

Spatio-temporal modelling of provincial Human Development Index in Indonesia: a comparison of six forecasting approaches

I Gusti Ngurah Sentana Putra

IPB University

Muh Akbar Idris

muhakbaridris@apps.ipb.ac.id

IPB University

Research Article

Keywords: Human Development Index, regional planning, spatial temporal forecasting, administrative boundaries, regional typology

Posted Date: September 8th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10939108/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Provincial Human Development Index data are often used to compare regional progress, yet their value for short-term planning depends on whether observed trajectories remain comparable across space and time. This study evaluates six spatial temporal models for one-year-ahead prediction of provincial HDI in Indonesia using annual data for 2017–2023. The models are geographically and temporally weighted regression, a Bayesian spatial temporal model, ConvLSTM, STGCN, GraphWaveNet, and a spatial temporal socioeconomic K-means baseline. A persistence forecast is retained as a non-spatial benchmark. The evaluation distinguishes the full 34-province panel from a stable-boundary sensitivity sample that excludes Papua and West Papua, where post-2021 territorial changes disrupt longitudinal comparability. On the stable sample, ST-SCKM and GTWR record RMSE values of 0.328 and 0.379, compared with 0.630 for persistence. The three deep-learning models do not improve on persistence, which is consistent with the limited number of annual training windows. Residual maps show that model errors are not significantly spatially autocorrelated in 2023, but their magnitude remains higher in Maluku and Papua when the full sample is retained. The findings support a boundary-aware and benchmark-based approach to regional forecasting. They do not establish a universally superior model, but show that simple local and cluster-adjusted forecasts can be more dependable than complex architectures when provincial panels are short and administrative units change.

Introduction

The Human Development Index (HDI) has become a familiar summary of progress in health, education and living standards. At the national level, it helps place economic growth within a wider account of human welfare. At the regional level, however, the index has an additional function. It reveals where development outcomes are concentrated, which areas remain behind, and whether gaps are narrowing at similar rates. These questions are directly relevant to regional planning because the same national improvement can arise from very different provincial trajectories.

Indonesia provides a demanding setting for this type of analysis. The country combines a large archipelagic territory with marked differences in settlement density, accessibility, economic structure and public-service provision. Provinces on Java and several metropolitan or resource-based regions generally record higher HDI, while parts of eastern Indonesia have historically faced lower levels and greater service-delivery constraints. The spatial pattern is not a simple west to east gradient, but neighbouring provinces often share infrastructure systems, labour markets and institutional conditions. A model that treats every province as an independent time series may therefore overlook information contained in regional proximity and similarity.

The last possibility is especially important in Indonesia. The creation of new provinces in Papua after 2021 changed the territory represented by the residual provinces of Papua and West Papua. A value labelled Papua in a recent table may refer to a smaller territory than the province observed earlier in the series. The distinction matters because a forecasting model assumes that the response at a location in

one year is comparable to the response at the same location in the next. When a boundary changes, an apparently large improvement can partly reflect a change in composition rather than development within a constant area.

This issue is easy to miss in model comparisons. Studies often devote considerable attention to algorithms while treating the spatial units as fixed containers. Yet regional evidence is produced through administrative definitions, data collection and boundary harmonisation. A sophisticated model cannot recover comparability that the input data do not provide. For planning research, data geography is therefore part of the model specification rather than a preliminary cartographic detail.

The second difficulty concerns the relationship between model complexity and data volume. Deep-learning architectures such as ConvLSTM, spatio-temporal graph convolutional networks and GraphWaveNet were developed for dense sequences, including rainfall or traffic observations collected at short intervals. Indonesian provincial HDI is annual. A seven-year panel offers only a few temporal windows for training, even though each window contains observations for 34 provinces. The network can learn cross-sectional structure, but the evidence available for temporal generalisation remains limited.

These conditions make a simple benchmark indispensable. HDI usually changes gradually, so a persistence forecast that carries the previous year's value forward can be difficult to beat. If a complex spatial model performs worse than persistence, its architecture has not provided practical predictive value for this sample. The comparison should also include methods that occupy a middle ground between persistence and deep learning. Geographically and temporally weighted regression (GTWR) allows local relationships to vary across space and time, while a cluster-adjusted forecast can preserve an interpretable connection between regional typology and expected change.

Research published in the *Journal of Regional and City Planning* offers a useful basis for that planning orientation. Baransano et al. (2016) showed that regional inequality in West Papua cannot be separated from the economic structure and position of particular districts. Latuconsina et al. (2018) used a development-region typology to connect multiple indicators with differentiated planning responses in Malang Regency. Otok et al. (2022) likewise demonstrated that clustering and mapping can make infrastructure disparities more legible for prioritisation. These studies differ in scale and method, but they share an interest in translating quantitative differentiation into region-specific interpretation.

The present study extends that line of work by examining prediction rather than classification alone. It compares six models under a common one-year-ahead evaluation, retains persistence as a reference, maps method-specific residuals, and tests how the ranking changes when provinces affected by territorial discontinuity are removed from evaluation. ST-SCKM is included as an interpretable cluster-adjusted baseline. It is not presented as a new general-purpose learning architecture. Its role is to ask whether provinces with similar lagged socioeconomic conditions can share a modest expected annual change without sacrificing the province's most recent HDI level.

Three research questions guide the analysis. First, how stable is the spatial pattern of provincial HDI between 2017 and 2023 when administrative comparability is considered? Second, which of the six models improves on persistence in the 2022–2023 hold-out period? Third, what do residual maps, regional error profiles and the ST-SCKM typology add to the aggregate accuracy ranking? Answering these questions allows the article to distinguish predictive performance, spatial interpretation and planning usefulness rather than treating them as the same criterion.

The contribution is empirical and evaluative. Two hold-out years cannot settle the wider contest between classical and deep spatial temporal models, but a boundary-aware comparison shows how a short annual panel and administrative change alter planning conclusions.

Human Development Forecasting as a Regional Planning Problem

HDI condenses several dimensions of welfare into a common scale. In Indonesia, the index combines longevity, knowledge and a decent standard of living, following the broad framework introduced by the United Nations Development Programme and adapted by Statistics Indonesia (UNDP, 1990; BPS, 2024). Mean years of schooling, one of the covariates used here, is also part of the knowledge dimension. Labour force participation and open unemployment are not components of HDI, but they describe the economic setting in which education and living-standard gains are realised.

Spatial analysis helps with this reading because nearby or connected regions do not develop in isolation. Infrastructure corridors, migration, commodity markets, university systems and administrative capacity can create similarities that cross provincial borders. Moran's I offers a summary test of whether similar values are located near one another more often than expected under spatial randomness (Moran, 1950; Anselin, 1988). A positive statistic supports the use of spatial information, although its magnitude and significance depend on the selected weight matrix and the stability of the units being compared.

Forecasting adds a different question because a model may describe historical spatial association well and still perform poorly in a future year. Randomly dividing province-year observations would overstate temporal generalisation, so the present study uses a temporal hold-out and reports RMSE, MAE, R^2 and MAPE together.

For a planning reader, residual geography is as important as the mean metric. An aggregate RMSE can be low even when the remaining errors concentrate in one island group or type of province. Positive residuals mean that observed HDI is higher than predicted, while negative residuals indicate overprediction. Mapping these differences shows where a model may be less dependable and where a planner should seek additional local evidence before using a forecast.

Regional typology serves another purpose. Latuconsina et al. (2018) argued that differentiated regional characteristics are needed to formulate alternatives suited to local conditions. Otok et al. (2022) used unsupervised learning to identify groups of underdeveloped areas and then mapped those groups to support priority setting. The present study uses clustering more conservatively. The clusters organise

lagged socioeconomic profiles and supply a group-specific annual HDI increment, but the last observed provincial HDI remains the forecast anchor.

A cluster-adjusted persistence forecast can represent gradual annual change without estimating a highly flexible response surface. It should not be mistaken for evidence that the clusters are permanent, causally homogeneous or suitable as administrative regions.

Data and Methods

Data and Spatial Units

The analysis uses a balanced panel covering 34 Indonesian provinces from 2017 to 2023. The national aggregate is removed, leaving 238 province-year observations. HDI is the response variable. Open unemployment (TPT), labour force participation (TPAK) and mean years of schooling (RLS) are used to describe the socioeconomic conditions associated with each provincial trajectory. February and August values of TPT and TPAK are averaged to obtain annual measures.

The source file compiles indicators published by Statistics Indonesia (BPS, 2024). Mean HDI is 71.254, with a standard deviation of 3.903 and a range from 59.090 to 82.460. Mean TPT is 4.957 percent, mean TPAK is 68.862 percent, and mean RLS is 8.616 years. These are unweighted panel summaries.

The spatial layer is the 2017 ADM1 boundary from geoBoundaries, containing the same 34 provincial units represented at the beginning of the panel (Runfolo et al., 2020). Province centroids are derived from this layer. A four-nearest-neighbour graph links each province to the four closest centroids and is symmetrised for graph modelling. This approach avoids isolating island provinces, which can occur under strict polygon contiguity, but it also treats centroid distance as a proxy for geographical connection.

Papua requires separate treatment. The supplied panel records Papua at 60.62 in 2021 and 71.76 in 2022, an increase of 11.14 points. A contemporaneous BPS Papua release reported 61.39 for the 2022 province under the earlier territorial configuration (BPS Papua, 2022). The value in the supplied panel therefore represents a different geographical composition from the pre-2022 series. West Papua was also affected by the creation of new provinces, even though its numerical break is less pronounced.

The analysis does not silently replace these values because doing so would require a complete reconstruction of all post-split indicators on constant boundaries. Instead, two evaluation samples are reported. The full evaluation retains all 34 province labels and exposes the effect of the break. The stable-boundary sensitivity sample excludes Papua and West Papua from the calculation of hold-out metrics while retaining the models trained on the common 2017–2021 panel. The second sample is used for the main comparison because its units are more comparable over time.

Prediction Design

All methods are evaluated as rolling one-year-ahead forecasts. The target years 2019–2021 form the supervised training period after two lags are constructed, producing 102 province-year training cases. The hold-out contains 68 predictions for 2022 and 2023. The 2022 forecast uses information available through 2021, while the 2023 forecast uses information through 2022. This is a rolling evaluation rather than a two-year recursive forecast.

The common lagged information includes the previous year's TPT, TPAK, RLS and HDI, the latest annual change in HDI, and the spatially weighted mean of neighbouring HDI in the previous year. GTWR, the Bayesian model and ST-SCKM use these summary features directly. The deep-learning models receive two annual input frames containing TPT, TPAK, RLS and HDI for all provinces. All standardisation parameters are estimated from training years and then applied to the hold-out.

The design reduces information leakage because a prediction for year t does not use socioeconomic indicators measured in year t . It remains a rolling evaluation because the 2023 prediction uses the observed 2022 frame.

Persistence predicts the HDI of year t with the observed HDI of year t minus one. It is not counted among the six spatial temporal methods. Its purpose is to establish the gain that a model must achieve beyond carrying the latest value forward. Because HDI changes slowly, this is a stronger benchmark than the overall sample mean.

Spatial Dependence

Global Moran's I is calculated with the row-standardised four-nearest-neighbour matrix and 999 permutations. Separate matrices are built for the 34-province panel and the stable 32-province sample so excluded provinces do not retain graph links.

The same procedure is applied to 2023 residuals. A significant positive value would indicate clusters of underprediction and overprediction, while a value near zero is more consistent with isolated misses. With 34 or fewer units, non-significance does not rule out local patterns.

Table 1
Global Moran's I for provincial HDI under full and stable-boundary samples, 2017–2023

Year	Full I	Full p	Stable I	Stable p
2017	0.190	0.024	0.114	0.099
2018	0.190	0.025	0.112	0.104
2019	0.180	0.029	0.105	0.115
2020	0.181	0.029	0.103	0.112
2021	0.181	0.032	0.102	0.116
2022	0.129	0.073	0.106	0.113
2023	0.124	0.078	0.102	0.115

Source: Authors' calculations from BPS data. Stable sample excludes Papua and West Papua; p-values are based on 999 permutations.

The full sample shows positive and statistically significant spatial autocorrelation during 2017–2021, with Moran's I around 0.18. Significance weakens in 2022 and 2023. After Papua and West Papua are removed, Moran's I remains positive but falls to approximately 0.10–0.11 and is not significant at the five percent level (Table 1). The difference shows that the apparent national spatial signal is sensitive to the eastern boundary configuration and should not be used as an unconditional justification for increasingly complex spatial models.

Six Forecasting Methods

GTWR estimates local coefficients by weighting observations according to both geographical and temporal distance (Huang et al., 2010; Fotheringham et al., 2015). The six lagged predictors are standardised before fitting. A Gaussian kernel is used, with a fixed spatial bandwidth and a temporal scaling ratio selected by the package search procedure. The fitted training model has an R² of 0.995, which confirms its flexibility but also warns against judging it by in-sample fit. Predictions at hold-out locations are calculated from locally weighted least squares using training observations only.

The Bayesian spatial temporal model combines fixed effects for the same lagged predictors with a linear time term, a structured spatial effect and an unstructured province effect. The structured effect uses an exponential distance correlation matrix based on provincial centroids, while a non-centred parameterisation improves sampling stability. Four chains with 1,500 tuning and 1,500 retained draws are estimated using the No-U-Turn Sampler. All monitored summary parameters have R-hat values below 1.01 and effective sample sizes above 600. Posterior predictive intervals incorporate parameter uncertainty and observation noise.

ConvLSTM was designed for sequences on regular grids (Shi et al., 2015). Indonesian provinces form an irregular archipelago, so the model requires a 6 by 6 pseudo-grid. Provinces are ordered by latitude and longitude, placed in 34 cells, and the remaining two cells are padded. The architecture uses two ConvLSTM layers with 16 hidden channels and a one-by-one convolutional output head. The approximation gives convolutional operations a spatial arrangement, but adjacency on the pseudo-grid does not always represent actual geographical or functional proximity.

STGCN represents provinces as nodes on the fixed four-nearest-neighbour graph (Yu et al., 2018). A model block applies a gated temporal convolution, graph propagation through the normalised adjacency matrix, and a second temporal convolution. The final hidden state is mapped to one HDI prediction per province. This structure is more faithful to irregular spatial units than the ConvLSTM grid, although it depends heavily on whether centroid proximity captures the relationships that matter for human development.

GraphWaveNet adds a learned adjacency matrix to the fixed graph (Wu et al., 2019). Two node-embedding matrices generate directed adaptive weights, which are combined with propagation through the geographical graph. A gated causal temporal convolution and residual connection complete the simplified architecture. This model can represent similarities between non-neighbouring provinces, but the adaptive graph also introduces additional parameters that must be learned from only three training target windows.

ST-SCKM denotes the spatial temporal socioeconomic K-means baseline used in this study. It is a hard K-means procedure applied to standardised lagged socioeconomic features, previous HDI dynamics, centroid coordinates and target year. Spatial coordinates and time receive weights of 0.45 and 0.35 relative to the standardised socioeconomic block. Four clusters are specified before evaluation to preserve a planning-oriented differentiation of provincial profiles. The sensitivity check from k equals two to six shows that k equals two has the largest silhouette value, so the four-cluster solution is retained for interpretability rather than claimed as the uniquely optimal partition.

For prediction, each hold-out province-year is assigned to its nearest training centroid. The forecast equals the province's most recent HDI plus the mean annual HDI change observed for its assigned training cluster. This anchor prevents a province from being pulled directly to a distant cluster mean and makes the baseline comparable to persistence. Its additional information is the cluster-specific increment. The method is therefore transparent, but it cannot learn province-specific nonlinear responses or time-varying graph connections.

The three neural models are trained for 400 epochs with the Adam optimiser, a learning rate of 0.005 and weight decay of 0.001. Hidden dimensions and training settings are deliberately modest because the temporal sample is small. Random seeds are fixed at 42 for model initialisation and clustering. A single seed cannot describe the full variability of neural optimisation, so the deep-learning results should be interpreted as reproducible implementations under one specified configuration rather than definitive estimates of each architecture's best possible performance.

Evaluation and Diagnostic Strategy

Model performance is reported for the full and stable-boundary samples. The main ranking uses the stable sample. To show uncertainty, 5,000 province-level bootstrap samples are drawn with replacement. Both hold-out years for a sampled province are kept together, preserving within-province dependence. The bootstrap yields a 95 percent interval for RMSE and for the difference between each method and persistence.

Predictions and residuals are mapped for every method in 2023 using common scales. Regional MAE is then calculated for six island groups, which are descriptive and are not fitted as model inputs. The 2023 ST-SCKM assignments are summarised by HDI, TPT, TPAK and RLS.

No causal interpretation is attached to the predictors. A positive or negative model coefficient can reflect correlation, temporal persistence or omitted regional processes. The purpose is forecast comparison and diagnostic mapping. Policy implications are therefore framed in terms of monitoring, model choice and the need for additional evidence, not claims that changing unemployment or schooling by a given amount would mechanically change HDI.

Results

Provincial Pattern and Boundary Discontinuity

Provincial HDI levels in 2023 remain uneven (Fig. 1). The highest values occur in Jakarta and the Special Region of Yogyakarta, while lower values appear in parts of eastern Indonesia. Several provinces outside Java occupy the middle and upper parts of the distribution, so the map does not support a simple binary division between Java and the rest of the country. It shows a combination of metropolitan advantage, provincial education profiles and persistent peripheral constraints.

Source: Authors' calculations from BPS data and geoBoundaries ADM1 geometry.

The change map needs a different reading. For 33 provinces, the increase from 2017 to 2023 lies between 2.18 and 3.62 HDI points. Papua records 13.32 points in the supplied panel and therefore sits far outside the rest of the distribution. The classified colours in Fig. 1 prevent this value from washing out the smaller differences among other provinces, while the annotation keeps the discontinuity visible. The map should not be interpreted as evidence that Papua achieved an internally comparable 13-point improvement.

A high 2023 level may reflect accumulated advantage, while a large change can indicate catch-up, a statistical revision or a boundary shift. Reading both panels prevents the Papua discontinuity from being mistaken for an unqualified development gain.

The Moran results reinforce that caution. The full sample has a modest positive spatial pattern before the hold-out, but the signal becomes weaker and statistically uncertain after the affected Papuan units are excluded (Table 1). Spatial models remain worth testing because they encode local and network information, but the data do not support a strong claim that provincial HDI is governed by stable global clustering throughout the period.

Predictive Performance

The model ranking changes in magnitude, but not substantially in order, when the stable-boundary sample replaces the full panel (Table 2 and Fig. 2). On the full sample, ST-SCKM and GTWR record RMSE values of 1.345 and 1.352. Persistence follows at 1.490. Much of this error comes from the 2022 Papua break, which none of the models could infer from pre-2022 data.

Table 2
One-year-ahead forecasting performance for 2022–2023

Model	Full RMSE	Stable RMSE	Stable RMSE 95% CI	Stable MAE	Stable R2
ST-SCKM	1.345	0.328	0.303–0.353	0.314	0.990
GTWR	1.352	0.379	0.343–0.416	0.358	0.987
Persistence (benchmark)	1.490	0.630	0.604–0.655	0.622	0.963
ConvLSTM	1.958	0.744	0.652–0.830	0.641	0.949
GraphWaveNet	1.723	0.824	0.754–0.900	0.781	0.937
Bayesian-ST	2.333	1.443	1.395–1.493	1.383	0.807
STGCN	4.201	3.945	2.534–5.219	2.632	-0.441

Source: Authors' calculations. The stable-boundary sample excludes Papua and West Papua from evaluation. Confidence intervals use 5,000 province-level bootstrap samples.

On the stable sample, ST-SCKM has the lowest RMSE at 0.328, followed by GTWR at 0.379. Persistence has an RMSE of 0.630. The bootstrap intervals for ST-SCKM and GTWR do not overlap the persistence interval, and their RMSE differences relative to persistence remain below zero across the 95 percent bootstrap ranges. The two methods therefore add predictive information in this evaluation, but their ranking is still tied to a short hold-out and one specification of the spatial graph and clusters.

Source: Authors' calculations.

ST-SCKM adds a training-period increment of roughly 0.30 to 0.40 points to the latest HDI, depending on cluster. Most stable provinces improved within that range. Its success comes from adjusting a strong persistence anchor, not from uncovering a hidden law of development.

GTWR reaches a similar conclusion through a different mechanism. Its local weighting combines lagged provincial conditions, recent HDI change and neighbouring HDI. The model is especially accurate in 2023, when its RMSE is 0.430 for the full sample. Its 2022 full-sample RMSE is 1.864 because Papua is underpredicted by 10.69 points. Once the affected Papuan units are removed, GTWR errors are small and geographically dispersed.

The neural models do not improve on persistence in the stable sample. ConvLSTM has an RMSE of 0.744 and GraphWaveNet 0.824, although both explain more than 93 percent of stable-sample hold-out variation by R^2 . Their relatively high R^2 coexists with worse RMSE because provincial rank and level differences are captured, while one-year changes are not estimated as closely as by the simpler baselines. In a slowly changing index, a model can reproduce the cross-sectional hierarchy yet still lose to persistence on annual increments.

The Bayesian model has an RMSE of 1.443 on the stable sample. Its posterior intervals provide information that point models do not, but the posterior mean is systematically conservative in 2023, with a full-sample mean residual of 2.00 points. The model's explicit linear time term and partial pooling do not adapt quickly enough to the observed 2022–2023 levels under this specification. Its value therefore lies more in uncertainty representation than point accuracy here.

STGCN performs worst, with a stable-sample RMSE of 3.945 and a negative R^2 . The fixed centroid-based graph propagates information among neighbours, but the model has only three training target windows. Large 2023 errors occur in Jakarta, Yogyakarta and parts of Kalimantan, indicating that the architecture smooths provinces whose levels differ sharply from their geographical neighbours. This is a data-architecture mismatch rather than evidence that graph convolution is generally unsuitable for regional forecasting.

Method Specific Prediction and Residual Maps

The 2023 maps show that all models recover the broad provincial hierarchy, but their residual fields differ markedly (Fig. 3). ST-SCKM and GTWR produce predictions close to the observed values in most provinces. Their 2023 maximum absolute errors are 0.701 and 0.751 points, respectively. The largest ST-SCKM error occurs in West Papua, while the largest GTWR error occurs in Jakarta.

Source: Authors' calculations from BPS data and geoBoundaries ADM1 geometry.

GraphWaveNet, ConvLSTM and the Bayesian model display a common underprediction pattern. Their mean 2023 residuals are 1.08, 1.18 and 2.00 points. Papua is the largest error for all three, but underprediction is also visible in several provinces with high or rapidly improving HDI. The shared direction suggests that the models learn the earlier level structure more readily than the later upward shift.

STGCN has a different residual pattern. It underpredicts Jakarta and Yogyakarta by more than 11 points while overpredicting West Kalimantan by approximately 9.6 points. These are not small deviations

around a generally accurate surface. They indicate excessive smoothing through the fixed graph and unstable learning from the short sequence. A planner using only a national RMSE would miss the particular types of regions where the model fails.

Residual Moran's I is not statistically significant for any model in 2023 (Table 3). The values range from minus 0.119 for GTWR to 0.108 for STGCN. This means that the remaining errors do not form a strong global cluster under the four-nearest-neighbour definition. It does not mean that the errors are unimportant. The method maps reveal isolated but policy-relevant misses that a global statistic averages away.

Table 3
Spatial diagnostics and magnitude of 2023 prediction residuals

Model	Moran's I	p	Mean bias	Maximum absolute error	Province
ST-SCKM	-0.051	0.436	0.344	0.701	West Papua
GTWR	-0.119	0.209	0.407	0.751	Jakarta
Persistence (benchmark)	0.046	0.219	0.654	1.000	West Papua
ConvLSTM	-0.046	0.302	1.181	10.499	Papua
GraphWaveNet	-0.008	0.282	1.078	6.102	Papua
Bayesian-ST	-0.025	0.414	2.004	10.250	Papua
STGCN	0.108	0.104	0.921	12.129	Jakarta

Source: Authors' calculations. Residual is observed minus predicted HDI; Moran p-values use 999 permutations.

GTWR and ST-SCKM do not gain only by removing an east-west residual pattern. Their advantage comes from smaller province-level errors across most of the country, while STGCN's large misses occur in distant regions.

ST SCKM Typology

The 2023 ST-SCKM map divides provinces into four profiles (Fig. 4 and Table 4). Cluster 1 contains East Nusa Tenggara and Papua. It has the lowest mean HDI at 69.545, the lowest mean years of schooling at 7.485, and the highest labour force participation at 76.055 percent. The profile shows why high participation should not be read automatically as strong labour-market opportunity. In provinces with lower schooling and substantial informal employment, participation can remain high while human development outcomes lag.

Source: Authors' calculations from BPS data and geoBoundaries ADM1 geometry.

Cluster 2 contains Maluku, North Maluku, South Sulawesi and North Sulawesi. Its mean HDI is 72.243 and mean RLS is 9.498 years, but TPAK is the lowest among the four clusters at 64.893 percent. The combination suggests an education-intensive profile with weaker labour-force participation. The cluster is geographically concentrated in eastern island provinces, yet it should be treated as a shared statistical profile rather than a single functional planning region.

Cluster 3 is the largest and most heterogeneous group, containing 19 provinces across Java, Bali, Kalimantan, Sulawesi, Sumatra, Nusa Tenggara, Maluku and West Papua. Its mean HDI is 72.148, mean TPT is 3.959 percent, and mean RLS is 8.567 years. The broad membership is a reminder that a four-cluster solution simplifies substantial internal variation. For planning, this cluster is better understood as the middle of the provincial distribution than as a uniform development regime.

Cluster 4 contains nine provinces, including Jakarta, West Java, Banten, several Sumatran provinces, the Riau Islands and East Kalimantan. It has the highest mean HDI at 75.549 and the highest mean RLS at 9.756 years. It also has the highest mean open unemployment rate at 6.347 percent and lower participation than Cluster 3. The profile is consistent with more urbanised and education-intensive labour markets where job search and formal-sector entry can raise measured unemployment even when human development is comparatively high.

Table 4
ST-SCKM cluster profiles for 2023

Cluster	Interpretive profile	n	HDI	TPT (%)	TPAK (%)	RLS (years)
1	Lower HDI, high participation	2	69.545	3.100	76.055	7.485
2	Education intensive, lower participation	4	72.242	5.397	64.892	9.498
3	Broad middle	19	72.148	3.959	70.543	8.567
4	Higher HDI, urban labour pressure	9	75.549	6.347	66.658	9.756

Source: Authors' calculations. Labels are interpretive summaries, not formal administrative categories.

The map explains the cluster increment used by ST-SCKM and identifies similar profiles across distant locations. It should not be used to assign identical policy packages, which still require district evidence and institutional context.

The sensitivity statistics also limit the interpretation. The silhouette value is 0.260 for two clusters and 0.196 for four clusters. Four groups were retained to avoid an overly coarse high-low division and to support differentiated discussion, but the partition is not sharply separated. This choice should be revisited with a longer panel or additional indicators before the typology is used operationally.

Regional Distribution of Error

The regional MAE heatmap shows that model performance is not spatially uniform (Fig. 5). ST-SCKM and GTWR have MAE values below 0.42 in Sumatra, Java and Bali, Nusa Tenggara, Kalimantan and Sulawesi. Their higher values in Maluku and Papua, 1.69 and 1.76, reflect the inclusion of the Papua break in the regional average. When Papua and West Papua are removed, the two methods retain their low national errors.

Source: Authors' calculations.

ConvLSTM and GraphWaveNet are reasonably stable across the five western and central regional groups, with MAE mostly between 0.50 and 1.02. Their errors rise to 2.94 and 2.72 in Maluku and Papua. The Bayesian model has a more uniform but larger underprediction, with MAE around 1.28–1.45 outside the eastern group and 3.63 in Maluku and Papua. This pattern suggests that partial pooling reduces sharp regional variation in error but does not correct the level bias.

STGCN is the least stable geographically. Its MAE reaches 4.81 in Java and Bali, 3.76 in Nusa Tenggara, 3.73 in Kalimantan and 3.52 in Maluku and Papua, while Sumatra is much lower at 1.16. The contrast is visible in the residual map and confirms that one graph configuration does not fit all regional relationships. Centroid proximity may connect provinces that are geographically near but structurally dissimilar, particularly around major metropolitan and island systems.

Regional aggregation does not identify causes, but it shows where a national score hides errors that require local validation.

Discussion

What the Comparison Shows

The central finding is modest but useful. In this short annual panel, the two most accurate forecasts are structurally simple. ST-SCKM adds a cluster-specific increment to the last observed HDI, and GTWR estimates a local relationship from lagged provincial and neighbouring conditions. Both improve on persistence in the stable-boundary sample. The deep-learning models capture much of the cross-sectional ordering but do not improve one-year-ahead accuracy relative to carrying the previous value forward.

This is not a general victory of simple models. Three supervised training target windows provide little evidence for temporal filters, recurrent gates or adaptive node embeddings, even when those architectures suit denser transport, weather or sensor data.

ST-SCKM benefits from a stronger prior assumption. It assumes that the most recent provincial level is informative and that annual change varies only modestly across four profiles. For a slowly moving index, that assumption is effective. The method would be less suitable for an outcome with abrupt local

shocks or rapidly changing seasonal structure. Its first-place ranking is evidence about annual provincial HDI in this evaluation, not proof that cluster-adjusted persistence is broadly superior to spatial learning.

GTWR provides a useful second result because it reaches nearly the same accuracy without relying on a discrete typology. Its local weighting can accommodate variation in the relationship between prior schooling, labour-market conditions, neighbouring HDI and the current outcome. At the same time, its training R^2 of 0.995 indicates substantial flexibility. A longer out-of-sample period is needed before local coefficient surfaces are interpreted substantively.

STGCN's result is the clearest warning against equating spatial structure with predictive gain. A graph makes the assumed relationships explicit, but the selected edges may be wrong for the process. Human development can be linked through migration, fiscal transfers, transport, commodity markets and institutional similarity, none of which is fully represented by centroid distance. Adding graph convolution to a weak proxy graph can transmit the wrong information more efficiently.

Implications for Regional Planning

The first practical implication concerns boundary management. A regional monitoring system should attach versioned geometries and stable-unit identifiers to every annual observation. When a province splits, the statistical system should either backcast the new units, aggregate them to the former boundary, or flag the break explicitly. Without this step, a forecast comparison may rank models by their inability to predict a redefinition rather than their ability to predict development.

The Papua case demonstrates the scale of that risk. The supplied series shows an 11.14-point increase between 2021 and 2022, while the contemporaneous BPS release for the earlier provincial territory reports an increase of 0.77 points (BPS Papua, 2022). No model trained on ordinary annual increments should be expected to predict the larger change. The appropriate response is not to increase model capacity. It is to reconcile the territorial definition.

The second implication concerns residual maps. A predicted HDI map often resembles the observed level map and can create a false sense of accuracy. The residual map is more informative for review because it directs attention to provinces where a forecast needs additional explanation. Here, it reveals underprediction in Papua and excessive STGCN smoothing around high-HDI provinces.

The third implication concerns regional typology. The JRCP studies by Latuconsina et al. (2018) and Otok et al. (2022) show that grouping becomes useful when it connects indicators to differentiated planning questions. The four ST-SCKM profiles serve that role at a preliminary level by distinguishing questions about schooling, labour-force entry, diversity within the middle group and unemployment in more urbanised labour markets.

These interpretations remain diagnostic. Baransano et al. (2016) showed that inequality in West Papua is shaped by sectoral and within-region differences that a provincial average cannot resolve. The same caution applies here. A cluster assignment does not specify which programme should be funded or

which district should receive priority. It identifies a set of relationships that should be investigated with more detailed evidence.

Administrative Change and the Meaning of Innovation

The comparison also speaks to methodological innovation in regional research. Innovation is often associated with a newer algorithm, but a novel architecture does not automatically answer a more important planning question. In this case, the main analytical gain comes from aligning evaluation with boundary stability, using a credible benchmark and examining the geography of error. These choices alter the conclusion more than adding another hidden layer.

GraphWaveNet's adaptive adjacency remains attractive because Indonesian regional links are not limited to nearest neighbours. District data, quarterly indicators or longer histories would allow that premise to be tested more convincingly.

The boundary sensitivity result is particularly relevant to emerging and transitional economies, where administrative reorganisation is common and statistical capacity varies. Regional models often assume that place names represent stable units. A boundary-aware workflow turns that assumption into a testable data condition. It also makes comparisons across studies clearer because readers can see whether a reported gain depends on harmonised geography.

Limitations

The study has several limitations. The first is temporal depth. Seven annual observations provide only three training target windows after two lags are used. The bootstrap quantifies uncertainty across provinces within the hold-out, but it cannot create additional independent years. A longer series is required to assess whether the observed ranking persists through different economic conditions.

The second limitation is spatial scale. Provinces are large and internally diverse, especially in Kalimantan, Sulawesi, Maluku and Papua. Provincial averages can hide district-level deprivation and metropolitan-rural contrasts. A district panel would provide more observations and greater planning relevance, but it would also require extensive boundary harmonisation and treatment of newly created districts.

The third limitation is the spatial graph. Four nearest centroid neighbours provide a reproducible graph for an archipelago, but they do not measure travel time, migration, trade, fiscal links or digital connectivity. Alternative graphs may change Moran's I and the performance of STGCN and GraphWaveNet. Future research should compare geographical, transport and socioeconomic similarity graphs rather than treating one matrix as definitive.

The fourth limitation concerns model tuning. Neural architectures are evaluated under a fixed compact configuration and random seed. More extensive hyperparameter search could improve their scores, but aggressive tuning against two hold-out years would itself risk overfitting. A longer rolling-origin evaluation is needed before large model searches are justified.

The fifth limitation concerns ST-SCKM. The four-cluster solution is chosen for interpretability despite a lower silhouette score than the two-cluster solution. The coordinate and time weights are fixed rather than estimated. The resulting typology is useful for reading the forecast, but it should not be treated as a validated regionalisation scheme.

Finally, the analysis does not estimate causal effects. TPT, TPAK and RLS are predictive covariates and may be jointly determined with HDI or related to omitted institutional factors. Policy should not be inferred from coefficient signs alone. The results support choices about monitoring and forecasting, while programme design requires causal and place-specific evidence.

Conclusions

This study compared six spatial temporal approaches for one-year-ahead prediction of provincial HDI in Indonesia and retained persistence as a reference. The comparison was conducted on the 2022–2023 hold-out and repeated on a stable-boundary sample that excludes Papua and West Papua. That distinction is essential because the post-2021 territorial reorganisation breaks the comparability of the Papuan series.

The stable-sample results favour modest models. ST-SCKM and GTWR produce RMSE values of 0.328 and 0.379, compared with 0.630 for persistence. ConvLSTM, GraphWaveNet, the Bayesian model and STGCN do not improve on persistence under the selected configurations. The result indicates that a strong level anchor and limited local adjustment fit this short annual panel better than architectures requiring richer temporal evidence.

The maps add information that the ranking alone cannot provide. They show which provinces drive error, whether a model smooths distinct regions excessively, and how the full-sample result is affected by Papua. Residual Moran's I is not significant in 2023, so the remaining errors are better understood as isolated provincial misses than as one national residual cluster. Regional MAE still rises in Maluku and Papua when the discontinuous units are retained.

ST-SCKM also provides a readable typology, but its role should remain limited. The four clusters identify combinations of HDI, education and labour-market conditions that can organise further inquiry. They do not define fixed development regions or prescribe common interventions. Their predictive success comes mainly from adding a small cluster increment to the latest provincial HDI.

For regional planning, the main lesson is procedural. Near-term forecasts should be evaluated against persistence, mapped as residuals, and checked on stable territorial units before model complexity is interpreted as progress. This workflow offers a more dependable basis for using provincial indicators while keeping statistical uncertainty and administrative change visible.

Declarations

Funding and Conflict of Interest

No external funding was received for this study. The author declares no conflict of interest.

Data and Code Availability

The analytical data, model code, prediction tables, figures and the 34-province boundary layer are available as supplementary materials. The boundary layer is distributed under the terms stated by geoBoundaries. Users should retain the stable-boundary warning when reusing the 2022-2023 Papuan observations.

Declaration of Generative AI and AI Assisted Technologies in the Manuscript Preparation Process

During the preparation of this work, the author used OpenAI Codex to assist with analytical code verification, figure preparation and improvement of language clarity. The author reviewed the code, results and text and takes full responsibility for the content of the article.

Author Contribution

M.A.I. conceived the study, curated the data, developed and implemented the analytical workflow, prepared the figures and tables, and drafted the manuscript. I.G.N.S.P. contributed to the methodology, validation, interpretation of the results, and critical revision of the manuscript. Both authors reviewed and approved the final manuscript.

References

1. Anselin, L. (1988). *Spatial Econometrics Methods and Models*. Kluwer Academic.
2. Badan Pusat Statistik (BPS) (2024). Indeks Pembangunan Manusia 2023. Jakarta: Badan Pusat Statistik. Available from <https://www.bps.go.id/id/publication/2024/05/13/8f77e73a66a6f484c655985a/indeks-pembangunan-manusia-2024.html>
3. Badan Pusat Statistik Provinsi Papua (BPS Papua) (2022). Indeks Pembangunan Manusia Provinsi Papua Tahun 2022. Jayapura: BPS Provinsi Papua. Available from <https://papua.bps.go.id/id/pressrelease/2022/12/01/658/indeks-pembangunan-manusia--ipm--prov>
4. Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data*. Second edition. Boca Raton: CRC Press.
5. Baransano, M. A., Putri, E. I. K., Achsani, N. A., & Kolopaking, L. (2016). The Role of Leading Sector to Reduce Regional Development Inequality in West Papua Province. *Journal of Regional and City Planning*, 27(2), 119–136. <https://doi.org/10.5614/jrcp.2016.27.2.4>
6. Birant, D., & Kut, A. (2007). ST-DBSCAN An Algorithm for Clustering Spatial Temporal Data. *Data and Knowledge Engineering* 60(1), 208–221. <https://doi.org/10.1016/j.datak.2006.01.013>.

7. Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically Weighted Regression A Method for Exploring Spatial Nonstationarity. *Geographical Analysis*, 28(4), 281–298. <https://doi.org/10.1111/j.1538-4632.1996.tb00936.x>
8. Cressie, N., & Wikle, C. K. (2011). *Statistics for Spatio Temporal Data*. Wiley.
9. Elhorst, J. P. (2014). *Spatial Econometrics From Cross Sectional Data to Spatial Panels*. Springer.
10. Fotheringham, A. S., Crespo, R., & Yao, J. (2015). Geographical and Temporal Weighted Regression. *Geographical Analysis*, 47(4), 431–452. <https://doi.org/10.1111/gean.12071>
11. Getis, A., & Aldstadt, J. (2004). Constructing the Spatial Weights Matrix Using a Local Statistic. *Geographical Analysis*, 36(2), 90–104. <https://doi.org/10.1111/j.1538-4632.2004.tb01127.x>
12. Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn Sampler Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15, 1593–1623.
13. Huang, B., Wu, B., & Barry, M. (2010). Geographically and Temporally Weighted Regression for Modeling Spatio Temporal Variation in House Prices. *International Journal of Geographical Information Science*, 24(3), 383–401. <https://doi.org/10.1080/13658810802672469>
14. Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting Principles and Practice*. Third edition. Melbourne: OTexts. Available from <https://otexts.com/fpp3/>
15. Jain, A. K. (2010). Data Clustering 50 Years Beyond K-Means. *Pattern Recognition Letters* 31(8), 651–666. <https://doi.org/10.1016/j.patrec.2009.09.011>.
16. Latuconsina, Z. M. Y., & Rustiadi, E., and Sahara (2018). An Analysis of the Level of Development in Malang Regency Based on a Typology of Development Regions. *Journal of Regional and City Planning*, 29(1), 1–17. <https://doi.org/10.5614/jrcp.2018.29.1.1>
17. MacQueen, J. (1967). Some Methods for Classification and Analysis of Multivariate Observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* 1, 281–297.
18. Meyer, H., Reudenbach, C., Wollenberg, K., Hengl, T., & Katurji, M. (2019). Improving Performance of Spatio Temporal Machine Learning Models Using Forward Feature Selection and Target Oriented Validation. *Environmental Modelling and Software*, 101, 1–9. <https://doi.org/10.1016/j.envsoft.2017.12.001>
19. Moran, P. A. P. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika*, 37(1/2), 17–23. <https://doi.org/10.2307/2332142>
20. Otok, B. W., Suharsono, A., Purhadi, Standsyah, R. E., & Al Azies, H. (2022). Partitional Clustering of Underdeveloped Area Infrastructure with Unsupervised Learning Approach A Case Study in the Island of Java Indonesia. *Journal of Regional and City Planning*, 33(2), 177–196. <https://doi.org/10.5614/jpwk.2022.33.2.3>
21. Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Aroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schroder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Cross Validation Strategies for Data with Temporal Spatial Hierarchical or Phylogenetic Structure. *Ecography*, 40(8), 913–929. <https://doi.org/10.1111/ecog.02881>

22. Runfola, D., Anderson, A., Baier, H., Crittenden, M., Dowker, E., Fuhrig, S., Goodman, S., Grimsley, G., Layko, R., Melville, G., Mulder, A., Oberman, R., Panganiban, J., Peck, A., Seitz, L., Shea, S., & Hobbs, L. (2020). geoBoundaries A Global Database of Political Administrative Boundaries. *Plos One*, 15(4), e0231866. <https://doi.org/10.1371/journal.pone.0231866>
23. Shi, X., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., & Woo, W. C. (2015). Convolutional LSTM Network A Machine Learning Approach for Precipitation Nowcasting. *Advances in Neural Information Processing Systems*, 28, 802–810.
24. United Nations Development Programme (UNDP). (1990). *Human Development Report 1990*. Oxford University Press.
25. Wu, Z., Pan, S., Long, G., Jiang, J., & Zhang, C. (2019). Graph WaveNet for Deep Spatial Temporal Graph Modeling. Proceedings of the Twenty Eighth International Joint Conference on Artificial Intelligence, 1907–1913. <https://doi.org/10.24963/ijcai.2019/264>
26. Yu, B., Yin, H., & Zhu, Z. (2018). Spatio Temporal Graph Convolutional Networks A Deep Learning Framework for Traffic Forecasting. Proceedings of the Twenty Seventh International Joint Conference on Artificial Intelligence, 3634–3640. <https://doi.org/10.24963/ijcai.2018/505>

Figures

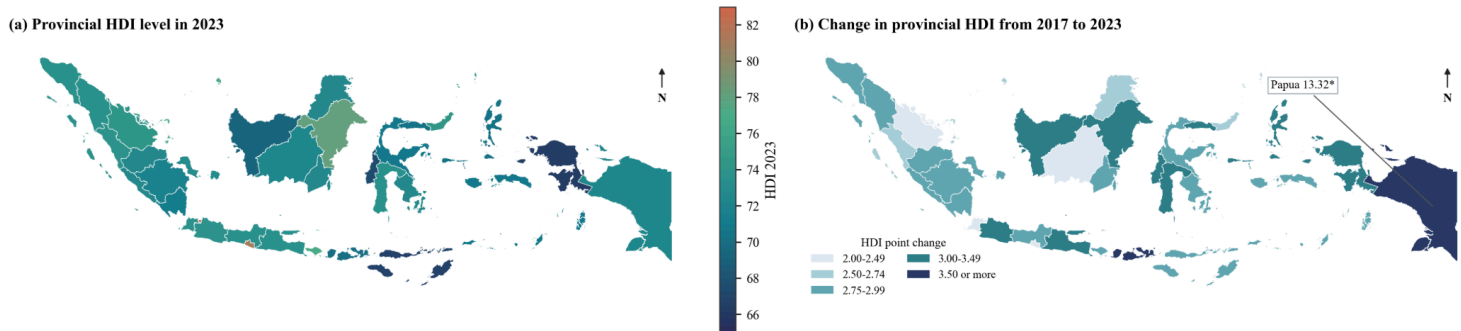


Figure 1

Provincial HDI level in 2023 and change from 2017 to 2023. Explanation: the asterisk marks Papua, where the supplied series is affected by a territorial discontinuity after 2021.

Source: Authors' calculations from BPS data and geoBoundaries ADM1 geometry.

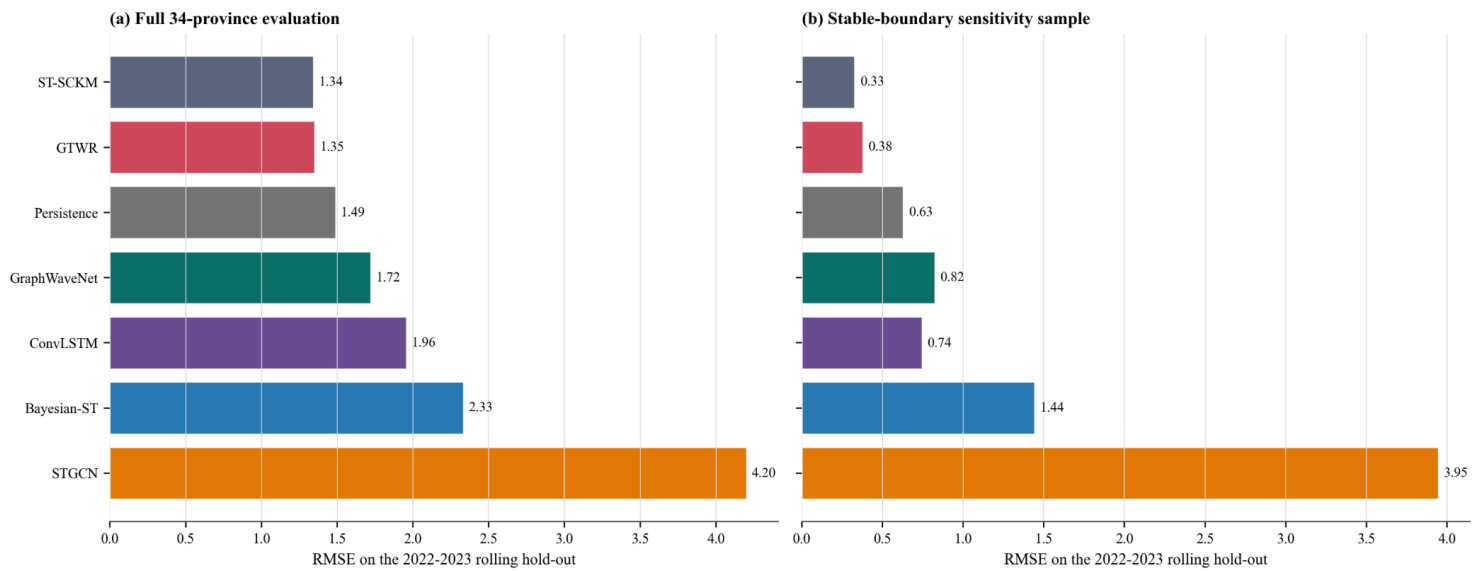


Figure 2

One-year-ahead RMSE for the full and stable-boundary evaluation samples. Explanation: lower values indicate better performance; the stable sample excludes Papua and West Papua.

Source: Authors' calculations.

1. ST-SCKM predicted HDI



ST-SCKM residuals RMSE 1.34



2. GTWR predicted HDI



GTWR residuals RMSE 1.35



3. GraphWaveNet predicted HDI



GraphWaveNet residuals RMSE 1.72



4. ConvLSTM predicted HDI



ConvLSTM residuals RMSE 1.96



5. Bayesian-ST predicted HDI



Bayesian-ST residuals RMSE 2.33



6. STGCN predicted HDI



STGCN residuals RMSE 4.20

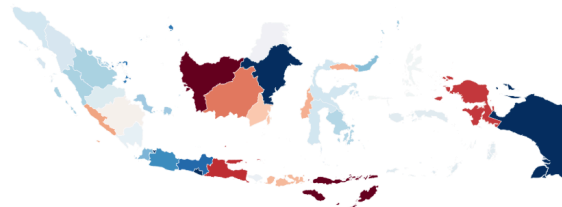


Figure 3

Predicted provincial HDI and residuals by method in 2023. Explanation: residual equals observed minus predicted HDI; positive values indicate underprediction; all methods use common colour scales.

Source: Authors' calculations from BPS data and geoBoundaries ADM1 geometry.

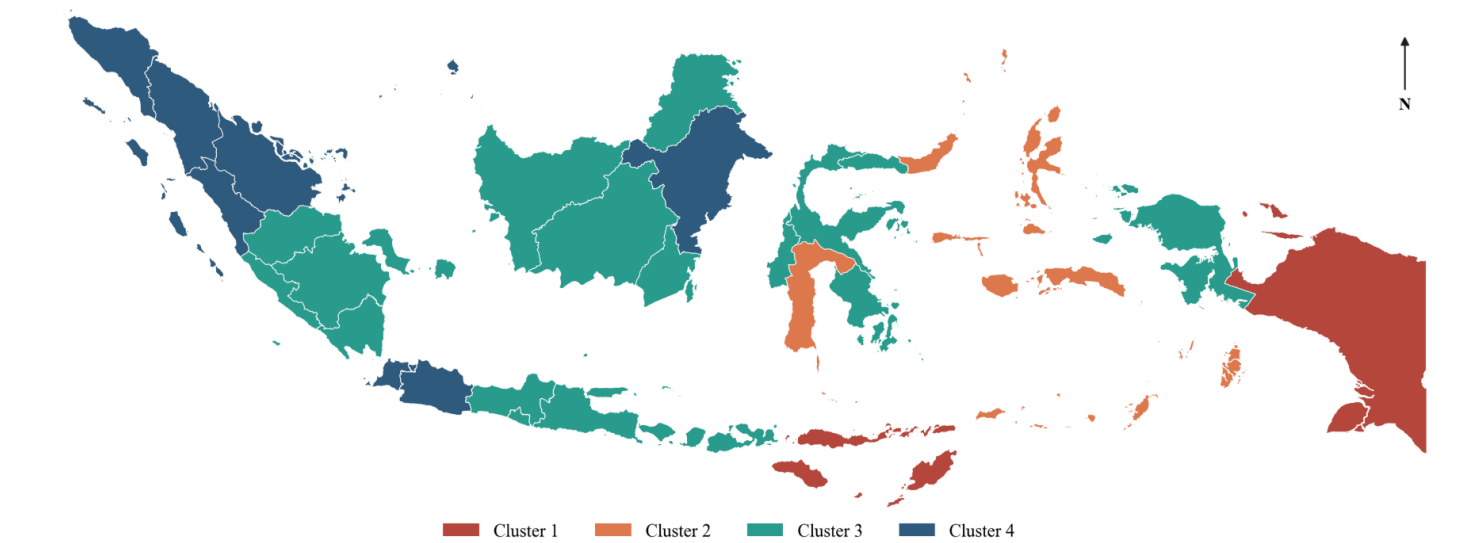


Figure 4

ST-SCKM provincial typology in 2023. Explanation: 1 - lower HDI and high participation; 2 - education intensive and lower participation; 3 - broad middle; 4 - higher HDI and urban labour pressure.

Source: Authors' calculations from BPS data and geoBoundaries ADM1 geometry.



Figure 5

Mean absolute prediction error by broad regional group and method, 2022-2023. Explanation: darker cells indicate larger errors; the displayed regional estimates retain all 34 province labels.

Source: Authors' calculations.