CAUCHY – Jurnal Matematika Murni dan Aplikasi
Volume 11 (2) (2026), Pages 2391–2410
p-ISSN: 2086-0382; e-ISSN: 2477-3344

Panel Data Break the Diagnostics of Geographically Weighted Regression: Five Failures, Their Repair, and a Rice-Productivity Application in Indonesia

Rahmawati Tri Iriani and Danang Ariyanto*

Department of Mathematics, Universitas Negeri Surabaya, Indonesia

Abstract

Geographically weighted panel regression (GWPR) is assembled from cross-sectional tools. Five of them fail on a balanced panel whose units keep their coordinates over time. Each failure is stated as a proposition, derived in closed form and repaired. An adaptive bandwidth in rows is not identified. A fold that holds out a whole unit cannot score a fixed-effects model without a plug-in level. A grid counted in units saturates at the inner training size, where its tie rule picks the bandwidth. A local and a global model share one information criterion only through a composite smoother. The Breusch–Pagan statistic reduces to a constant set by the sample size. The repairs are applied to 272 province-years, 34 Indonesian provinces from 2018 to 2025, with log yield as the outcome. Under nested leave-one-province-out cross-validation, the global two-way fixed-effects model attains a cross-validated R^2 of 0.100 and the GWPR 0.064. The GWPR loses in every variant tried. A province-block permutation test cannot distinguish Indonesia's real map from a random relabelling, and none of the 136 local slopes differs from its global counterpart. The R^2 of 0.9997 in the earlier version of this study comes from the accounting identity between production and harvested area.

Keywords: bandwidth identification; geographically weighted panel regression; grouped cross-validation; heteroskedasticity testing; rice productivity

Copyright © 2026 by Authors, Published by CAUCHY Group. This is an open access article under the CC BY-SA License (https://creativecommons.org/licenses/by-sa/4.0)

1. Introduction

Rice is Indonesia's staple food and central to national food security, and a growing population needs sustained production to keep the country self-sufficient [1, 2]. The sector faces land conversion, climate change and falling production capacity in several regions [3]. BPS reports production as GKG (*gabah kering giling*), which is dried unhusked paddy ready for milling. Its data show national rice production falling from 59.20 million tonnes of GKG (2018) to 54.60 million tonnes (2019) and 53.14 million tonnes (2024) [4], before rising to a provisional 60.21 million tonnes in 2025 [5].

Production is the product of harvested area and productivity, and the two carry different information. Harvested area behaves like an accounting record, so regressing production on it mainly reconstructs that identity rather than revealing agronomic relationships. Productivity depends on climate, pest pressure and farm management. This study therefore models productivity

*Corresponding author. E-mail: danangariyanto@unesa.ac.id

Received: June 22, 2026 Revised: October 12, 2026 Accepted: October 20, 2026
DOI: https://doi.org/10.18860/cauchy.v11i2.44351

and keeps harvested area as a scale control. The distinction comes first because it determines interpretation. A strong fit to production can be inherited entirely from the identity and say nothing about geography.

Indonesian provinces differ in geography, land, agricultural infrastructure and cultivation practice, so the relationships between rice productivity and its determinants may vary by region [6]. Geographically weighted regression was built to estimate such location-specific relationships. Whether the data support varying coefficients should be tested rather than assumed, and for panel data the tools for that test must first be corrected. This study tests it on a panel of 34 Indonesian provinces observed annually from 2018 to 2025. Rainfall supports water availability, and irregular rainfall disrupts planting [7, 8]; temperature matters beyond the optimal range [9], and pests and diseases damage crops [10, 11]. These effects may differ by province through irrigation, planting calendars, local climate and farm management, and earlier studies found that rice production does not respond uniformly across space [8, 9, 12, 13].

Earlier panel and spatial studies of rice were limited in space, time or outcome. Utari et al. [14] fitted a fixed-effects panel model to 19 sub-districts of Ngawi Regency (2006–2014) with constant slopes. Wicaksono et al. [15] fitted GWR to 34 provinces in 2021 and Astari et al. [16] fitted geographically weighted lasso to 2019 provincial data, each for a single year. Gamayanti et al. [17] fitted fixed-effects GWPR to one province, Central Sulawesi (2014–2020), with a fixed exponential kernel. Sifriyani et al. [18] compared weighting functions in a within-estimator GWPR across 34 provinces, but for a food security index rather than rice productivity.

GWR was designed for a single cross-section [19], and so were its diagnostics, bandwidth selection, validation and model comparison. Applied directly to panel data, as in much applied GWPR work including the earlier version of this study, five of these tools fail because of how the estimator works, not because of any dataset. This paper states those five failures as propositions (P1–P5 of Section 2) and derives each in closed form for a balanced $n \times T$ panel with repeated coordinates. Each carries a repair: a bandwidth defined and searched in units rather than rows (P1); a grouped fold with a plug-in level from the held-out unit's other periods (P2); a bandwidth grid indexed by a fraction of the training units, whose ceiling is fold-independent (P3); a composite smoother that places a local and a global panel model on one information criterion (P4); and the studentised Koenker form of the heteroskedasticity test in place of a statistic that collapses to a constant (P5). The contribution is methodological. The Indonesian rice panel is where the failures were found and the repairs demonstrated. Its substantive finding, that the local model earns nothing over a global two-way fixed-effects specification out of sample, follows from the corrected procedure and is not the claim the paper rests on.

2. Methods

This section has five parts. §2.1 builds the panel. §2.2 specifies the global two-way fixed-effects model that GWPR is judged against. §2.3 states five propositions, P1 to P5, on where GWR fails on panel data, each with its remedy. §2.4 assembles the remedies into the corrected procedure, and §2.5 lists the prohibitions that bind the analysis.

2.1. Data and panel construction

2.1.1. Frame, span and unit of observation

The panel is balanced, with $n = 34$ provinces $\times$ $T = 8$ years (2018–2025) and $nT = 272$ province-year observations on the pre-split 34-province frame. BPS issues two vintages of rice production and harvested area, a provisional figure within the reference year and a final figure in later tables. The panel uses the final vintage wherever one exists: the 2018 figures come from *Statistik Indonesia 2020*, Table 5.1.1 (pp. 269–270) [20], and the 2019–2024 figures come from the corresponding editions. An earlier draft of this panel carried the provisional 2018 release, and this revision corrects it. 2025 has no final vintage yet, so its figures come from BPS's provisional

release, *Berita Resmi Statistik* No. 2545 of 2 February 2026 [5]. The panel is therefore final for seven of its eight years and provisional for the last. Reverting the earlier years for uniformity would discard the better measurement.

Deleting 2025 leaves the direction of the model comparison in §3 unchanged. The global baseline's cross-validated R^2 falls from 0.1003 to 0.0685, and its margin over the best local specification narrows from 0.0084 to 0.0023 without reversing. The pest-rate coefficient keeps its sign and significance ($+0.518 \rightarrow +0.568$, Driscoll–Kraay $p < 0.001$). No conclusion depends on 2025, so a revision of the 2025 figures cannot overturn one.

2.1.2. Province-frame harmonization

Four provinces (Papua Selatan, Papua Tengah, Papua Pegunungan and Papua Barat Daya) were separated from Papua and Papua Barat in 2022, taking the national count from 34 to 38. The panel aggregates published figures for 2022 onward back onto the pre-split frame. Papua Tengah, Papua Pegunungan and Papua Selatan return to Papua, and Papua Barat Daya returns to Papua Barat. No other province or boundary changes between 2018 and 2025.

The aggregation rule depends on the variable. Production, harvested area and pest-attack area are summed across constituent units, with missing values treated as zero. Rainfall and temperature are averaged in two stages. Station values are averaged within each constituent province, and those provincial means are then averaged without weights. Stations with missing values are excluded from the denominator. Papua's 2025 rainfall is 3916.67 mm under the two-stage mean against 3478.10 mm under a pooled mean.

BPS uses the 38-province frame for production and harvested area from 2022, although no values are reported for the four new provinces that year. The pest-attack series uses it from 2023, and the BMKG station tables only in 2025. The supplementary file gives the unit-level crosswalk, the rule for each variable and the years in which each rule applies. The replication script rebuilds every harmonized cell from the published sources and reconciles all 54 of 54 cells with the analysis panel.

2.1.3. Climate variables and their design change

Rainfall and temperature come from BPS, *Statistik Indonesia*, Table 1.2.1 (*Pengamatan Unsur Iklim di Stasiun Badan Meteorologi, Klimatologi, dan Geofisika Menurut Provinsi*), whose design changes within the study period. For reference years 2018–2022 the table reports one designated climatology station per province. From 2023 it reports every station in the province, following Peraturan Badan Meteorologi, Klimatologi dan Geofisika Nomor 1 Tahun 2022, as the table's own note records [21]. The panel reproduces the table, so a province's annual value is one station's for 2018–2022 and the unweighted mean over its stations from 2023. Papua and Papua Barat in 2025 take the two-stage rule of §2.1.2. The series is therefore not a homogeneous climate record, and no trend in Indonesian climate is inferred from it.

Three limitations follow, and the paper reports them rather than correcting them. First, the unweighted second stage gives each constituent province equal weight whatever its cultivated extent. Highland Papua Pegunungan holds 0.1% of Papua's 2025 harvested area and one quarter of the weight, which places Papua's 2025 temperature 2.14 °C below the area-weighted value from the same stations, more than two panel standard deviations. The area-weighted second stage was estimated and not adopted (§3.10). Second, the change of table design shifts the level at the 2022/2023 boundary, by a median province change of −866 mm and −0.25 °C, in the same direction in 27 and 22 of 34 provinces. The year fixed effects absorb the common part of that shift but not the province-specific part. Letting the weather coefficients differ across the boundary does not improve the model under province-clustered inference ($\chi^2 = 1.52$ on 2 df, $p = 0.47$), and the boundary is not the most discrepant of the seven contiguous splits, so it is not adjusted. No interpretation rests on movement across it. Third, Papua's harmonized 2025 values reflect the change of provincial frame rather than weather and are not read as climate change.

All three bear on the eastern provinces, where the local estimates are largest and least precise.

2.1.4. Pest-attack area: three vintages, two pest sets, one truncated year

OPT stands for *organisme pengganggu tanaman*, the Indonesian term for crop pests and diseases. The pest-attack (OPT) series comes from three Pusdatin editions of *Statistik Iklim, Organisme Pengganggu Tanaman dan Dampak Perubahan Iklim*: panel years 2018–2020 from the 2017–2020 edition [22], 2021–2022 from the 2020–2023 edition [23], and 2023–2025 from the 2023–2025 edition, released on 31 October 2025 [24]. Five workbook cells were transcription slips against the printed source. They were corrected before estimation, and the replication package records each one.

1. *Revision policy.* Overlap years disagree slightly between editions, and the later edition is used.
2. *The variable is not the same sum throughout.* It is a six-pest sum for 2018–2022 and a seven-pest sum for 2023–2025, because the 2023–2025 edition is the first to tabulate *kerdil rumput*, the rice grassy stunt virus disease (0.04–0.20% of the national total).
3. *2025 covers ten months.* The edition reports data through 24 October 2025, so 2025 is not a full calendar year.
4. *The rate is sensitive to provinces with little harvested area.* OPT is pest-affected area divided by harvested area, and DKI Jakarta has an average harvested area of 572.75 ha. The province is kept and disclosed rather than excluded post hoc, and the results report a 33-province sensitivity beside the 34-province fit.

2.1.5. Data availability at higher frequency

Pest-attack data are published annually. BPS reports monthly harvested area and production by province throughout the study period. The panel therefore uses an annual frequency to match the availability of the pest-attack series.

2.1.6. Outcome specification

The main outcome is log yield, production per harvested hectare, with harvested area as a scale control and pest-attack area entered as a rate. The production model is kept as a documented baseline only, on specification evidence. On production the model estimated the accounting identity $P = A \times Y$. The log–log within-province area elasticity was indistinguishable from 1 under Driscoll–Kraay errors, and harvested area alone gave a near-perfect *global* pooled fit. The evidence standard for this switch was declared before any fit, and the replication package holds the declaration.

2.1.7. Multicollinearity

No conclusion here depends on a variance-inflation threshold. §3 reports the measured condition numbers and variance-inflation factors with the fits. They are small enough that no rule of thumb matters, and O'Brien argues that such rules should not decide anything [25].

2.2. The global panel model

The global model has province and year fixed effects:

$$\log y_{it} \;=\; \mathbf{x}_{it}'\boldsymbol{\beta} + \alpha_i + \gamma_t + \varepsilon_{it}. \tag{1}$$

Here y_{it} is the yield of province i in year t in tonnes per hectare, $\mathbf{x}_{it}$ holds log harvested area, the OPT rate, rainfall and temperature, α_i and γ_t are the province and year effects, and ε_{it} is the error. The within transformation estimates Eq. (1) [26]. Computations use R 4.6.1, `plm` 2.6-7 [27] for the global model and `GWmodel` 2.4-1 [28] for the local fits.

The earlier version used province effects only and never tested for year effects. They are jointly significant, so the one-way model is rejected, and they absorb the panel's cross-sectional dependence. All standard errors are Driscoll–Kraay [29]. §3 reports the diagnostics for heteroskedasticity (§2.3.5), serial correlation and cross-sectional dependence.

2.3. Five propositions

This section states five propositions. Each holds for a balanced panel of n units observed in T periods, in which all T rows of a unit share one pair of coordinates. Each proposition comes with a remedy. Appendix A of the supplementary file holds the derivations, the corollaries of P5 and the check of the kernel formulas. P1, P3 and P5 are closed-form results about the estimator and need no data. P2 is an identification result, and P4 is a construction. All five were found in the rice panel and are shown on it, but none of them depends on it. Together, the propositions show that geographically weighted regression, applied unchanged to panel data, fails at five steps. These are the diagnostic that motivates a local model (P5), the bandwidth that sets each local neighbourhood (P1), the validation that scores the model (P2), the search that selects the bandwidth (P3) and the criterion that compares the local model with a global one (P4).

2.3.1. Adaptive bandwidth on a stacked panel (P1)

> **Proposition 1.** *Stack a balanced $n \times T$ panel into nT rows, so that the T rows of each unit share the same coordinates. Fit a geographically weighted model with a kernel of compact support and an adaptive bandwidth of k nearest rows, with $T < k \leq nT$. Let $m = \lfloor (k-1)/T \rfloor$. Then:*
> *(a) exactly Tm rows receive nonzero weight. They are all the rows of the m nearest units, and the focal unit is one of them, at distance zero;*
> *(b) every k from $Tm+1$ to $T(m+1)$ gives the same weights and so the same fitted model. The selection criterion is therefore a step function of k with at most n steps;*
> *(c) a bandwidth reported as a number of rows is not identified. It names a range of T values, not a single value. A golden-section search assumes a criterion with one minimum, and it offers no guarantee on a step function.*

Parts (a) and (b) apply to kernels of compact support, such as the bisquare and the tricube. The Gaussian kernel in the software used here is not truncated. It gives every row nonzero weight at every bandwidth, so it has no neighbourhood count to misstate. For the Gaussian, the problems are the boundary optimum of §2.3.3 and the search floor described next.

The derivation also revealed two errors in the earlier version of this study. First, its bandwidth search used a routine whose lower limit for adaptive bandwidths is a constant fixed in the package code. The reported Gaussian bandwidth was that lower limit, not an estimate. The golden-section search evaluated 9 of the 253 admissible values of k and stopped when two of them tied on one step. Second, the earlier version left the distance metric to a software default. The corrected analysis uses great-circle distance throughout and sets it in code (`gw.dist(..., longlat = TRUE)`). The earlier version's §3.6 claimed that the results do not depend on the metric, and that sentence is removed.

The earlier version's Eqs. (13)–(15) printed the three kernels. Its bisquare and tricube match the software, but its Gaussian was truncated at the bandwidth, where the software's Gaussian is not. The corrected formula changes no estimate.

The remedy is to define, search and report the bandwidth as a number of units, which here are provinces, not rows. The estimator still runs on the stacked panel. For each unit count m, the search evaluates one value of k in the middle of its step, $k = Tm + T/2$. The criterion then becomes an ordinary function of n points instead of a step function over nT rows. Every bandwidth in this paper is a province count. A province-year count appears only where a software argument is described, and then in parentheses. Where a distance matrix is built from the unique

province coordinates, the text says so.

2.3.2. Intercept identification under grouped folds (P2)

> **Proposition 2.** *Let a model be estimated on unit-demeaned data, and let cross-validation hold out a whole unit. Then the training data do not identify the held-out unit's level α_i, so a grouped fold scores no model until an auxiliary estimator supplies that level. Setting $\alpha_i = 0$ is defined but uninformative. It scores the ability to predict unit* levels*, which is the nuisance the fixed effects were introduced to absorb, so a local and a global demeaned model fail on it for the same reason and the comparison says nothing about slopes.*

Two alternatives were considered and rejected. Modelling α_i from between-unit covariates, a Mundlak device, adds a component that neither competing model contains, so it would validate a third model. Period effects are predicted rather than plugged in, because they stay identified in the training set. A unit holdout removes no year. The earlier version of this study validated by leave-one-row-out on the stacked panel, a criterion under which a third of the local weight at a held-out row belongs to that row's own unit, and every bandwidth it reported was selected by minimising it.

The remedy is to score observation (i, t) against a plug-in level computed from unit i's other $T - 1$ periods, with the target period excluded. No value of the scored observation enters the slopes, the level or the period effects. The criterion then validates the transferability of slope coefficients and not of unit levels, and that limitation is printed rather than left implicit. This criterion replaces the earlier version's Eq. (16). The design built on that remedy is set out below.

- The primary scheme is leave-one-unit-out with $n = 34$ deterministic folds, each training on 33 provinces and 264 rows. Determinism removes randomness rather than persisting a seed.
- Spatial blocking is not primary, and the reason is evidence. Residual spatial autocorrelation is strong in the pooled residuals and absent once province effects are in, and at small bandwidths removing a whole island removes a focal point's entire neighbourhood. An island-blocked scheme, with blocks of 10, 6, 3, 5, 6 and 4 provinces, is reported as a stress test.
- A holdout of 2025 against 2018–2024 is secondary and decides nothing, because the 2025 figures are a provisional vintage and 34 observations are not a basis for selecting a model.
- Nesting is mandatory. The local model has a tuned hyperparameter and the global baseline has none, so selecting that hyperparameter on the folds that produce the reported score biases the comparison in one direction. Each outer fold selects its own bandwidth from its own training units, split into 5 grouped inner folds (seed 20260808, the only seed in the design).
- The reported bandwidth is not the validation bandwidth. The reported one is the argmin of the full-data CV error, which is the argmax of the CV R^2 curve, while the out-of-sample numbers come from the nested procedure in which each outer fold chose its own. The spread of those selected bandwidths across outer folds is reported as a stability diagnostic.
- The criterion is CV $R^2 = 1 - \text{PRESS}/\text{TSS}$ on log yield, aggregated over all 272 predictions, with TSS taken on the same plug-in-demeaned target, so the null model is the unit mean and nothing else. Overfitting is reported as in-sample R^2 minus CV R^2 on that target and null. The back-transformed figures in tonnes per hectare are for readability and are never the criterion.
- Five leakage assertions are enforced, and a failure of any one halts the pipeline: a held-out unit never appears in its own training set; a scored period is never among its own level plug-in periods; the scored observations equal the panel exactly, one to one; the inner folds partition each outer training set; and the design leaves no same-unit row at distance zero, against $T - 1 = 7$ under leave-one-row-out. Fold files are persisted with per-file MD5

checksums in a manifest.

2.3.3. Grid saturation and the fold-dependent ceiling (P3)

Proposition 3. *Let an adaptive bandwidth be indexed by a unit count m and implemented as $k = Tm + c$ rows with $0 < c < T$, capped at the number of rows available, and let an inner training set hold M units and so MT rows. Then:*
(a) every $m \geq M$ asks for more rows than exist and returns the same weights, so a grid in absolute m admits at most $M - 1$ distinct weight vectors and saturates below its own upper limit, whatever that limit is;
(b) on the plateau the criterion is flat, so the selected bandwidth is not identified and the tie rule, an implementation detail, decides it;
(c) the binding constraint is the inner training size, not the upper limit. Raising the limit alone changes nothing while the grid is saturated, so a single ceiling effect is not a quantity this design can report.

The parameterisation carries one further disclosure. Because $k = Tm + c$ cuts inside the tied rows of the $(m + 1)$-th unit, an adaptive kernel of compact support at $q = 1$ gives the farthest unit zero weight. Then $q = 1$ and $q = (M - 1)/M$ are the same fit, and the search ranges over $M - 1$ distinct compact neighbourhoods. The text says so wherever the grid is described. The neighbourhood sizes the earlier version of this study reported for these data, one per kernel, came from a grid that saturated below its own upper limit, and this paper discloses that rather than correcting it silently.

The remedy is to search the bandwidth as a fraction $q \in (0, 1]$ of the training units, applied at each nesting level as $m = \text{round}(q \cdot M)$. Its upper limit is then the whole training set at every level and depends on no fold-design constant, so one value of the hyperparameter denotes the same relative neighbourhood inside and outside the fold. This is a reparameterisation of the same estimator. Fits that use no inner folds, which are the full-data bandwidth curve, the permutation test, the local inference and the information criteria, are unchanged by it, and so are the persisted inner-fold files of the validation design. The saturation point of the inner training set is stated wherever a nested search is described in this paper. §2.5 points to the two prohibitions this proposition imposes.

2.3.4. One criterion for a local and a global panel model (P4)

Proposition 4. *Let P be the two-way within projection, which demeans by unit and by period, and let S be the smoother matrix of a geographically weighted fit computed on the demeaned response. Define the composite smoother*

$$S^* \;=\; P + S(I - P). \tag{2}$$

Then $I - S^* = (I - S)(I - P)$, so the fitted residuals are unchanged and only the complexity charge moves. For the global two-way fixed-effects model, where $S = 0$ on the demeaned residual space and there are p slopes, the composite effective parameter count is $\text{tr}(P) + p = (n + T - 1) + p$, the shared level counted once. On this panel that is $34 + 8 - 1 + 4 = 45.000$, measured on the realised design rather than imposed.

Two declarations go with the composite smoother, both required by the rule that no criterion is reported without its scope. First, the complexity figure printed beside an information criterion is not the figure that criterion is penalised by. `GWmodel` penalises AIC, AICc [30] and BIC by $\text{tr}(S)$ while printing beside them the effective number of parameters of the earlier version's

Eq. (19),

$$\mathrm{ENP} \;=\; 2\,\mathrm{tr}(S) - \mathrm{tr}(S^{\top}S), \tag{3}$$

and the two differ by 12 to 24 percent on these fits. This paper re-implements those formulas, checks them against the package's saved output for all three kernels, and uses $\mathrm{tr}(S^{*})$ as the complexity penalty at composite scope. ENP is reported as Eq. (3) at the same scope, and the text says which figure is which. Second, two statistics share the name R^2 and two more share the name CV R^2. A *within* R^2 takes zero as its null and a *resubstitution* R^2 takes the unit plug-in mean, and only the second is directly comparable with the cross-validated statistic. They differ by a factor of 3.5 to 5 on the same fit through the null alone. A *grouped* CV R^2 selects the hyperparameter across all folds and a *nested* one selects it inside each fold. Every reported quantity carries its scope and its null as a label.

Without Eq. (2) the comparison the review asks for cannot be performed, and the earlier version of this study put a *within* residual sum of squares beside a *levels* residual sum of squares, a mismatch worth more criterion points than the models differ by. The remedy is to place every model on one response vector and charge it the composite smoother. The primary criterion is then the nested leave-one-unit-out CV R^2 on log yield, with the bandwidth and the kernel both selected inside the fold. It answers the paper's transfer claim, is the only criterion that can be made leakage-safe, and was declared before any fit. It also charges the search, where $\mathrm{tr}(S)$ prices the flexibility of a single fit. The information criteria, by contrast, assume an independent Gaussian likelihood over nT rows when the data are n units observed in T periods. AICc is the primary criterion nowhere in this paper. The earlier version's nomination of it cannot be called blind, because the comparison table existed when the nomination was made, and it did not separate the models either way. Every comparison quotes the margin and not only the rank.

2.3.5. The Breusch–Pagan statistic on a balanced panel (P5)

Proposition 5. *Let the Breusch–Pagan statistic be computed as printed in the geographically weighted panel literature,* $\mathrm{BP} = \tfrac{1}{2}\, f'Z(Z'Z)^{-1}Z'f$ *with* $f_{it} = e_{it}^2/\hat{\sigma}^2 - 1$*, and let the error variance be estimated as* RSS/n*, dividing by the number of* units*, while* f *runs over all* nT *rows. Then*

$$f_{\mathrm{coded}} \;=\; f_{\mathrm{defined}}/T - (T-1)/T, \tag{4}$$

so Eq. (4) compresses every scaled squared residual onto the constant $-(T-1)/T$*. The first column of* Z *is an intercept, so the projection reproduces that constant exactly and*

$$\mathrm{BP} \;\approx\; \tfrac{1}{2}\, nT\,((T-1)/T)^2\,, \tag{5}$$

a sample-size constant that no residual influences. Eq. (5) holds to within a tenth of a percent on this panel and is identical across kernels to machine precision.

The earlier version's null hypothesis indexed the error variances by p, the number of regressors, whereas the required hypothesis indexes them by observation. The remedy is to replace the earlier version's Eq. (7), not to relabel it. The replacement is the studentised form of Koenker [31],

$$\mathrm{BP}^{*} \;=\; nT \cdot R^2_{\mathrm{aux}}, \tag{6}$$

where R^2_{aux} is the coefficient of determination of the auxiliary regression of the scaled squared residuals $e_{it}^2/(\mathrm{RSS}/nT)$ on a constant and the regressors, referred to χ^2 on p degrees of freedom and applied to the residuals of the two-way fixed-effects model once the province and year effects have been removed. Under its null the error variance is constant across province-year observations. The reason for studentising is measured rather than stylistic. The corrected model's residuals carry heavy excess kurtosis, and the original statistic [32] exists to be studentised because it

mistakes excess kurtosis for heteroskedasticity. The earlier version's Eq. (8) is deleted, and the heading of the earlier version's §3.5 becomes *Heteroskedasticity.*

Whatever verdict the corrected statistic returns, and §3 reports it, it bears only on the choice of standard errors, and it does not earn the word *spatial.* A heteroskedasticity test is a statement about the error term, while spatially varying coefficients are a statement about the mean function, which a spatial-error test cannot answer either. Two sentences of the earlier version are deleted in consequence, one from its Results and one from its Discussion.

2.4. The corrected procedure, assembled

The propositions are constructive, so the corrected procedure applies their remedies in sequence. It covers bandwidth specification, grouped validation, model comparison and diagnostic testing. Appendix A of the supplementary file lists its eleven steps, each linked to the proposition it applies.

2.5. Standing prohibitions carried by this section

Appendix A of the supplementary file lists the seven prohibitions that bind the rest of the paper. They constrain both the analytical choices and the interpretation of the results. For example, a province is not excluded after its influence is observed, and a heteroskedasticity test is not used to justify geographically varying coefficients.

3. Results and Discussion

The results follow the corrected procedure. §3.1 to §3.3 describe the panel and test the global model. §3.4 to §3.7 select the bandwidth, compare GWPR with the global model and test whether the geography carries information. §3.8 to §3.10 report the local coefficients, the pest term and the sensitivity checks. §3.11 to §3.14 discuss the five propositions, the null result, the pest finding and the limitations of the study.

3.1. The panel

Every criterion below is printed with its scope and its null (P4). The panel has 34 provinces $\times$ 8 years (2018–2025), 272 province-years on the pre-split frame (descriptive statistics in the supplementary file). Mean yield is 4.55 t/ha (standard deviation 0.87, range 2.37 to 7.28). Harvested area runs from 113 ha to 1,841,346 ha, four orders of magnitude. The OPT rate has a mean of 0.045, a median of 0.029 and a maximum of 0.339. Papua's 2025 temperature fall is a frame change and not weather, because the same thirteen stations are 0.06 °C *warmer* than in 2024, so it is reported and never interpreted.

3.2. The global baseline and its specification tests

Province effects are required (Chow [33] $F(33, 234) = 41.74$, $p = 5.1 \times 10^{-80}$), and so are year effects, $F(7, 227) = 7.858$, $p = 1.6 \times 10^{-8}$ (Table 1), a test the earlier version never ran. Random effects are rejected (Hausman [34] 12.96, df 4, $p = 0.0115$; robust auxiliary 12.72, $p = 0.0127$). On the subpanel 2019–2025 the rejection disappears (5.107, $p = 0.2765$). That sensitivity is to 2018 as a year and not to its vintage, since it appears under both the provisional and the corrected 2018 column, so the fixed-effects choice costs efficiency and not validity.

Table 2 reports the two-way fixed-effects model on log yield with Driscoll–Kraay standard errors. The OPT-rate coefficient is 0.5180 (DK SE 0.1412, $p = 0.0003$), and the coefficients on rainfall, temperature and log harvested area are indistinguishable from zero ($p = 0.578$, 0.786 and 0.983). Driscoll–Kraay standard errors are chosen for the dependence structure, not the variance structure. The year effects absorb the one-way model's cross-sectional dependence (Pesaran CD 12.20, then CD -0.291, $p = 0.771$). Mild serial correlation remains under two-way effects (Breusch–Godfrey(2) 6.777, $p = 0.034$; Wooldridge AR(1) $F = 4.396$, $p = 0.037$), and

Driscoll–Kraay standard errors are robust to it. The corrected heteroskedasticity statistic does not reject (§3.3).

On log *production*, the province-effects model with pest-attack area in totals gives an area elasticity of 0.9426 (SE 0.0286). It differs from 1 at the 5% level under naive errors ($t = -2.007$ against 1; two-sided $p = 0.0459$, df 234), but not under Driscoll–Kraay errors, where the same model's log-yield area coefficient, −0.0574, has $p = 0.135$. Harvested area alone gives a *global pooled* R^2 of 0.99518, and the within-R^2 of the four-regressor production model falls from 0.864 to 0.009 when area is dropped. The near-perfect fit therefore predates any local weighting. It is the accounting identity, not GWPR flexibility and not a small bandwidth (§2.1.6). No statistic of the production model is placed beside a log-yield statistic.

Table 1: Specification and diagnostic tests, canonical panel

test	statistic	df	p value
Province effects (Chow F), pooled vs one-way within	41.74	33,234	5.09×10^{-80}
Year effects (F), not tested in the earlier version	7.858	7,227	1.59×10^{-8}
Hausman, FE vs RE (classical)	12.96	4	0.01148
Hausman, FE vs RE (robust auxiliary)	12.72	4	0.01272
Serial correlation, Breusch–Godfrey(2), two-way FE residuals	6.777	2	0.03375
Serial correlation, Wooldridge AR(1) for FE	4.396	1,236	0.03709
Cross-sectional dependence, Pesaran CD, one-way FE	12.20	–	3.17×10^{-34}
Cross-sectional dependence, Pesaran CD, two-way FE	−0.2913	–	0.7708

Note: Read together: province effects and year effects are both required; RE is rejected; the two-way specification removes the cross-sectional dependence the one-way model shows; mild serial correlation remains, and Driscoll–Kraay standard errors are robust to it.

Table 2: Global two-way fixed-effects model on log yield (272 province-years, 34 provinces, 8 years)

term	estimate	SE Driscoll–Kraay	t DK	p DK	SE naive	p naive	SE clustered	p clustered
OPT rate	0.5180	0.1412	3.668	0.0003043	0.1582	0.001221	0.2218	0.02037
rainfall (mm)	3.88×10^{-6}	6.96×10^{-6}	0.5576	0.5777	9.43×10^{-6}	0.6810	1.05×10^{-5}	0.7111
temperature (°C)	−0.0008259	0.003037	−0.2720	0.7859	0.007222	0.9090	0.005744	0.8858
log harvested area	−0.0005825	0.02662	−0.02188	0.9826	0.02992	0.9845	0.04896	0.9905

Note: Driscoll–Kraay (DK) standard errors are the reported ones. Naive and province-clustered columns are printed so the reader can see that the OPT-rate coefficient is not a fragile t-statistic (Section 3.9: it rests on a single leverage point).

3.3. Heteroskedasticity and what it cannot show

On the corrected design (log yield, two-way within residuals), the studentised Koenker statistic Eq. (6) is 7.318 on 4 df, $p = 0.120$, so the corrected panel shows no detectable heteroskedasticity (supplementary file). The residuals carry heavy excess kurtosis (Jarque–Bera 286.7, kurtosis 8.00), which the unstudentised statistic mistakes for non-constant variance, and on the same residuals it gives 25.84 ($p = 3.4 \times 10^{-5}$). On the specification the earlier version's routine tested, pooled OLS in raw levels, the statistic is 62.08 unstudentised and 38.90 studentised. A Breusch–Pagan routine applied to a panel-model object returns that pooled statistic whatever the panel specification is, so the 38.90 the earlier version printed never had a province effect removed.

The heteroskedasticity table in the supplementary file has one row per specification, not one per kernel. Under the earlier version's printed Eq. (7), the statistic cannot depend on the kernel (P5), and the kernel-to-kernel variation the earlier version reported came from a second, undisclosed statistic summed onto the first. Two of its sentences are therefore deleted, "Consequently, GWPR was applied to examine relationships that may vary across provinces" from its Results and "The spatial heterogeneity test supported the use of a local modelling approach" from its Discussion. No heteroskedasticity test and no spatial-error test can support a claim about coefficient variation.

3.4. The bandwidth curve

Searched in provinces over $m = 1, \ldots, 33$ with great-circle distance, the grouped CV R^2 curve rises almost monotonically for every kernel, and its full-data argmax is $m = 32$ of 34 for all three (Fig. 1). The earlier version's adaptive bandwidth of 26 rows is, on a stacked 34×8 panel, a neighbourhood of three provinces, and by P1(b) one of at most 34 distinguishable choices, not one of 253. The nested selection demonstrates P3, and the supplementary file lists what every fold selected. Each outer fold trains on 33 provinces split into five inner folds, so the inner training set holds 26 provinces and the absolute-m grid saturates there. The selected bandwidth is at the top of the grid in 30 of 34 folds (bisquare), 27 of 34 (tricube) and 18 of 34 (Gaussian). The bandwidth is therefore *also* reported as a fraction q of the training provinces, whose ceiling depends on no fold constant (Table 3, last three rows).

Table 3: Model comparison on one response vector, all models charged the composite smoother (Eq. (2))

model	global two-way FE	GWPR Gaussian	GWPR bisquare	GWPR tricube
bandwidth m (provinces)	–	32	32	32
ENP (Eq. (3)) composite	45.0000	46.1284	49.7289	48.7579
AICc composite	−664.755	−665.154	−663.494	−663.536
ΔAICc	0.398824	0	1.66024	1.61765
Akaike weight	0.303346	0.370289	0.161445	0.164920
BIC composite	−795.712	−794.477	−786.203	−787.558
R^2 resubstitution	0.241481	0.247520	0.263243	0.259293
CV R^2 nested LOPO	0.100321	0.0919677	0.0301611	0.0300973
RMSE log in sample	0.0581134	0.0578816	0.0572737	0.0574270
RMSE log CV	0.0723317	0.0726667	0.0750991	0.0751016
CV R^2 nested fractional q	–	0.0956902	0.0855849	0.0889033
selection optimism absolute grid	–	0.00597486	0.0547967	0.0580530
selection optimism fractional q	–	0.00225230	−0.000627074	−0.000753033

Note: Scope and null, stated for every row: the resubstitution R^2 takes the province plug-in mean as its null and is the only in-sample R^2 directly comparable with the cross-validated rows. The nested LOPO CV R^2 is the primary criterion, with the bandwidth selected inside each of 34 leave-one-province-out folds. RMSE is on the log scale only. AICc, BIC and the composite ENP are the implementation's own formulas verified against its saved output. The criteria are penalised by $\mathrm{tr}(S^*)$, while ENP is Eq. (3) at composite scope.

3.5. The decision rule

The rule was fixed before any fit: *if the GWPR does not beat the global panel model out of sample under grouped cross-validation, the global model is reported as preferred.* Table 4 reports nested leave-one-province-out CV R^2 on log yield against the province plug-in-mean null. The global two-way fixed-effects model attains 0.1003, the GWPR 0.0920 (Gaussian, $m = 32$), 0.0302 (bisquare), 0.0301 (tricube), and 0.0643 with kernel and bandwidth selected in-fold.

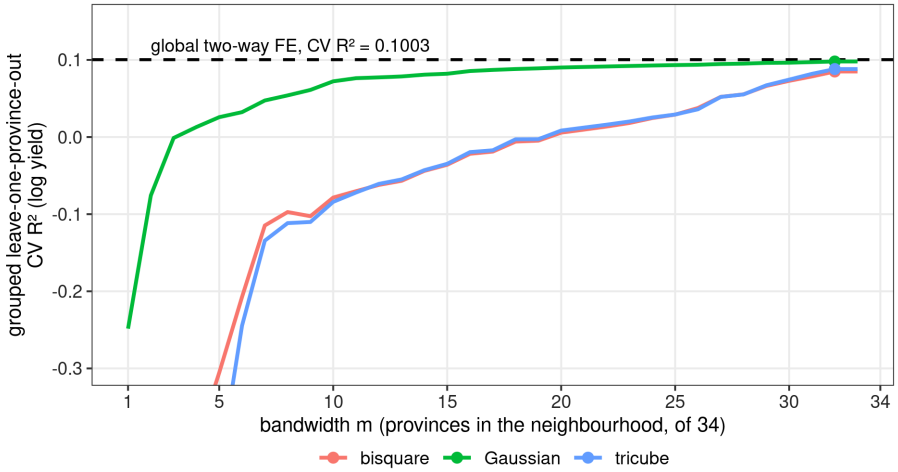


Fig. 1: Every kernel wants almost the whole country in every neighbourhood. Great-circle distance. The earlier version reported a neighbourhood of three provinces. Curve values are grouped CV; the values used to apply the decision rule are nested (Table 4). The baseline has no tuned hyperparameter, so its grouped and nested values coincide.

The GWPR loses in all four rows. The rule is applied to the last row, because the kernel is a tuned hyperparameter too, and selecting it outside the fold would tilt the comparison (P2). Under the fold-independent fractional grid the four rows are 0.0957, 0.0856, 0.0889 and 0.0930, all still below 0.1003 (Table 3).

A paired province-clustered bootstrap (2000 replicates, seed 20260808) puts the Gaussian at -0.0084 $[-0.0236, +0.0038]$ and the in-fold selector at -0.0360 $[-0.1022, +0.0036]$, intervals that straddle zero. The two compact kernels lose *significantly*, bisquare -0.0702 $[-0.1399, -0.0115]$ ($p = 0.009$) and tricube -0.0702 $[-0.1401, -0.0099]$ ($p = 0.013$). Only 13, 12, 12 and 13 of the 34 provinces improve under the four rows. The data either do not separate the two models or separate them against the GWPR. In-sample minus nested CV R^2 is 0.141 for the global model against 0.156 (Gaussian), 0.233 (bisquare) and 0.229 (tricube), so the local models do not keep the fit they buy in sample.

Table 4: The pre-declared decision rule: nested leave-one-province-out CV R^2 on log yield

model	full data argmin m	CV R^2 at argmin grouped	CV R^2 nested	baseline two-way FE	Δ vs baseline	beats baseline
global two-way FE (baseline)	–	–	0.100321	0.100321	0	–
GWPR Gaussian	32	0.0979425	0.0919677	0.100321	−0.00835322	no
GWPR bisquare	32	0.0849578	0.0301611	0.100321	−0.0701598	no
GWPR tricube	32	0.0881503	0.0300973	0.100321	−0.0702235	no
GWPR, kernel and bandwidth selected in-fold	–	–	0.0642787	0.100321	−0.0360422	no

Note: The rule was fixed before any fit: if the GWPR does not beat the global panel model out of sample under grouped cross-validation, the global model is reported as preferred. The GWPR loses in all four rows. The rule is applied to the last row, where the kernel is a tuned hyperparameter too.

3.6. One criterion for both models

All four models are placed on one response vector and charged the composite smoother $S^* = P + S(I - P)$ of Eq. (2), which leaves residuals untouched and moves only the complexity charge (Table 3). The global model's composite ENP is exactly $45.000 = \mathrm{tr}(P) + p = 41 + 4$, measured rather than imposed. Without it the reconciliation cannot be performed, because the earlier version's Table 6 has no FEM row, the earlier version's Table 7 prints no AIC or AICc, and the

two sides are a *within* and a *levels* residual sum of squares.

No criterion has a decisive margin. The AICc spread over the four models is 1.660 points, inside its own $\Delta\text{AICc} < 2$ indifference band, with a leading Akaike weight of 0.370 against 0.25 for indifference. BIC prefers the global model by 1.235 points, also inside its band. Over $m = 1, \ldots, 33$ (Fig. 2), no BIC-selected bandwidth of any kernel beats the global model, while 11 of 99 AICc points fall below it, all Gaussian. AICc is nominated as primary nowhere in this paper.

AIC, AICc *and* BIC use the `GWmodel` formulas with $\text{tr}(S^*)$ in place of $\text{tr}(S)$, while the effective-parameter column is Eq. (3) at composite scope, and the two are 0.5 to 1.6 parameters apart. The resubstitution R^2 takes the province plug-in mean as its null and is the only in-sample figure directly comparable with the cross-validated one. The local ENP at the selected bandwidths is 6.47 (Gaussian), 11.21 (bisquare) and 9.87 (tricube), against 92.3 in the earlier version, which came from the stacked-coordinate configuration with an untransformed response. Selection optimism, grouped minus nested CV R^2, is 0.0060, 0.0548 and 0.0581 on the absolute-m grid and 0.0023, -0.0006 and -0.0008 on the fractional grid, so most of the compact kernels' apparent optimism came from the grid.

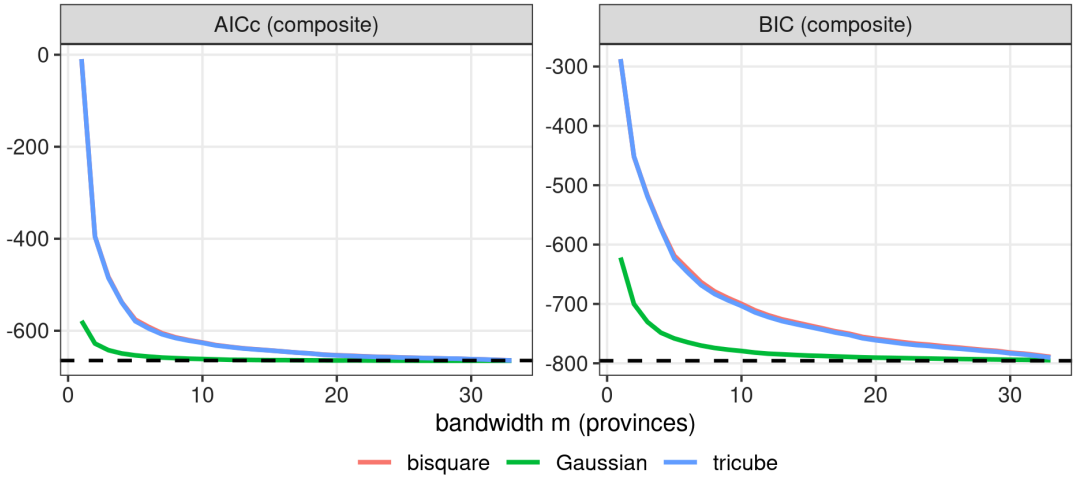


Fig. 2: Information criteria over the whole bandwidth grid, all models charged the composite smoother. Dashed line: the global two-way fixed-effects model, charged the same composite smoother. No BIC-selected bandwidth of any kernel beats it.

3.7. Does the geography carry information?

The goodness-of-fit F-test of the earlier version's Eqs. (21) and (22) is withdrawn. Its replacement, which needs no degrees of freedom, relabels the 34 provinces on their own coordinates in 999 province-block permutations (seed 20260808) and refits at the selected bandwidth (Table 5, Panel A). The observed residual sum of squares yields $p = 0.390$ (Gaussian), 0.535 (bisquare) and 0.483 (tricube), so the real map of Indonesia fits no better than a random one (supplementary file). Leung, Mei and Zhang's F_1 [35] agrees ($p \approx 0.46$–0.48), and so does the correctly formed differenced nested F ($p = 0.489$ to 0.510 under the parameter count of Eq. (3)). Out of sample (Table 5, Panel B), permuted maps at fixed bandwidth give $p = 0.723$ (bisquare), 0.713 (tricube) and 0.701 (Gaussian), and about a third of random maps beat the global model. An out-of-sample gain from a near-global compact kernel is therefore a shrinkage effect that any map can buy, and this paper never reads such a gain as evidence of geographically varying coefficients.

3.8. Local coefficients and intervals, without significance stars

The inference family is 34 provinces $\times$ 4 slopes = 136. The da Silva–Fotheringham critical values [36], recomputed on the corrected fit, are 2.080, 2.301 and 2.251 (Gaussian, bisquare, tricube),

Table 5: Province-block permutation test of the geography (999 permutations, seed 20260808), in sample and out of sample, replacing the withdrawn goodness-of-fit F-test

Panel A. In sample: residual sum of squares at the selected bandwidth

kernel	bandwidth m	permutations	RSS observed	RSS null median	RSS global model	p one-sided
Gaussian	32	999	0.91128	0.91221	0.91859	0.39000
bisquare	32	999	0.89223	0.89125	0.91859	0.53500
tricube	32	999	0.89702	0.89755	0.91859	0.48300

Panel B. Out of sample: CV R^2 under permuted maps

kernel	bandwidth m	observed CV R^2	null median CV R^2	null q95	share of random maps beating the global model	p one-sided
bisquare	30	0.07287	0.08985	0.1367	0.3373	0.7230
tricube	30	0.07445	0.08975	0.1364	0.3303	0.7130
Gaussian	21	0.09084	0.09735	0.1208	0.3934	0.7010
bisquare	17	−0.01887	0.05463	0.1500	0.2382	0.8390
tricube	17	−0.01715	0.05370	0.1505	0.2352	0.8160

Note: The null relabels the 34 provinces on their own coordinates, so it needs no degrees of freedom. The observed fit buys no smaller residual sum of squares than a random map of Indonesia. About a third of random maps beat the global model, so an out-of-sample gain from a near-global compact kernel is a shrinkage effect any map can buy and is not evidence of geographically varying coefficients.

against 1.973 in the earlier version, a value that appears nowhere among the revised critical values (Table 6). At the CV-selected fit (Gaussian, $m = 32$), under Benjamini–Hochberg at FDR 5% with clustered errors, 0 of 136 local slopes reach significance. Under the compact kernels 24 (bisquare) and 23 (tricube) do, all of them positive OPT-rate slopes spanning only 0.210–0.665 (bisquare) and 0.183–0.667 (tricube) around a global coefficient of 0.518.

Against the global coefficient, 0 of 136 local slopes differ from their global counterpart under any kernel, nominally or under Benjamini–Hochberg, with a largest $|t|$ of 1.562. The supplementary file prints the full 34×4 table of local slopes with coefficient, clustered interval and no significance stars. Fig. 3 draws the pest surface, and the earlier version's significance-grouping table is withdrawn.

Table 6: Local inference: what survives, in every kernel

kernel	family size	n BH vs zero	n BH vs zero that are OPT	n nominal vs global	n BH vs global	max abs t vs global	critical t
bisquare	136	24	24	0	0	1.462	2.301
Gaussian	136	0	0	0	0	0.4214	2.080
tricube	136	23	23	0	0	1.562	2.251

Note: Against zero, under Benjamini–Hochberg (BH) at FDR 5%, with province-clustered standard errors. Against the global coefficient, the comparison a geographically weighted paper actually makes, nothing survives in any kernel, nominally or under BH. The critical t is the da Silva–Fotheringham value.

3.9. The pest term: what it measures, and the one province that carries it

In the ladder of pest specifications (supplementary file), CV R^2 is 0.1003 with OPT as a rate, 0.1183 in totals and 0.1239 with no pest term, the best of all. Two placebos declared in advance do as well (supplementary file), so the advantage of the totals carries no pest information. The rate is kept because the pest variable is the paper's subject, not because it is informative. In PRESS, the rate costs 0.0372 against the no-pest model (supplementary file).

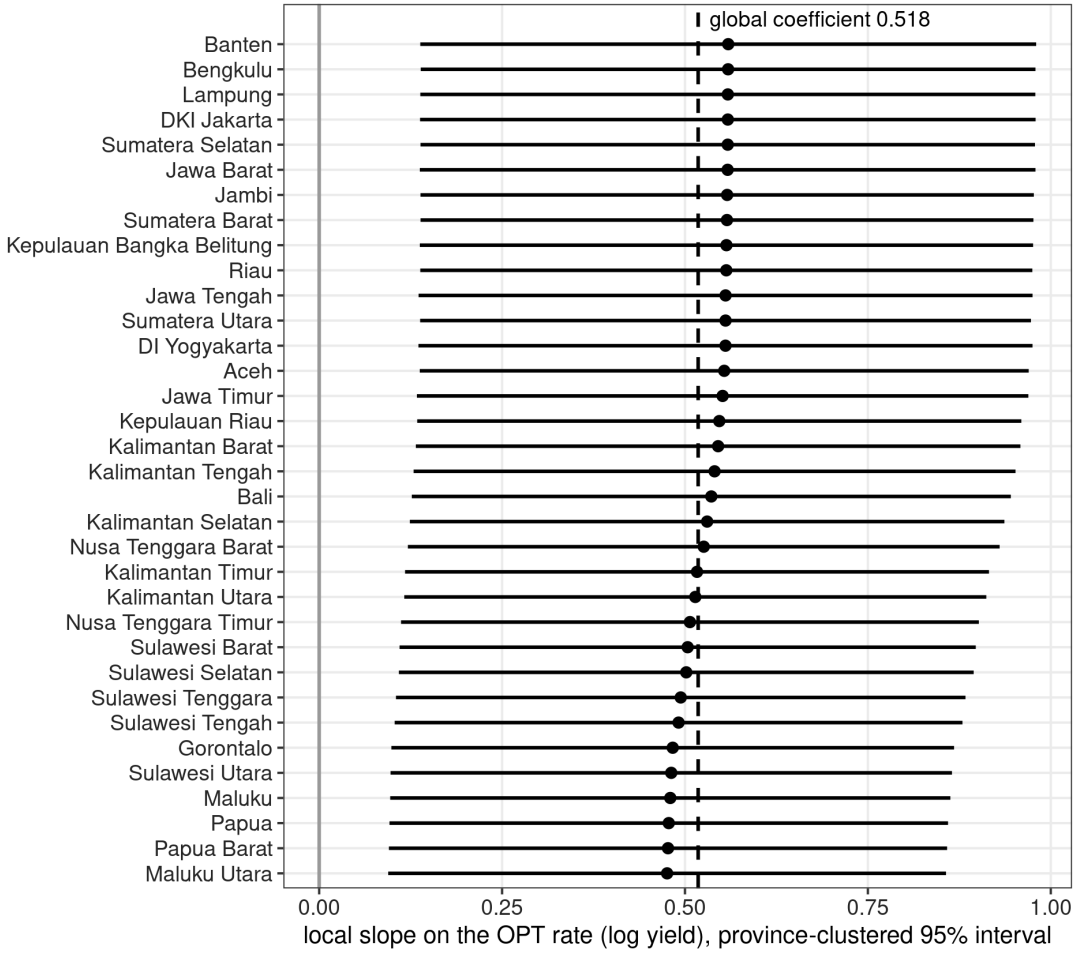


Fig. 3: Local pest slopes at the CV-selected fit: no stars, and none distinguishable from the global coefficient. Gaussian kernel, $m = 32$ of 34 provinces. Intervals are province-clustered; there are no significance stars anywhere in this figure. The whole surface spans a range narrower than one standard error of the global coefficient.

The global coefficient on the OPT rate is +0.518 (Driscoll–Kraay $p = 0.0003$, naive $p = 0.0012$, clustered $p = 0.0204$; Table 2). It holds under every standard error but rests on one leverage point, DKI Jakarta. Before correction the coefficient was −0.053 (Driscoll–Kraay $p = 0.838$). The final 2018 vintage (§2.1.1) alone carries it to +0.367, and the five corrected workbook cells (§2.1.4) alone leave it at +0.054 (supplementary file). In the leave-one-province-out refits (supplementary file), deleting DKI Jakarta moves the coefficient by −0.532, more than its own size, to $\beta = -0.014$ ($p = 0.934$), while the other 33 deletions leave it between +0.494 and +0.567. The final 2018 table gives DKI Jakarta 673 ha harvested, so its 2018 OPT rate is 0.260, the fourth largest of all 272 province-years, in a panel whose median rate is 0.029.

DKI Jakarta is disclosed, not dropped, because deleting a province after finding that it produces the result is the post hoc exclusion that §2.1.4 rules out. With §3.8, the finding is an absence. This panel cannot detect an effect of recorded pest-attack area on rice yield, at any geography, under any specification tried.

3.10. Sensitivities

Island-blocked folds (six blocks) give 0.0771 for the global model, 0.0745 for the Gaussian, 0.0689 for the tricube and 0.0649 for the bisquare, the ranking of §3.5 under a scheme that penalises the local model for geography rather than flexibility. Trained on 2018–2024 and tested on 2025, every model scores below the province-mean null. The 2025 year effect is unidentified in training

and set to zero for both models, so the result decides nothing. The area-weighted variant (§2.1.3) moves the global model from 0.1003 to 0.1007 and leaves the ranking unchanged, with the in-fold GWPR at 0.0862. The largest local condition number is 1.473 and the largest VIF is 1.097.

3.11. What P1–P5 mean for GWPR practice

The five failures have one cause, a cross-sectional tool applied to non-exchangeable rows.

- P1. An adaptive bandwidth of k observations on a stacked panel is $\lfloor (k-1)/T \rfloor$ units, so a reported 26 nearest neighbours is three provinces when $T = 8$. Papers should report the bandwidth in units with T beside it.
- P2. Cross-validation of a unit-demeaned panel GWPR tests slope transfer, never levels. Leave-one-row-out on a panel leaks, because the held-out row's own unit supplies $T - 1$ neighbours at distance zero.
- P3. A grid whose ceiling depends on the fold design searches the validation scheme, not the estimator. A fractional bandwidth avoids this.
- P4. Local and global panel models are routinely compared across residual sums of squares on different scales, which the composite smoother puts on one.
- P5. A statistic that reduces to $\frac{1}{2}\, nT\, ((T-1)/T)^2$ is evidence of nothing. Even computed correctly, a heteroskedasticity test says nothing about the mean function, so it cannot motivate a local model.

3.12. Why a null here is not a null everywhere

This panel gives no support for geographically varying coefficients, for reasons specific to it. The outcome is a ratio with little within-province variation (within-R^2 of the four-regressor model ≈ 0.048). The regressors that vary within a province, weather and a pest rate, show little measured association with yield here. Eight periods and 34 units are few. Pooled Moran's I is $+0.384$ and significant in all eight annual cross-sections, but the fixed effects remove it, and under province effects the statistic is degenerate. A local model is left with variation in slopes, not in levels. A larger or longer panel, or regressors whose within-unit variation differs by region, could reach the opposite verdict under the same procedure. The earlier version's finding does not transfer, because removing the five defects of P1 to P5 removes the result.

3.13. What the pest finding is, and what it is not

The missing pest association concerns recorded pest-attack area, an annual administrative series whose pest set, coverage and denominator change across the panel (§2.1.4), not pests. No coefficient is interpreted causally. The findings do not support conclusions about crop protection policy.

3.14. Limitations

The climate series changes design inside the window under Peraturan BMKG No. 1/2022, so it is not a homogeneous climate record (§2.1.3). No weather slope is reported on a short subpanel. Harvested area and production exist monthly in the KSA frame, so the earlier version's claim that higher-frequency data were unavailable is narrowed to the pest series, which is annual (§2.1.5).

4. Conclusion

Geographically weighted panel regression takes five pieces of machinery from the cross-sectional case, and all five break on a balanced panel with repeated coordinates. This paper states the five as propositions, derives them, repairs them, and reruns the analysis of the earlier version of this study through the repaired procedure. The bandwidth becomes a count of units, on a criterion with at most n steps. The validation gains a plug-in level and validates slope transfer. The grid

stops depending on the folds, one criterion compares the models, and the diagnostic becomes a test that can reject.

On the Indonesian rice panel the repaired procedure prefers the global two-way fixed-effects model, on the criterion declared before any fit and on every variant of it computed since. A random relabelling of the provinces fits as well as the real map, and no local slope differs from its global counterpart. This outcome is reported as the demonstration, not as the contribution. The propositions are what generalise. Any GWPR study on panel data can check its bandwidth semantics, fold design, grid ceiling, criterion scope and motivating diagnostic against them, using the arithmetic in Section 2 and no new data.

CRediT Authorship Contribution Statement

The authors contributed as follows. **Rahmawati Tri Iriani:** Conceptualization, Methodology, Software, Formal Analysis, Investigation, Data Curation, Visualization, Writing–Original Draft Preparation. **Danang Ariyanto:** Validation, Writing–Review & Editing, Supervision, Project Administration.

Declaration of Generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors used DeepL and Grammarly as AI-assisted tools for language translation, grammar checking and improving sentence clarity, and ChatGPT (OpenAI) to paraphrase and draft text, comparing and combining its suggestions with those of Claude. Claude (Anthropic; models Claude Opus 5 and Claude Opus 5.5) was also used to assist with the development and drafting of the R scripts, and to rewrite the text of this revision in plain English and shorten it to the journal's length. These tools were used to support language refinement, drafting and coding tasks and did not determine the research design, statistical interpretation, results or scientific conclusions. The authors reviewed, verified and revised all of the resulting text and code, and take full responsibility for the accuracy, integrity and final content of the manuscript.

Declaration of Competing Interest

Rahmawati Tri Iriani declares no competing interests. Danang Ariyanto also declares no competing interests. The authors therefore report no competing interests for this study.

Ethics Approval and Consent to Participate

This study analyses published province-level data. The unit of observation is the province-year. Ethics approval was not required for this study.

Funding and Acknowledgments

This research received no external funding. The analysis uses published statistical data from BPS and Pusdatin. These sources are acknowledged through citations in the text and the reference list.

Data and Code Availability

The authors' source workbooks, the derived province-year panel, the province crosswalk and the R scripts that produce every table and figure are provided as a supplementary file with this article. The workbooks were transcribed from published BPS and Pusdatin tables. These sources and their web addresses are cited in §2.1 and the reference list.

References

[1] Sutardi et al. "The Transformation of Rice Crop Technology in Indonesia: Innovation and Sustainable Food Security". In: *Agronomy* 13.1 (2023). Art. no. 1. DOI: 10.3390/agronomy13010001.

[2] M. Mariyah, J. Juraemi, and R. Mariati. "Analysis of Market Integration and Price Transmission of Rice in East Kalimantan". In: *Sosiohumaniora: Jurnal Ilmu-Ilmu Sosial dan Humaniora* 26.3 (2024), pp. 430–438. DOI: 10.24198/sosiohumaniora.v26i3.53418.

[3] L. P. Suciati and P. J. Kunu. "Deconstructing National Food Security: A Spatial Analysis of Paddy Field Conversion and Rice Production Volatility Based on Area Frame Sampling (AFS) Data". In: *Agricultural Power Journal* 3.1 (2026), pp. 12–21. DOI: 10.70076/apj.v3i1.150.

[4] Badan Pusat Statistik. *Statistik Indonesia 2025 (Statistical Yearbook of Indonesia 2025)*. Catalogue 1101001, Publication no. 03200.25004, ISSN 0126-2912, xlii+852 pp., p. 336 (Table 5.1.1). Jakarta: BPS-Statistics Indonesia, 2025. https://www.bps.go.id/en/publication/2025/02/28/8cfe1a589ad3693396d3db9f/statistical-yearbook-of-indonesia-2025.html.

[5] Badan Pusat Statistik. *Luas Panen Padi pada Tahun 2025 Mencapai Sekitar 11,32 Juta Hektare dengan Produksi Padi sebanyak 60,21 Juta Ton Gabah Kering Giling (GKG)*. Berita Resmi Statistik 2545. 2 February 2026. 2026. https://www.bps.go.id/id/pressrelease/2026/02/02/2545/luas-panen-padi-pada-tahun-2025-mencapai-sekitar-11-32-juta-hektare-dengan-produksi-padi-sebanyak-60-21-juta-ton-gabah-kering-giling--gkg-.html.

[6] Q. Wang, L. Sun, and X. Yang. "Identifying Spatial Determinants of Rice Yields in Main Producing Areas of China Using Geospatial Machine Learning". In: *ISPRS International Journal of Geo-Information* 13.3 (2024). Art. no. 76. DOI: 10.3390/ijgi13030076.

[7] A. Ansari, A. Pranesti, M. Telaumbanua, T. Alam, Taryono, R. A. Wulandari, B. D. A. Nugroho, and Supriyanta. "Evaluating the Effect of Climate Change on Rice Production in Indonesia Using Multimodelling Approach". In: *Heliyon* 9.9 (2023). Art. no. e19639. DOI: 10.1016/j.heliyon.2023.e19639.

[8] R. R. Rachmawati et al. "Regional Variability in Climate–Rice Yield Relationships across Java, Indonesia: A 30-Year Spatial-Temporal Analysis". In: *International Journal of Design & Nature and Ecodynamics* 20.8 (2025), pp. 1815–1825. DOI: 10.18280/ijdne.200812.

[9] Z. Kurniawan, R. Tiaharyadini, P. Jayadi, and W. Widhyanty. "Analyzing Climate Impacts on Rice Production in Sumatra through Spatiotemporal Machine Learning Models". In: *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)* 11.2 (2025), pp. 550–559. DOI: 10.33480/jitk.v11i2.7344.

[10] Nurhayati, A. M. Hasibuan, and M. S. Rahayu. "Climate Variability and Its Effects on Rice (Oryza sativa L.) Productivity, Pest Incidence, and Disease Occurrence in Pematang Setrak Village, Teluk Mengkudu District, Serdang Bedagai Regency, Indonesia". In: *Jurnal Penelitian Pendidikan IPA* 11.9 (2025), pp. 1160–1170. DOI: 10.29303/jppipa.v11i9.11771.

[11] A. S. Modjo, I. Dunggio, Hasim, and S. Rahim. "Pengaruh Luas Panen terhadap Produksi Padi di Indonesia: Analisis Regresi Data Tahun 2018–2024". In: *JPIM: Jurnal Penelitian Ilmiah Multidisipliner* 1.3 (2025), pp. 299–308. https://ojs.ruangpublikasi.com/index.php/jpim/article/view/397.

[12] T. Mulyaqin, R. Nurmalina, N. Kusnadi, and B. H. Trisasongko. "Spatial Heterogeneity of Rice Production Responses to ENSO Anomalies in Banten Province, Indonesia". In: *Journal of Applied Agricultural Science and Technology* 10.1 (2026), pp. 105–122. DOI: 10.55043/jaast.v10i1.503.

[13] Y. Sulaeman et al. "Yield Gap Variation in Rice Cultivation in Indonesia". In: *Open Agriculture* 9.1 (2024). Art. no. 20220241. DOI: 10.1515/opag-2022-0241.

[14] D. T. Utari, T. Yuliana, and A. P. Hendradewa. "A Panel Data Analysis of Rice Production in Ngawi Regency, East Java". In: *Proceedings of the 2nd International Seminar on Science and Technology (ISSTEC 2019)*. Vol. 474. Advances in Social Science, Education and Humanities Research. Atlantis Press, 2020, pp. 212–217. DOI: 10.2991/assehr.k.201010.031.

[15] D. S. Wicaksono, P. A. Kusumasari, H. M. Fajar, R. Fajritia, I. G. A. P. Anggraini, and Budiasih. "Implementasi Geographically Weighted Regression (GWR) pada Determinasi Faktor Produksi Beras di Indonesia Tahun 2021". In: *Seminar Nasional Official Statistics* 2023.1 (2023). Politeknik Statistika STIS, pp. 537–546. DOI: 10.34123/semnasoffstat.v2023i1.1714.

[16] R. A. Astari, Megawati, and S. Wahyudi. "Implementation of Geographically Weighted Lasso (GWL) in Analyzing Rice Production Factors in Indonesia". In: *Journal of Statistical Methods and Data Science* 1.2 (2024), pp. 44–51. DOI: 10.31258/jsmds.v1i2.16.

[17] N. F. Gamayanti, Junaidi, Fadjryani, and Nur'eni. "Analysis of Spatial Effects on Factors Affecting Rice Production in Central Sulawesi Using Geographically Weighted Panel Regression". In: *BAREKENG: Jurnal Ilmu Matematika dan Terapan* 17.1 (2023), pp. 361–370. DOI: 10.30598/barekengvol17iss1pp0361-0370.

[18] Sifriyani, I. N. Budiantara, M. F. F. Mardianto, and Asnita. "Determination of the Best Geographic Weighted Function and Estimation of Spatio Temporal Model–Geographically Weighted Panel Regression Using Weighted Least Square". In: *MethodsX* 12 (2024). Art. no. 102605. DOI: 10.1016/j.mex.2024.102605.

[19] C. Brunsdon, A. S. Fotheringham, and M. E. Charlton. "Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity". In: *Geographical Analysis* 28.4 (1996), pp. 281–298. DOI: 10.1111/j.1538-4632.1996.tb00936.x.

[20] Badan Pusat Statistik. *Statistik Indonesia 2020 (Statistical Yearbook of Indonesia 2020)*. Catalogue 1101001, Publication no. 03220.2007, ISSN 0126-2912, xl+748 pp., p. 270 (Table 5.1.1). Jakarta: BPS-Statistics Indonesia, 2020. https://www.bps.go.id/en/publication/2020/04/29/e9011b3155d45d70823c141f/statistical-yearbook-of-indonesia-2020.html.

[21] Badan Pusat Statistik. *Statistik Indonesia 2024 (Statistical Yearbook of Indonesia 2024)*. Catalogue 1101001, Publication no. 03200.24003, ISSN 0126-2912, xliv+804 pp., pp. 30–64 (Table 1.2.1); the Peraturan BMKG No. 1/2022 note is on p. 64. Jakarta: BPS-Statistics Indonesia, 2024. https://www.bps.go.id/id/publication/2024/02/28/c1bacde03256343b2bf769b0/statistik-indonesia-2024.html.

[22] Pusat Data dan Sistem Informasi Pertanian. *Statistik Iklim, Organisme Pengganggu Tanaman dan Dampak Perubahan Iklim 2017–2020*. 271 pp., pp. 21–26 (Tables 2.1.1.1–2.1.1.6, column T). Jakarta: Sekretariat Jenderal, Kementerian Pertanian, 2020. https://satudata.pertanian.go.id/assets/docs/publikasi/Statistik_Iklim,_OPT_dan_DPI_2017-2020.pdf.

[23] Pusat Data dan Sistem Informasi Pertanian. *Statistik Iklim, Organisme Pengganggu Tanaman dan Dampak Perubahan Iklim 2020–2023*. 265 pp., pp. 41–46 (Tables 2.1.1.1–2.1.1.6, column T). Jakarta: Sekretariat Jenderal, Kementerian Pertanian, 2023. https://satudata.pertanian.go.id/assets/docs/publikasi/Statistik_IKLIM,_OPT_dan_DPI_2023.pdf.

[24] Pusat Data dan Sistem Informasi Pertanian. *Statistik Iklim, Organisme Pengganggu Tanaman dan Dampak Perubahan Iklim 2023–2025*. 298 pp., Tabel 2.1.1.1–2.1.1.7, column T; the 2025 column is data through 24 October 2025; released 31 Oct 2025. Jakarta: Sekretariat Jenderal, Kementerian Pertanian, Oct. 2025. https://satudata.pertanian.go.id/assets/docs/publikasi/ba64cc8e022dbdc43862f2c839bc82b8.pdf.

[25] R. M. O'Brien. "A Caution Regarding Rules of Thumb for Variance Inflation Factors". In: *Quality & Quantity* 41.5 (2007), pp. 673–690. DOI: 10.1007/s11135-006-9018-6.

[26] B. H. Baltagi. *Econometric Analysis of Panel Data*. 6th ed. Cham: Springer, 2021. DOI: 10.1007/978-3-030-53953-5.

[27] Y. Croissant and G. Millo. "Panel Data Econometrics in R: The plm Package". In: *Journal of Statistical Software* 27.2 (2008), pp. 1–43. DOI: 10.18637/jss.v027.i02.

[28] I. Gollini, B. Lu, M. Charlton, C. Brunsdon, and P. Harris. "GWmodel: An R Package for Exploring Spatial Heterogeneity Using Geographically Weighted Models". In: *Journal of Statistical Software* 63.17 (2015), pp. 1–50. DOI: 10.18637/jss.v063.i17.

[29] J. C. Driscoll and A. C. Kraay. "Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data". In: *Review of Economics and Statistics* 80.4 (1998), pp. 549–560. DOI: 10.1162/003465398557825.

[30] C. M. Hurvich, J. S. Simonoff, and C.-L. Tsai. "Smoothing Parameter Selection in Nonparametric Regression Using an Improved Akaike Information Criterion". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 60.2 (1998), pp. 271–293. DOI: 10.1111/1467-9868.00125.

[31] R. Koenker. "A Note on Studentizing a Test for Heteroscedasticity". In: *Journal of Econometrics* 17.1 (1981), pp. 107–112. DOI: 10.1016/0304-4076(81)90062-2.

[32] T. S. Breusch and A. R. Pagan. "A Simple Test for Heteroscedasticity and Random Coefficient Variation". In: *Econometrica* 47.5 (1979), pp. 1287–1294. DOI: 10.2307/1911963.

[33] G. C. Chow. "Tests of Equality between Sets of Coefficients in Two Linear Regressions". In: *Econometrica* 28.3 (1960), pp. 591–605. DOI: 10.2307/1910133.

[34] J. A. Hausman. "Specification Tests in Econometrics". In: *Econometrica* 46.6 (1978), pp. 1251–1271. DOI: 10.2307/1913827.

[35] Y. Leung, C.-L. Mei, and W.-X. Zhang. "Statistical Tests for Spatial Nonstationarity Based on the Geographically Weighted Regression Model". In: *Environment and Planning A: Economy and Space* 32.1 (2000), pp. 9–32. DOI: 10.1068/a3162.

[36] A. R. da Silva and A. S. Fotheringham. "The Multiple Testing Issue in Geographically Weighted Regression". In: *Geographical Analysis* 48.3 (2016), pp. 233–247. DOI: 10.1111/gean.12084.