang, Y.; et al. Drivers of improved PM2.5 air quality in China from 2013 to 2017. Proc. Natl. Acad. Sci. USA 2019, 116, 24463–24469.
[30] Pandey, A.; Brauer, M.; Cropper, M.L.; Balakrishnan, K.; Mathur, P.; Dey, S.; et al. Health and economic impact of air pollution in the states of India: The Global Burden of Disease Study 2019. Lancet Planet. Health 2021, 5, e25–e38.

 Copyright @ STEMM Institute Press