



Received: 20-08-2026

Accepted: 30-09-2026

International Journal of Advanced Multidisciplinary Research and Studies

ISSN: 2583-049X

Inverse Optimal Transport Modelling of Hidden Barriers to Secondary-School Access in Dispersed Rural Settlements

¹ Ndapewa Johannes, ² Philipus Nangolo, ³ Immanuel Nakambale, ⁴ George Johannes Katambila

¹ Department of Strategic Management, International University of Management, Nkurenkuru Campus, Namibia

² Department of Mathematics and Economics, International University of Management, Nkurenkuru Campus, Namibia

³ Department of Strategic Management, International University of Management, Nkurenkuru Campus, Namibia

⁴ Department of Information and Communications Technology, International University of Management, Nkurenkuru Campus, Namibia

Corresponding Author: Ndapewa Johannes

Abstract

Distance is observable, but the effective cost of secondary-school access in dispersed rural settlements is shaped by barriers that are partly hidden in enrolment and attendance flows. This study develops an entropic inverse optimal transport framework that recovers latent generalized access costs from origin–school flow matrices while respecting learner totals and school capacities. The model combines network distance with river crossings, road condition, safety exposure, transport fares and household–school compatibility. Existence, uniqueness, Sinkhorn representation, invariance, parameter identifiability, stability and accessibility bounds are established. A reproducible synthetic experiment with 24 settlements, five schools and 1,230 capacity-weighted places demonstrates the method without representing simulated observations as field

evidence. The full inverse model reduced Kullback–Leibler flow error from 0.38579 for a distance-only specification to 0.00431 and recovered standardized barrier coefficients with an Euclidean error of 0.3157. River crossings and transport fares produced the largest non-distance penalties. A soft-min accessibility index translated the recovered cost surface into settlement priorities. Counterfactual transport subsidies and a hostel reduced both average and upper-tail generalized costs, whereas a candidate school-placement rule exposed an equity efficiency trade-off. The framework therefore converts routinely available origin–destination counts into interpretable planning evidence, while explicitly separating structural inference from causal claims. Field validation, uncertainty quantification and participatory specification remain necessary before operational allocation.

Keywords: Inverse Optimal Transport, Rural Education, Sinkhorn Algorithm, Latent Access Costs, School Planning, Accessibility Equity

Introduction

Secondary-school access is rarely governed by straight-line or road-network distance alone. Two learners living equally far from a school may face substantially different effective costs because one must cross a seasonal river, travel on an unreliable road, pay several fares, walk through an unsafe corridor or combine schooling with household responsibilities. These barriers are difficult to observe at scale and are commonly collapsed into distance bands. The resulting maps can therefore describe proximity while misrepresenting feasible access.

The distinction is especially important in dispersed rural settlements. Sparse demand, irregular transport, limited school capacity and seasonal infrastructure create a many-origin, few-destination allocation problem. School-selection and attendance flows contain information about this allocation: repeated avoidance of a geographically close school is evidence of an unmeasured cost, provided capacity, cohort size and alternative destinations are accounted for. Conventional gravity and discrete-choice models exploit flows, but they do not automatically enforce both origin totals and destination capacities. Optimal transport supplies precisely that constrained geometry ^[1-4].

Entropic optimal transport adds a strictly convex entropy term to the transportation problem. Its solution is smooth, unique and computable by matrix scaling through the Sinkhorn algorithm ^[3, 5, 6]. These properties make the forward map differentiable and hence suitable for inversion. Instead of assuming a cost matrix and predicting school flows, inverse optimal transport observes flows and estimates the cost parameters that make them optimal under the marginal constraints. Recent inverse-transport studies show that latent costs can be learned from matching or mobility data ^[7-10], but educational planning has not

yet fully used this structure to distinguish visible distance from river, safety, fare and household barriers.

The educational relevance extends beyond mathematical fit. Village-based schools can reduce access barriers^[11], and transport interventions can materially affect secondary enrolment^[12]. Rural policy also recognises the joint importance of roads, transport, services and household welfare rather than treating distance as an isolated constraint^[13]. A useful planning model should therefore recover a multidimensional cost, quantify uncertainty and permit counterfactual comparison of school placement, transport subsidies and hostels. It must also avoid a common interpretive error: a recovered cost is an observationally equivalent generalized barrier, not by itself a causal effect.

This article contributes a unified mathematical and planning framework. First, it formulates rural school access as a capacity-constrained entropic transport problem with observable and latent cost components. Second, it proves existence and uniqueness of the forward plan, gives its Sinkhorn factorization, identifies the row-column potential ambiguity, states a usable identifiability condition, derives the inverse gradient and establishes stability of recovered parameters and accessibility indices. Third, it defines settlement-level soft generalized costs and an equity-sensitive intervention objective. Fourth, it illustrates the entire pipeline in a transparent simulation calibrated to the scale and spatial logic of dispersed settlements. The objectives are to determine when latent barriers can be recovered, evaluate whether the full model outperforms distance alone, and show how recovered costs can inform interventions without overstating empirical validity.

Materials and Methods

Let the set of settlements be $I = \{1, \dots, n\}$ and the set of secondary schools be $J = \{1, \dots, m\}$. The vector $\mathbf{a} \in \mathbb{R}_{++}^n$ contains the shares of eligible learners at origins, and $\mathbf{b} \in \mathbb{R}_{++}^m$ contains normalized school capacities. Both sum to one. A transport plan $P = (P_{ij})$ records the share of learners from settlement i assigned to school j . Feasible plans belong to the transportation polytope:

$$\mathcal{U}(\mathbf{a}, \mathbf{b}) = \{P \in \mathbb{R}_{+}^{n \times m} : P\mathbf{1}_m = \mathbf{a}, P^T\mathbf{1}_n = \mathbf{b}\} \quad (1)$$

The marginal constraints are substantive. The first preserves the cohort represented by each settlement; the second prevents the model from assigning unlimited learners to an attractive school. Where capacity is not binding, $\mathbf{b}$ may instead be estimated from observed enrolment or extended through an unbalanced formulation. This article uses balanced transport because it exposes the central inverse problem cleanly.

For each origin-school pair, let x_{ijr} denote standardized feature r . The feature tensor includes distance, river-crossing exposure, poor-road exposure, safety exposure, transport fare and a household-school compatibility measure. The latter varies across both origin and destination; a purely origin-specific household term is not identifiable from a balanced plan because it is absorbed by a row potential. The generalized cost is linear in the first implementation,

$$C_{ij}(\theta) = \sum_{r=1}^q \theta_r x_{ijr}, \quad \theta \in \Theta \subset \mathbb{R}^q \quad (2)$$

Nonnegative coefficients are imposed for features coded as barriers. Interactions, splines or neural cost maps can replace the linear form, but the linear specification keeps the recovered weights interpretable and permits direct identifiability analysis. Distance is standardized and its coefficient is fixed at one. This normalization is essential because $\mathbf{C}$ and the entropy parameter can otherwise be multiplied by the same positive constant without changing the kernel ratios.

Given $\varepsilon > 0$, the forward entropic transport plan solves:

$$P_\theta = \operatorname{argmin}_{P \in \mathcal{U}(\mathbf{a}, \mathbf{b})} \{C(\theta), P\} + \varepsilon \sum_{i=1}^n \sum_{j=1}^m P_{ij} (\log P_{ij} - 1) \quad (3)$$

The inner product is Frobenius. Small ε concentrates mass on low-cost routes, while larger ε represents greater unobserved heterogeneity and yields smoother flows. The parameter is therefore not merely numerical; it controls the scale at which observed deviations from minimum-cost allocation are treated as heterogeneity rather than systematic barriers.

Lemma 1 (existence and uniqueness of the forward plan): If $\mathbf{a}$ and $\mathbf{b}$ are strictly positive probability vectors and $\mathbf{C}$ is finite, then for every $\varepsilon > 0$ the forward problem has a unique solution P_θ , and every entry of that solution is strictly positive.

Proof: The transportation polytope is nonempty because $\mathbf{a}\mathbf{b}^T \in \mathcal{U}(\mathbf{a}, \mathbf{b})$, and it is closed and bounded in finite-dimensional space, hence compact. The objective is continuous on the polytope after extending $t \log t$ continuously at zero. Existence follows from the Weierstrass theorem. On the positive orthant, the entropy contribution has diagonal Hessian entries $\varepsilon/P_{ij} > 0$, so the objective is strictly convex along every nonzero feasible direction. Consequently two distinct minimizers cannot exist. Finally, the derivative of $t(\log t - 1)$ tends to negative infinity as t approaches zero. Because the marginals are strictly positive, a feasible cycle perturbation can move mass into any zero cell while preserving both marginals and strictly decrease the objective. Thus the unique minimizer is positive.

Proposition 1 (Sinkhorn representation): Define the positive Gibbs kernel by:

$$K_{ij}(\theta) = \exp \left[-\frac{C_{ij}(\theta)}{\varepsilon} \right] \quad (4)$$

There exist positive scaling vectors $\mathbf{u} \in \mathbb{R}_{++}^n$ and $\mathbf{v} \in \mathbb{R}_{++}^m$, unique up to a reciprocal scalar, such that:

$$P_\theta = \text{diag}(u)K(\theta)\text{diag}(v), \quad u \odot (Kv) = a, \quad v \odot (K^\top u) = b \quad (5)$$

Proof: Introduce multipliers α and β for the row and column constraints. Stationarity of the Lagrangian gives $C_{ij} + \varepsilon \log P_{ij} - \alpha_i - \beta_j = 0$. Hence $P_{ij} = \exp(\alpha_i/\varepsilon) \exp(-C_{ij}/\varepsilon) \exp(\beta_j/\varepsilon) = u_i K_{ij} v_j$. Substitution of the marginal constraints yields the two scaling equations. Positivity of K and the matrix-scaling theorem give existence and uniqueness up to $u \mapsto tu$, $v \mapsto v/t$ [5].

The scaling equations yield the Sinkhorn iteration:

$$u^{(t+1)} = a \oslash (Kv^{(t)}), \quad v^{(t+1)} = b \oslash (K^\top u^{(t+1)}) \quad (6)$$

Where $\oslash$ denotes componentwise division. Log-domain updates should be used when costs are large relative to ε . Iteration stopped when the maximum absolute marginal residual was below 10^{-10} or after 10,000 iterations.

Lemma 2 (row–column invariance): For arbitrary vectors $r \in \mathbb{R}^n$ and $s \in \mathbb{R}^m$, replacing C_{ij} by $C_{ij} + r_i + s_j$ does not change the optimal plan.

Proof: For every feasible P ,

$$\sum_{i,j} (r_i + s_j) P_{ij} = \sum_i r_i a_i + \sum_j s_j b_j \quad (7)$$

Which is constant over $\mathcal{U}(a, b)$. The two optimization problems therefore differ only by a constant and have the same unique minimizer.

Lemma 2 is central to responsible interpretation. A hardship feature that varies only by settlement cannot be recovered from one balanced origin–destination table: it adds the same value to every route leaving that settlement. Likewise, a destination-only prestige term is absorbed by a column potential. Identification requires route variation or information outside the fixed marginals, such as attendance rates, rejected applications or longitudinal changes.

Let $\mathcal{R} = \{r \mathbf{1}_m^\top + \mathbf{1}_n s^\top : r \in \mathbb{R}^n, s \in \mathbb{R}^m\}$ be the row–column potential space, and let $\Pi_{\mathcal{R}^\perp}$ denote orthogonal projection onto its complement. Write $X_r = (x_{ijr})$.

Theorem 1 (identifiability modulo potentials): Fix ε and the coefficient scale. If the matrices $\Pi_{\mathcal{R}^\perp} X_1, \dots, \Pi_{\mathcal{R}^\perp} X_q$ are linearly independent, then $P_\theta = P_{\theta'}$ implies $\theta = \theta'$.

Proof. By Proposition 1,

$$\log P_{ij} = \log u_i + \log v_j - \frac{C_{ij}(\theta)}{\varepsilon} \quad (8)$$

If $P_\theta = P_{\theta'}$, subtracting the two representations shows that $C(\theta) - C(\theta')$ equals a row potential plus a column potential and thus belongs to $\mathcal{R}$. Applying $\Pi_{\mathcal{R}^\perp}$ gives $\sum_r (\theta_r - \theta'_r) \Pi_{\mathcal{R}^\perp} X_r = 0$. Linear independence forces every coefficient difference to be zero.

Corollary 1. A feature matrix lying in $\mathcal{R}$ is not identifiable from a single balanced flow matrix, regardless of sample size or optimization accuracy.

Let $P^{\text{obs}} \in \mathcal{U}(a, b)$ denote the observed flow matrix after confidentiality-preserving aggregation and, if necessary, iterative proportional fitting to common marginals. Define the entropy-conjugate value:

$$\Phi_\varepsilon(-C) = \max_{P \in \mathcal{U}(a, b)} \{ -\langle C, P \rangle - \varepsilon \sum_{i,j} P_{ij} (\log P_{ij} - 1) \} \quad (9)$$

The inverse estimator minimizes the convex discrepancy:

$$\hat{\theta} = \operatorname{argmin}_{\theta \in \Theta} \left\{ \mathcal{L}_\lambda(\theta) = \Phi_\varepsilon[-C(\theta)] + \langle C(\theta), P^{\text{obs}} \rangle + \frac{\lambda}{2} \|\theta\|_2^2 \right\} \quad (10)$$

This objective is a Fenchel–Young loss. It is nonnegative up to an additive constant and equals zero at a perfectly represented plan when $\lambda = 0$. The regularizer stabilizes correlated barrier features rather than replacing the structural identifiability condition.

Proposition 2 (inverse gradient and moment matching): For a linear cost and interior parameter point,

$$\frac{\partial \mathcal{L}_\lambda}{\partial \theta_r} = \langle P^{\text{obs}} - P_\theta, X_r \rangle + \lambda \theta_r \quad (11)$$

If $\lambda = 0$ and the optimum is interior, fitted and observed feature moments agree: $\langle P_\theta, X_r \rangle = \langle P^{\text{obs}}, X_r \rangle$ for every identifiable feature.

Proof: The maximizer defining Φ_ε is unique by Lemma 1. Danskin’s theorem therefore gives the derivative of $\Phi_\varepsilon[-C(\theta)]$ as $-(P_\theta, X_r)$. Differentiating the remaining terms gives the displayed gradient. The moment equations follow by setting the unregularized gradient to zero.

Theorem 2 (unique and stable regularized estimator): Suppose Θ is closed and convex and $\lambda > 0$. Then $\mathcal{L}_\lambda$ has at most one minimizer. If minimizers based on feasible flow matrices P and Q are interior, then:

$$\|\hat{\theta}(P) - \hat{\theta}(Q)\|_2 \leq \frac{1}{\lambda} \|(\langle P - Q, X_r \rangle)_{r=1}^q\|_2 \quad (12)$$

Proof: The conjugate value is convex as a supremum of affine functions of $-C$, and composition with a linear cost preserves convexity. Adding $\lambda \| \theta \|_2^2/2$ makes the objective λ -strongly convex, proving uniqueness. Let $d = \hat{\theta}(P) - \hat{\theta}(Q)$. Strong monotonicity of the gradient and the two first-order conditions imply $\lambda \|d\|_2^2 \leq \langle d, z(P - Q) \rangle$, where $z(P - Q)_r = \langle P - Q, X_r \rangle$. Cauchy–Schwarz and cancellation of $\|d\|_2$ give the bound.

To translate pairwise costs into a settlement indicator, define the soft generalized access cost:

$$g_i(C) = -\varepsilon \log \left[\sum_{j=1}^m b_j \exp \left(-\frac{C_{ij}}{\varepsilon} \right) \right] \quad (13)$$

And the unit-free accessibility score $A_i = \exp[-g_i(C)]$. The log-sum-exp form rewards access to several feasible schools rather than only the nearest one. The capacity weights discourage an interpretation in which a small school offers unlimited access.

Theorem 3 (robustness of the accessibility index): For any two finite cost matrices C and D ,

$$|g_i(C) - g_i(D)| \leq \max_j |C_{ij} - D_{ij}| \quad (14)$$

Proof: Let $\delta = \max_j |C_{ij} - D_{ij}|$. Then $D_{ij} - \delta \leq C_{ij} \leq D_{ij} + \delta$. Exponentiation, capacity-weighted summation and application of $-\varepsilon \log$ preserve the corresponding bound $g_i(D) - \delta \leq g_i(C) \leq g_i(D) + \delta$.

Corollary 2: Uniformly accurate route-cost recovery implies uniformly accurate settlement accessibility: $\|g(C) - g(D)\|_\infty \leq \|C - D\|_\infty$.

For policy design, let z encode an intervention and $C(z)$ its modified cost matrix. A planner may minimize a combination of population-weighted mean cost, upper-tail deprivation and inequality,

$$\mathcal{J}(z) = \sum_i a_i g_i[C(z)] + \eta G_a\{g[C(z)]\} + \xi Q_{0.90,a}\{g[C(z)]\}, \quad z \in \mathcal{Z}(B) \quad (15)$$

Where G_a is a weighted Gini index, $Q_{0.90,a}$ is the weighted 90th percentile and $\mathcal{Z}(B)$ is the set of interventions satisfying budget B . The weights η and ξ are normative and must be stated, not estimated from flows.

Proposition 3 (marginal value of a route subsidy): If an intervention reduces route i, j by subsidy s_{ij} , then the derivative of g_i with respect to that subsidy is negative and equals the soft assignment share,

$$\frac{\partial g_i}{\partial s_{ij}} = -\frac{b_j \exp[-C_{ij}/\varepsilon]}{\sum_k b_k \exp[-C_{ik}/\varepsilon]} \quad (16)$$

Proof: Substitute $C_{ij} - s_{ij}$ into the definition of g_i and differentiate the log-sum-exp expression. The fraction is positive and sums to one over destinations, proving monotonicity and giving the exact local benefit.

The empirical workflow requires: geocoded settlement and school identifiers; eligible-learner or attendance totals by origin; destination capacities; an origin–school flow table; and route features measured for the same time window. River and road variables should be season-specific when access changes across the year. Safety data should use privacy-preserving spatial aggregation. Fares should measure the complete journey rather than a single vehicle leg. Household information should be included only with ethical approval and aggregation sufficient to prevent re-identification.

The numerical illustration used 24 synthetic settlements in a 60 km by 45 km region, five schools, positive origin masses generated from 28–77 learners and destination capacities proportional to 230–265 places. A seasonal river divided the region. Six route features were scaled to $[0,1]$. The data-generating coefficient vector was $(1.00, 1.65, 1.10, 0.80, 1.30, 0.70)$ and $\varepsilon = 0.32$. A positive multiplicative perturbation was applied to the generated plan and rebalanced to the same margins, mimicking aggregated sampling and reporting noise without creating impossible totals. The inverse model fixed the distance coefficient at one, constrained all barrier coefficients to $[0,4]$ and used $\lambda = 0.0005$. Optimization used L-BFGS-B with the analytic gradient. The simulation seed was 863.

Model fit was assessed by Kullback–Leibler divergence, cellwise root-mean-square error and settlement-level top-school agreement. The full model was compared with a distance-only model and the known oracle. Entropy sensitivity used $\varepsilon \in \{0.16, 0.24, 0.32, 0.48, 0.64\}$. Policy scenarios reduced fare exposure for the eight least-accessible settlements, introduced a hostel effect at School 4, or added a 12% capacity school selected from a predeclared 48-point grid by mean-plus-inequality score. These are mechanism demonstrations, not forecasts.

Results and Discussion

Fig 1 displays the synthetic spatial system. Settlement marker size represents learner mass, colour represents recovered soft generalized cost and triangles locate existing schools. High-cost settlements are not simply the most distant: the river and route-specific barriers alter the effective surface. This is the central practical reason for inversion. A map based only on kilometres would assign the same policy meaning to geographically similar routes even when their observed allocations consistently differ.

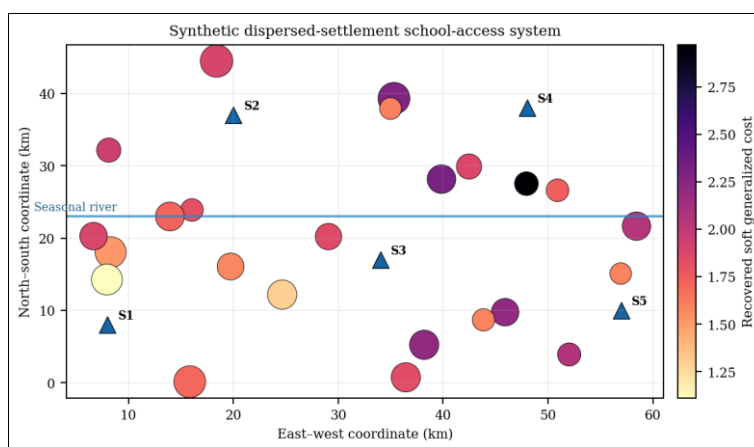


Fig 1: Synthetic settlements, schools, river barrier and recovered generalized access cost. Marker size is proportional to the settlement learner share

Table 1: Variables and their mathematical roles

Quantity	Meaning	Planning interpretation
a_i	Origin learner share	Size of eligible cohort in settlement i
b_j	Destination capacity share	Available places at school j
p_{ij}^{obs}	Observed flow share	Enrolment or regular-attendance movement
x_{ijr}	Route feature	Distance, river, road, safety, fare or compatibility
θ_r	Recovered coefficient	Generalized penalty per standardized feature unit
ϵ	Entropy scale	Unobserved heterogeneity and plan smoothness
g_i	Soft generalized cost	Settlement-level accessibility burden

The recovered coefficients were close to the data-generating values (Table 2 and Fig 2). The coefficient error was 0.3157 in Euclidean norm. River crossing remained the largest non-distance barrier, followed by fare and poor-road exposure. The slight downward shrinkage is expected from regularization and correlated route features. Importantly, the household term was constructed as a household–school compatibility interaction. Had it been constant across schools for each origin, Corollary 1 would make it unrecoverable even with noiseless data.

Table 2: Recovery of standardized barrier coefficients

Barrier	Data-generating value	Recovered value	Absolute error
Distance (scale fixed)	1.0000	1.0000	0.0000
River crossing	1.6500	1.3810	0.2690
Poor-road exposure	1.1000	0.9782	0.1218
Safety exposure	0.8000	0.7128	0.0872
Transport fare	1.3000	1.2306	0.0694
Household–school compatibility	0.7000	0.6691	0.0309

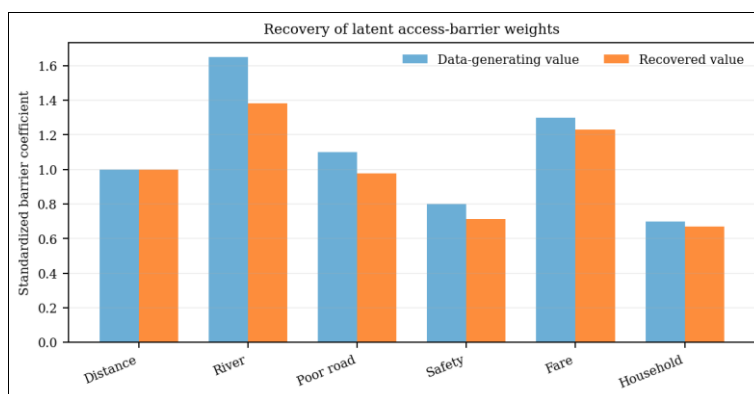


Fig 2: Data-generating and recovered standardized barrier coefficients

The distance-only model achieved 87.5% top-school agreement, which could appear acceptable if evaluated only by the modal destination. Its distributional fit was nevertheless poor: Kullback–Leibler divergence was 0.38579 and cellwise root-mean-square error was 0.007867. The full inverse model reduced these values to 0.00431 and 0.000806, respectively, while matching the top school in all settlements. Its divergence approached the oracle value of 0.00409. Thus distance captured the broad spatial ordering, but not the allocation intensities from which capacity and policy decisions are made.

Table 3: Flow reconstruction performance

Model	KL divergence	Cellwise RMSE	Top-school agreement
Distance only	0.38579	0.007867	87.50%
Full inverse model	0.00431	0.000806	100.00%
Oracle cost	0.00409	0.000822	100.00%

Fig 3 shows that the gain was not a graphical artefact. Predicted and observed cells clustered near the equality line under the full model. Sensitivity analysis also revealed a U-shaped pattern: $\epsilon = 0.32$, the generating value, minimized divergence and coefficient error among the tested values. At $\epsilon = 0.16$, the plan was too concentrated and forced coefficients to compensate; at $\epsilon = 0.64$, excessive smoothing again distorted the cost scale. Entropy should consequently be selected by held-out flows, repeated time periods or a substantively justified heterogeneity scale, not solely by computational convenience.

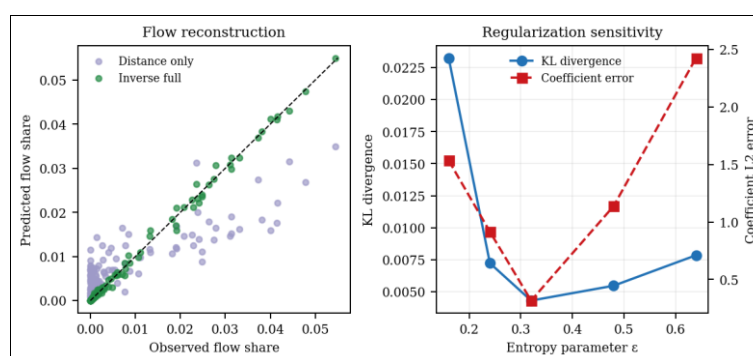


Fig 3: Flow reconstruction and sensitivity to the entropy parameter

The intervention calculations illustrate how latent costs can change decisions. The baseline population-weighted mean soft cost was 1.8509, the weighted 90th percentile was 2.2773 and the weighted Gini was 0.2262. A fare subsidy focused on the eight least-accessible origins reduced all three measures. The School 4 hostel produced a larger mean and upper-tail reduction and also lowered inequality in this realization. The candidate school lowered inequality to the same level as the hostel but slightly increased the mean and upper tail after capacities were rebalanced. This is not a contradiction: shifting 12% of destination capacity to a new site benefits isolated settlements while slightly reducing the capacity weight of existing schools used by many learners. The example demonstrates why placement should be judged against an explicit social objective rather than a distance-minimizing rule.

Table 4: Counterfactual planning indicators

Scenario	Weighted mean cost	Weighted 90th percentile	Weighted Gini
Baseline	1.8509	2.2773	0.2262
Targeted transport subsidy	1.8193	2.2345	0.2103
Hostel at School 4	1.7479	2.1976	0.2093
Equity-oriented new school	1.8598	2.3181	0.2093

These policy values are dimensionless because the input features were standardized. For budgeting, one coefficient should be anchored to a monetary or time unit. If the distance coefficient is calibrated as cost per kilometre, every other coefficient can be expressed as a kilometre-equivalent penalty. A recovered river coefficient of 1.3810 would then mean that, over the observed choice set and conditional on included features, a full-scale river exposure has the same allocation effect as 1.3810 standardized distance units. This is a structural equivalence, not a claim about psychological valuation.

The framework offers three advantages over a nearest-school indicator. First, it respects competition for finite places. Second, it distinguishes route barriers through the features required to rationalize actual flows. Third, it produces differentiable policy metrics. Proposition 3 identifies where a small subsidy has the highest immediate accessibility return, while the objective J makes distributional preferences transparent. These properties connect mathematical estimation to operational planning.

Several limitations delimit the claims. The experiment is synthetic and establishes numerical behaviour under controlled conditions; it does not estimate barriers in any named community. Observational flows conflate constrained choice, information, perceived school quality, language, social networks and administrative placement. Omitted school-quality variables may therefore appear as transport costs. Capacity records may themselves reflect policy decisions and should not automatically be treated as exogenous. Balanced transport ignores learners who never enrol or attend; unbalanced transport or an explicit outside option is needed when exclusion is a central outcome. Cross-sectional inversion does not identify causal effects of subsidies or hostels. Those effects require longitudinal variation, natural experiments or randomized designs.

The proofs also clarify data-design priorities. More observations do not solve a structural invariance. Origin-only hardship, destination-only attractiveness and common cost shifts cannot be separated from balanced margins without additional information. Repeated seasons can help because river and road exposure vary while settlement identity remains fixed. Application and rejection data can help distinguish preference from capacity rationing. Attendance-frequency flows can reveal barriers hidden by annual enrolment. Participatory route audits can validate whether a high recovered residual corresponds to a real crossing, fare, safety concern or omitted educational attribute.

For a field study, uncertainty should be reported through a cluster bootstrap over settlements or time blocks. Each bootstrap sample must repeat marginal construction, feature scaling, entropy selection and inverse estimation. Spatial cross-validation should hold out contiguous areas rather than random cells, because nearby routes share infrastructure. Model comparison should include distance-only, gravity, multinomial choice and inverse transport baselines. A barrier should be regarded as decision-relevant only when its sign is stable, its predictive gain persists out of sample and local stakeholders recognize a plausible mechanism.

Conclusion

This study formulated secondary-school access in dispersed rural settlements as an inverse entropic optimal transport problem. The approach infers a multidimensional generalized cost from origin–school flows while preserving learner totals and school capacities. The mathematical analysis established a unique positive forward solution, its Sinkhorn scaling form, the row–column invariance that limits interpretation, an identifiability criterion, a differentiable inverse loss, regularized stability and robust soft-accessibility bounds.

In the reproducible simulation, distance alone recovered the dominant school for most settlements but failed to reproduce the full flow distribution. Adding river, road, safety, fare and household–school interaction features reduced Kullback–Leibler error by almost two orders of magnitude and recovered the latent barrier ordering. Counterfactuals showed that a targeted transport subsidy and a hostel could reduce average, upper-tail and unequal access, while school placement could involve an explicit equity–efficiency trade-off.

The principal contribution is therefore not a claim that flows reveal every hardship. It is a disciplined method for stating which barriers can be inferred, which are observationally confounded and how recovered costs can enter transparent planning objectives. The next research step is field validation using seasonally resolved attendance flows, verified capacities, route audits and intervention data. With those safeguards, inverse optimal transport can complement not replace community knowledge in school placement, subsidy targeting and hostel planning.

Recommendations

Education authorities should construct privacy-preserving origin–school flow matrices at term level and link them to contemporaneous road, river, safety and fare data. Distance should remain a benchmark but not the sole accessibility measure. Capacity figures should be audited before fitting a balanced model, and an outside option should be included where non-enrolment or irregular attendance is substantial.

Model implementation should fix a cost scale, test feature independence after removing row and column potentials, select ϵ using temporal or spatial validation and report sensitivity across plausible values. Recovered coefficients should be translated into time, monetary or kilometre-equivalent units only after an external anchor is supplied. Policy rankings should report mean, upper-tail and inequality outcomes separately.

Planning decisions should combine model outputs with route inspections and structured consultation with learners, caregivers, school leaders and transport providers. Subsidies should be evaluated for take-up and reliability; hostels should be assessed for safeguarding, staffing and recurrent cost; and new schools should be compared with flexible transport and satellite provision. Prospective pilots or phased implementation are recommended before scale-up.

Declarations

Acknowledgement: The authors have no acknowledgements to declare.

Author Contributions: All authors contributed to the conception and design of the study, development of the mathematical framework, interpretation of the results, preparation of the manuscript and critical revision of its intellectual content. All authors read and approved the final manuscript.

Funding: The authors received no external financial support for the research, authorship or publication of this article.

Conflict of Interest: The authors declare that they have no competing interests.

Ethical Statement: This study used a synthetic and reproducible numerical experiment and did not involve human participants, personal records or personally identifiable information. Ethical approval and informed consent were therefore not required. Any future application involving learner-level or household-level data will require appropriate institutional ethical approval, data-protection authorisation and informed consent or an approved waiver.

Data and Code Availability: No external participant dataset was used in this study. The synthetic data-generating procedure, model parameters and random seed required to reproduce the numerical results are reported in the Materials and Methods section. Supporting computational code is available from the corresponding author upon reasonable request.

References

1. Kantorovich L.V. On the translocation of masses. *Journal of Mathematical Sciences*. 2006; 133(4):1381–1382.
2. Villani C. *Optimal Transport: Old and New*. Berlin, Springer, 2009.
3. Peyré G, Cuturi M. Computational optimal transport. *Foundations and Trends in Machine Learning*. 2019; 11(5-6):355–607.

4. Galichon A. Optimal Transport Methods in Economics. Princeton, Princeton University Press, 2016.
5. Sinkhorn R, Knopp P. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics*. 1967; 21(2):343-348.
6. Cuturi M. Sinkhorn distances: Lightspeed computation of optimal transportation distances. *Advances in Neural Information Processing Systems*. 2013; 26:2292-2300.
7. Dupuy A, Galichon A. Personality traits and the marriage market. *Journal of Political Economy*. 2014; 122(6):1271-1319.
8. Galichon A, Salanié B. Cupid's invisible hand: Social surplus and identification in matching models. *Review of Economic Studies*. 2022; 89(5):2600-2629.
9. Persiianov M, Asadulaev A, Andreev A, Zadorozhny K, Karkhanis D, Panov M, *et al*. Inverse entropic optimal transport solves semi-supervised learning via data likelihood maximization. *arXiv*, 2024. arXiv:2410.02628.
10. Martinez M. Learning urban access costs from origin-destination flows via inverse optimal transport. *arXiv*, 2026. arXiv:2606.14157.
11. Burde D, Linden LL. Bringing education to Afghan girls: A randomized controlled trial of village-based schools. *American Economic Journal: Applied Economics*. 2013; 5(3):27-40.
12. Muralidharan K, Prakash N. Cycling to school: Increasing secondary school enrollment for girls in India. *American Economic Journal: Applied Economics*. 2017; 9(3):321-350.
13. National Planning Commission. National Rural Development Policy. Windhoek, Government of the Republic of Namibia, 2012, p. 64.
14. UNESCO. Global Education Monitoring Report 2020: Inclusion and Education-All Means All. Paris, UNESCO, 2020, p. 424.
15. World Bank. World Development Report 2018: Learning to Realize Education's Promise. Washington DC, World Bank, 2018, p. 216.
16. Genevay A, Peyré G, Cuturi M. Learning generative models with Sinkhorn divergences. *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*. 2018; 84:1608-1617.
17. Feydy J, Séjourné T, Vialard FX, Amari S, Trouvé A, Peyré G. Interpolating between optimal transport and maximum mean discrepancy using Sinkhorn divergences. *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*. 2019; 89:2681-2690.
18. Chizat L, Peyré G, Schmitzer B, Vialard FX. Scaling algorithms for unbalanced optimal transport problems. *Mathematics of Computation*. 2018; 87(314):2563-2609.
19. Cominetti R, San Martín J. Asymptotic analysis of the exponential penalty trajectory in linear programming. *Mathematical Programming*. 1994; 67:169-187.
20. Weed J. An explicit analysis of the entropic penalty in linear programming. *Proceedings of Machine Learning Research*. 2018; 75:1841-1855.