CLASSICAL CONTROL STRENGTH AND EXACT/CERTIFIED ANCHORS IN NEURAL TSP EVALUATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Across four separately controlled Euclidean TSP blocks spanning $n \in \{20, 50\}$ and uniform/clustered spatial generators, fixed-start nearest-neighbor construction lies roughly 17–25% above the exact or certified optimization anchor, whereas the specified audited local-search controls lie roughly 1.2–3.6% above it. The clustered protocol and primary endpoint were fixed before the 50,000-step training runs, with the held-out evaluation sets sampled only afterward. The clustered replication reproduces this separation at both sizes. This replicated separation shows how comparator strength changes the scale on which comparisons between learned and classical methods are interpreted; interpretation also depends on the selected checkpoint, inference multiplicity, and the form of the optimization anchor. We use common-instance evaluation, independently recomputed float64 objectives, exact float64 Held–Karp anchors at TSP20, and certified float64 optimum intervals from exact scaled-integer Concorde solves at TSP50. Across both tested distributions, the selected local POMO model has the smaller gap at $n = 20$, whereas audited 2-opt has the smaller gap at $n = 50$. Because the blocks use separately trained models and different inference/search protocols, this observation is descriptive and does not support a causal scaling claim or a ranking of method families.

1 INTRODUCTION

A learned TSP result can appear different when compared with fixed-start nearest-neighbor construction or with an audited local-search control. The relative ordering of learned and classical methods alone does not indicate how close either result is to optimality. Exact anchors and certified optimum intervals calibrate the magnitude of these differences.

Across four separately controlled Euclidean TSP blocks spanning $n \in \{20, 50\}$ and uniform/clustered spatial generators, fixed-start nearest-neighbor construction lies roughly 17–25% above the exact or certified optimization anchor, whereas the specified local-search controls lie roughly 1.2–3.6% above it. The clustered replication follows a protocol fixed before the 50,000-step training runs, with held-out evaluation performed only after checkpoint identities were finalized.

Interpretation also requires explicit model and protocol context. The local AM and POMO models are trained separately for each size/distribution block, and inference and search protocols differ. We therefore report their observed orderings descriptively, with common-instance comparisons within each block and explicit anchor semantics.

Strong baselines and transparent protocols are established evaluation principles. The empirical contribution here is a controlled demonstration of the large NN–local-search separation, its pre-specified replication under a second spatial distribution, and a repeated ordering between learned and classical methods whose interpretation remains tied to the selected models and protocols.

Our contributions are:

- **Four-block empirical result.** A common-instance study across uniform and clustered TSP20/TSP50 blocks shows roughly 17–25% anchor-relative gaps for fixed-start nearest-neighbor construction and roughly 1.2–3.6% for the specified local-search controls.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

- **Pre-specified replication.** The clustered protocol and primary endpoint were fixed before the 50,000-step training runs, and the held-out evaluation sets were sampled only after checkpoint selection was complete; the primary NN–local-search separation replicated at both clustered sizes.
- **Protocol-specific ordering.** Under both tested distributions, the selected local POMO model has the smaller gap at $n = 20$, whereas 2-opt has the smaller gap at $n = 50$.
- **Supporting evaluation infrastructure.** Common-instance correctness checks, exact float64 Held–Karp anchors at TSP20, certified float64 optimum intervals from exact scaled-integer Concorde solves at TSP50, and explicit reporting of model provenance and protocol details make these empirical comparisons interpretable.

2 Related Work

The Attention Model (AM) and POMO are representative learned construction approaches for routing (Kool et al., 2019; Kwon et al., 2020). RL4CO provides a common implementation framework (Berto et al., 2025), while Jumanji provides environments and example agents used in accelerator-oriented reinforcement learning (Bonnet et al., 2024; InstaDeepAI, 2023). Our local AM and POMO models are controlled implementations rather than exact reproductions of the original training recipes.

Classical routing practice ranges from construction heuristics through move-based improvement (Croes, 1958; Lin, 1965; Lin & Kernighan, 1973) to strong heuristic and exact TSP software (Helsgaun, 2000; Applegate et al., 2006). Held–Karp subset dynamic programming supplies the exact TSP20 formulation (Held & Karp, 1962); Concorde supplies exact optimization of the explicit scaled-integer TSP50 objectives. We propagate quantization error back to the study's float64 objective instead of identifying the integer optimum with a float64 point optimum.

Evaluation guidance has emphasized protocol disclosure, reproducibility, and generalization boundaries (Accorsi et al., 2022; Joshi et al., 2022; Ma et al., 2025). Unify ML4TSP introduces ML4TSPBench, a modular framework for examining learning and search components under unified protocols (Li et al., 2025). Our study addresses a narrower, complementary question: how comparator choice changes the scale on which reported neural-TSP results are interpreted when methods are evaluated on common instances against exact or certified optimization anchors. We examine this question through the observed NN–local-search separation and a pre-specified second-distribution replication. Established evaluation principles supply the context for this empirical evidence over two sizes and two synthetic spatial distributions.

3 Problem, Distributions, and Anchor Semantics

For coordinates $x_i \in [0, 1]^2$, we evaluate a Hamiltonian cycle T by the closed-tour objective

$$L_{\text{float}}(T) = \sum_{(i,j)\in T} \sqrt{(x_{i0} - x_{j0})^2 + (x_{i1} - x_{j1})^2}, \tag{1}$$

using the stored coordinates and the common float64 evaluation convention.

Spatial distributions. The uniform blocks use independent coordinates from the unit square. The clustered blocks use a pre-specified RL4CO 0.6.0 three-cluster generator: three centers are sampled independently in $[0.2, 0.8)^2$, points are drawn with isotropic standard deviation 0.07, and coordinates are clipped to $[0, 1]^2$. Cluster sizes are $[7, 7, 6]$ at TSP20 and $[17, 17, 16]$ at TSP50; stored coordinates are float32 and are evaluated through the same float64 objective above. The clustered generator, seeds, endpoint, and analysis protocol were fixed before the 50,000-step training runs; held-out evaluation instances were sampled only after checkpoint selection was complete.

TSP20. Held–Karp subset dynamic programming returns the exact optimum of this stored-coordinate float64 objective for every instance (Held & Karp, 1962). "Exact" here means exact with respect to the common float64 distance matrix and closed-tour objective, not symbolic exact arithmetic.

TSP50. Concorde solves an explicit symmetric integer matrix exactly (Applegate et al., 2006). For scale $S = 10^6$, we define $w_{ij} = \lfloor Sd_{ij} + 0.5 \rfloor$, set $\widetilde{d}_{ij} = w_{ij}/S$, and calculate $\delta_{\max} = \max_{ij} |d_{ij} - \widetilde{d}_{ij}|$ for each instance. Since a TSP50 tour has 50 edges, $\epsilon = 50\delta_{\max}$ bounds the discrepancy for every tour. If $C^\star = W^\star/S$ is the exact scaled-integer optimum and T_q its returned tour, then

$$C^\star - \epsilon \le L^\star_{\text{float}} \le L_{\text{float}}(T_q) \le C^\star + \epsilon. \tag{2}$$

We therefore report $[LB, UB] = [C^\star - \epsilon, L_{\text{float}}(T_q)]$, not a float64 point optimum. For method length L_m, the certified relative-gap interval is

$$\left[100\frac{L_m - UB}{UB},\ 100\frac{L_m - LB}{LB}\right]. \tag{3}$$

Across the two TSP50 blocks, certificate uncertainty is numerically narrow but is always retained as an interval rather than collapsed to a point estimate.

4 Methods and Controls

Table 1 separates model and implementation provenance, distribution/size scope, selection, inference multiplicity, classical budget, and anchor role. Fixed-start nearest neighbor starts from city 0. Deterministic JAX 2-opt uses best improvement over all admissible edge pairs and a 100-sweep cap; every final tour passes an all-pairs local-optimum audit. JAX bounded sampled 3-opt draws 64 admissible triples per sweep for five sweeps under a fixed random protocol. It is bounded and non-exhaustive.

The bounded sampled 3-opt budget is deliberately unchanged between sizes: 64 sampled triples for five sweeps, corresponding to approximately 8.000% triple coverage at TSP20 and 0.371% at TSP50. This preserves the absolute control budget and avoids post-hoc tuning, but precludes interpreting the cross-size change as a pure scaling property of 3-opt. The 2-opt protocol retains its 100-sweep cap at both sizes; every reported final tour nevertheless passes the complete all-pairs local-optimum audit. Neither choice establishes compute fairness.

5 Learned Checkpoints and Training

The local AM and POMO models are trained for 50,000 optimization steps with three seeds for each size/distribution block. Checkpoints are selected independently by the highest mean reward on the fixed validation set, with ties broken in favor of the earlier checkpoint. The held-out evaluation sets were sampled only after checkpoint selection was complete. In the uniform blocks, the primary selected checkpoints are AM seed 1 at 40k and POMO seed 1 at 45k for TSP20, and AM seed 0 at 45k and POMO seed 0 at 50k for TSP50. In the clustered blocks, they are AM seed 1 and POMO seed 1 at 50k for TSP20, and AM seed 0 at 45k and POMO seed 0 at 50k for TSP50.

AM uses single-start greedy decoding without augmentation. POMO uses inherent multistart greedy decoding and maximum-reward selection: 20 starts at TSP20 and 50 starts at TSP50. The public Jumanji TSP-v1 A2C example model appears only in the uniform TSP20 block and uses greedy decoding. Although both local AM and POMO are trained for 50,000 optimizer steps, this does not imply matched training FLOPs, wall-clock compute, or inference effort. Inference multiplicity is not equalized. Separately trained models across sizes and distributions do not constitute a zero-shot generalization test.

6 Evaluation Protocol

The uniform TSP20 block contains 2,048 fresh instances; each of the uniform TSP50, clustered TSP20, and clustered TSP50 blocks contains 1,024 fresh instances. Within each block, all methods receive identical coordinate arrays and use the same stored-coordinate float64 objective. Instance identifiers and coordinate hashes persist through outputs. Learned tours are checked for permutation validity, invalid actions, finite rewards, and reward/objective agreement. Classical and anchor tours receive independent objective recomputation. All reported method–instance evaluations pass these checks.

Table 1: Method-level protocol and provenance shared across uniform and clustered blocks. Common instances and objectives apply within each block; training compute and inference effort are not matched across learned models. Primary checkpoint identities differ by block and are listed in the supplementary manifest.

Method / model	Provenance / training	Selection / multiplicity	Classical budget	Anchor / role
TSP20, uniform and clustered				
Held–Karp anchor	Study implementation; not trained	N/A	Exact subset DP	Exact float64 point; anchor
Selected local AM model	Local RL4CO; 50k, 3 seeds/block	Validation; 1 greedy	N/A	Held–Karp; secondary controlled model
Selected local POMO model	Local RL4CO; 50k, 3 seeds/block	Validation; 20 starts	N/A	Held–Karp; selected learned model
Jumanji A2C example	Public example; external training	Public checkpoint; 1 greedy	N/A	Uniform only; learned context
Fixed-start NN	Study implementation; not trained	N/A; 1 start	Start city 0	Held–Karp; weak construction
JAX 2-opt	Study implementation; not trained	N/A; 1 run	Best improvement, 100 sweeps	Held–Karp; local control
Bounded sampled 3-opt	Study implementation; not trained	N/A; 1 run	64 triples $\times$ 5 sweeps (8.000%)	Held–Karp; local control
TSP50, uniform and clustered				
Concorde certificate	Local Concorde build; research use; not trained	N/A	$S = 10^6$ integer matrix	Certified float64 interval; anchor
Selected local AM model	Local RL4CO; 50k, 3 seeds/block	Validation; 1 greedy	N/A	Certified interval; secondary controlled model
Selected local POMO model	Local RL4CO; 50k, 3 seeds/block	Validation; 50 starts	N/A	Certified interval; selected learned model
Fixed-start NN	Study implementation; not trained	N/A; 1 start	Start city 0	Certified interval; weak construction
JAX 2-opt	Study implementation; not trained	N/A; 1 run	Best improvement, 100 sweeps	Certified interval; local control
Bounded sampled 3-opt	Study implementation; not trained	N/A; 1 run	64 triples $\times$ 5 sweeps (0.371%)	Certified interval; local control

We report each block separately. Means, medians, and anchor-relative gaps are never pooled across sizes or distributions. Descriptive paired intervals use 5,000 deterministic instance-level resamples within a block and are not formal significance tests. The uniform TSP50 paired analysis uses base seed 20260924; the clustered paired analysis uses base seed 20260926. Both use tie tolerance 10^{-9}. Detailed paired results are supplementary because the main analysis emphasizes the anchor-calibrated quality scale rather than method rankings.

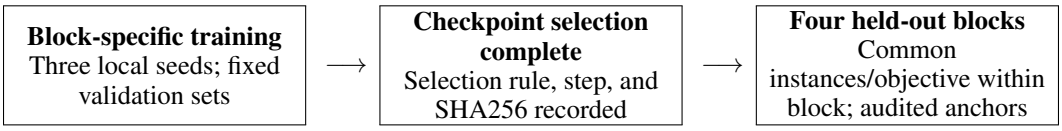


Figure 1: Evaluation sequence applied independently to each size/distribution block. The held-out evaluation sets are sampled only after checkpoint selection is complete. Common instances within a block do not imply matched training compute or inference effort.

7 RESULTS

7.1 FOUR-BLOCK SEPARATION AND CLUSTERED REPLICATION

The primary finding is the separation between fixed-start nearest-neighbor construction and the specified local-search controls across all four blocks: roughly 17–25% versus 1.2–3.6% above the exact or certified optimization anchor. Figure 2 displays this contrast for 2-opt alongside local POMO; the detailed tables retain bounded sampled 3-opt and local AM.

The clustered replication reproduces the primary NN–local-search separation at both TSP20 and TSP50. Its protocol and primary endpoint were fixed before the 50,000-step training runs, and the held-out evaluation sets were sampled only after checkpoint selection was complete, making the second-distribution result a pre-specified replication within the tested synthetic setting.

Anchor-calibrated interpretation. The exact float64 Held–Karp anchors at TSP20 and certified float64 optimum intervals from exact scaled-integer Concorde solves at TSP50 locate the specified comparators on an optimization-quality scale within each block. The large separation changes the reference level supplied by the classical control when interpreting a learned result. Common-instance evaluation and independently recomputed objectives support this comparison; the learned models and protocols remain distinct across the four blocks.

7.2 PROTOCOL-SPECIFIC LEARNED/CLASSICAL ORDERING

Across both tested spatial distributions, the selected local POMO model has the smaller gap at $n = 20$, whereas 2-opt has the smaller gap at $n = 50$. This repeated ordering is a secondary descriptive observation. The blocks use separately trained models, POMO uses 20 versus 50 starts, and bounded sampled 3-opt retains a fixed absolute search budget with different relative coverage. These differences preclude interpreting the ordering as a causal size effect or as a ranking of method families.

7.3 DETAILED BLOCK RESULTS AND MODEL CONTEXT

Uniform blocks. The uniform TSP20 block places fixed-start nearest neighbor at 16.9072% mean relative gap, compared with 1.4851% for JAX 2-opt and 1.3712% for bounded sampled 3-opt. The selected local POMO and AM models lie at 0.7767% and 7.2787%; the public A2C example model, available only in this block, lies at 1.7992%. In uniform TSP50, nearest neighbor has a mean certified gap of [22.7846%, 22.7852%], while 2-opt and bounded sampled 3-opt have [3.5561%, 3.5565%] and [3.5511%, 3.5516%]. The selected local POMO and AM models have [4.6413%, 4.6417%] and [21.6243%, 21.6248%]. These results describe the selected checkpoints rather than the performance of the corresponding method families.

Clustered blocks. The clustered protocol and primary endpoint were fixed before the 50,000-step training runs, and the held-out evaluation sets were sampled only after checkpoint selection was complete. At TSP20, fixed-start nearest neighbor lies at 17.8436% relative gap to the exact anchor, while 2-opt and bounded sampled 3-opt lie at 1.2656% and 1.1683%. The selected local POMO and AM models lie at 1.0596% and 10.9629%. At TSP50, nearest neighbor has a certified gap of [25.0155%, 25.0167%], versus [3.2141%, 3.2151%] for 2-opt and [3.2043%, 3.2053%] for bounded sampled 3-opt. The selected local POMO and AM models have [7.1463%, 7.1474%]

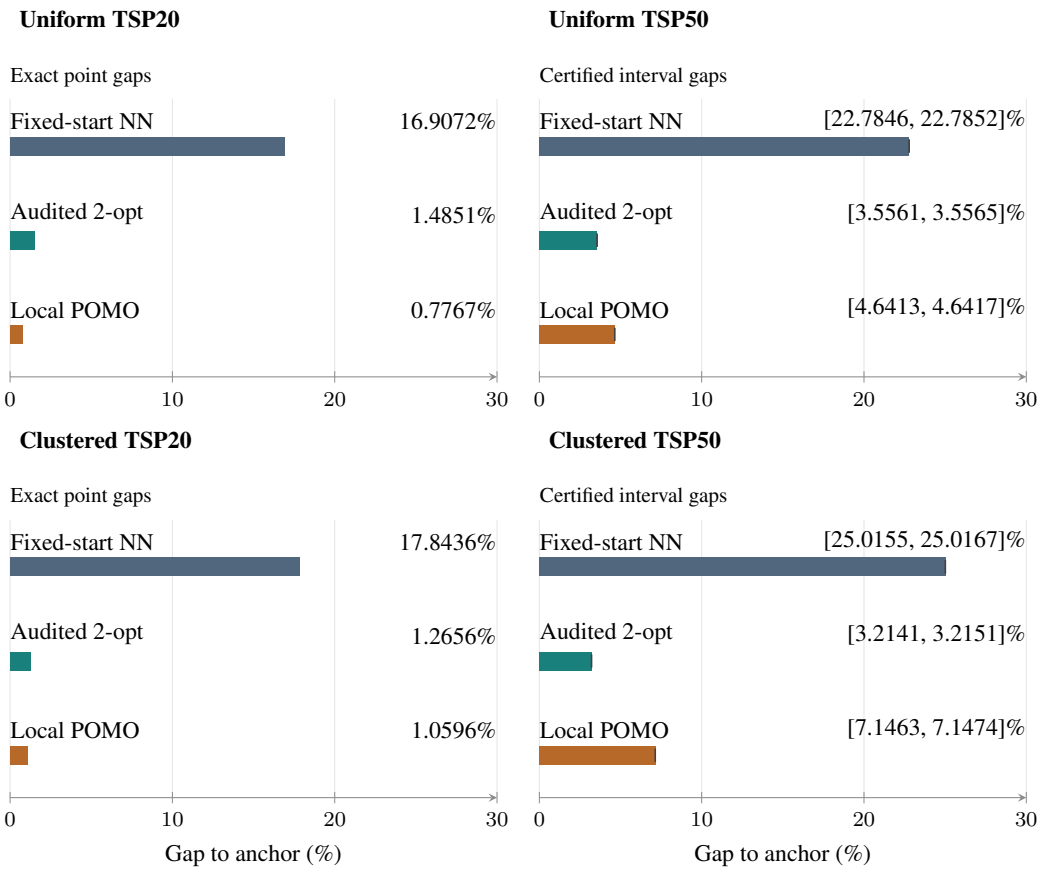


Figure 2: Comparator strength and the anchor-relative evaluation scale in four separately controlled blocks: fixed-start NN is substantially farther from the anchor than 2-opt. TSP20 uses exact float64 Held–Karp anchors; TSP50 gaps use certified float64 optimum intervals derived from exact scaled-integer Concorde solves. Local POMO models were separately trained and validation-selected within each size/distribution block. The shared axis compares the observed gap scales; the panels do not establish a causal scaling trajectory or family-level ranking. TSP50 bar lengths show interval midpoints for display, while bracketed labels give certified gap endpoints. Local AM and bounded sampled 3-opt are omitted from this view for clarity and remain in Table 2 and the detailed appendix tables.

and [30.4763%, 30.4775%]. The local AM results describe only these specific controlled training configurations.

Local AM is retained as a secondary controlled model with its checkpoint and training configuration reported explicitly. Detailed values and paired comparisons, together with variation across three seeds for the learned models, remain in the appendix.

8 Discussion and Limitations

Empirical interpretation. The direct result is a substantial separation between the anchor-relative gaps of fixed-start construction and the specified local-search controls in all four tested blocks. The pre-specified clustered replication reproduces this separation at both sizes. Thus, comparator choice changes the reference level supplied by the classical control when interpreting a learned result. The secondary observation is that local POMO has the smaller gap at $n = 20$ and 2-opt at $n = 50$ under both tested distributions. This protocol-specific ordering does not establish a size-caused crossover, show that local search scales better, demonstrate universal generalization across distributions, or support a method-family ranking.

Scope and protocol limits. The study spans two problem sizes and two synthetic Euclidean spatial generator families. This provides size and distribution breadth within the tested synthetic setting, but does not establish universal distributional validity, real-world routing generalization, or a zero-shot distribution-shift result. Learned models are trained independently for every size/distribution block. The specific local AM configurations and 50k training protocols are not reproductions of the original Attention Model training setup; their observed gaps cannot identify an architectural effect. In clustered TSP50 the AM mean spread across three seeds is 0.140510, compared with 0.000708 for POMO, but these selected local models do not characterize either family. POMO inference naturally uses 20 starts at TSP20 and 50 at TSP50, while local AM is single-start, so learned inference effort is unmatched.

Classical budgets are also protocol choices. Bounded sampled 3-opt uses 64 triples for five sweeps at both sizes, reducing approximate triple coverage from 8.000% to 0.371%. This preserves the absolute control budget and avoids post-hoc tuning, but precludes interpreting the cross-size pattern as a pure scaling property of 3-opt. The fixed 100-sweep 2-opt cap may have size-dependent cost implications, although final all-pairs audits pass. Anchor semantics differ: TSP20 has an exact float64 point optimum, whereas TSP50 has a certified float64 interval derived from exact scaled-integer optimization. A2C appears only in uniform TSP20. There is no real-world distribution, learned CVRP replication, broader learned-comparator suite, or formal compute-normalized comparison. The study is therefore a scoped empirical evaluation rather than a comprehensive neural combinatorial optimization benchmark and supports no broad neural/classical superiority claim.

9 Conclusion

Across four controlled TSP blocks spanning two sizes and two pre-specified synthetic Euclidean spatial distributions, fixed-start nearest-neighbor construction remained roughly 17–25% above the exact or certified optimization anchor, while the specified local-search controls were roughly 1.2–3.6% above it. The pre-specified clustered replication reproduced this separation at both sizes. This empirical finding shows how comparator choice materially changes the scale on which learned routing results are interpreted. Model, protocol, and anchor context make the comparison meaningful; family labels alone do not determine its interpretation. The selected local POMO model has the smaller gap at $n = 20$ and 2-opt at $n = 50$ under both tested distributions. This ordering remains specific to the selected models and protocols, and the evidence covers only the two tested sizes and two synthetic Euclidean generators.

10 Reproducibility

The study provides strong provenance and auditability, although the release is not a fully self-contained end-to-end reproduction package. Uniform-block records include held-out coordinate hashes, learned checkpoint hashes, fixed validation selections, anchor implementations, and deterministic bootstrap settings and seeds. Clustered-block records add the generator ID and configuration hash, fixed validation hashes, all twelve selected-checkpoint identities, held-out-set hashes, a record that the clustered distribution was fixed in advance, exact or certified anchor outputs, and the deterministic bootstrap configuration. The clustered endpoint and protocol were fixed before the 50,000-step training runs, and the held-out evaluation sets were sampled only after checkpoint selection was complete. This was an internal pre-specification rather than a preregistration in an external registry. The TSP50 anchor executable and QSopt dependency remain local research-only under the documented licensing constraints and are not redistributed. The clean manuscript archive contains source, tables, figures, references, and a supplementary artifact manifest, but no checkpoints, raw per-instance data, logs, environments, QSopt files, Concorde source, or linked executable. Checkpoints and per-instance evaluated rows are planned components of a future research artifact, subject to repository and dependency licenses; they are not bundled here.

AI Use Statement. Generative AI tools were used to assist with language editing and limited code development and debugging. All AI-assisted code was reviewed, tested, and validated by the author. The author takes full responsibility for the final content of this work.

REFERENCES

Luca Accorsi, Andrea Lodi, and Daniele Vigo. Guidelines for the computational testing of machine learning approaches to vehicle routing problems. *Operations Research Letters*, 50(2):229–234, 2022. doi: 10.1016/j.orl.2022.01.018. URL https://doi.org/10.1016/j.orl.2022.01.018.

David L. Applegate, Robert E. Bixby, Vašek Chvátal, and William J. Cook. *The Traveling Salesman Problem: A Computational Study*. Princeton Series in Applied Mathematics. Princeton University Press, 2006.

Federico Berto, Chuanbo Hua, Junyoung Park, Laurin Luttmann, Yining Ma, Fanchen Bu, Jiarui Wang, Haoran Ye, Minsu Kim, Sanghyeok Choi, Nayeli Gast Zepeda, Andre Hottung, Jianan Zhou, Jieyi Bi, Yu Hu, Fei Liu, Hyeonah Kim, Jiwoo Son, Haeyeon Kim, Davide Angioni, Wouter Kool, Zhiguang Cao, Qingfu Zhang, Joungho Kim, Jie Zhang, Kijung Shin, Cathy Wu, Sungsoo Ahn, Guojie Song, Changhyun Kwon, Kevin Tierney, Lin Xie, and Jinkyoo Park. RL4CO: An extensive reinforcement learning for combinatorial optimization benchmark. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2*, pp. 5278–5289. ACM, 2025. doi: 10.1145/3711896.3737433. URL https://doi.org/10.1145/3711896.3737433.

Clement Bonnet, Daniel Luo, Donal Byrne, Shikha Surana, Sasha Abramowitz, Paul Duckworth, Vincent Coyette, Laurence I. Midgley, Elshadai Tegegn, Tristan Kalloniatis, Omayma Mahjoub, Matthew Macfarlane, Andries P. Smit, Nathan Grinsztajn, Raphael Boige, Cemlyn N. Waters, Mohamed A. Mimouni, Ulrich A. Mbou Sob, Ruan de Kock, Siddarth Singh, Daniel Furelos-Blanco, Victor Le, Arnu Pretorius, and Alexandre Laterre. Jumanji: A diverse suite of scalable reinforcement learning environments in JAX. In *International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=C4CxQmp9wc.

G. A. Croes. A method for solving traveling-salesman problems. *Operations Research*, 6(6):791–812, 1958. doi: 10.1287/opre.6.6.791. URL https://doi.org/10.1287/opre.6.6.791.

Michael Held and Richard M. Karp. A dynamic programming approach to sequencing problems. *Journal of the Society for Industrial and Applied Mathematics*, 10(1):196–210, 1962. doi: 10.1137/0110015. URL https://doi.org/10.1137/0110015.

Keld Helsgaun. An effective implementation of the Lin–Kernighan traveling salesman heuristic. *European Journal of Operational Research*, 126(1):106–130, 2000. doi: 10.1016/S0377-2217(99)00284-2. URL https://doi.org/10.1016/S0377-2217(99)00284-2.

InstaDeepAI. Jumanji benchmark A2C TSP-v1. Hugging Face model repository, 2023. URL https://huggingface.co/InstaDeepAI/jumanji-benchmark-a2c-TSP-v1. Apache-2.0 metadata; accessed 2026-09-12.

Chaitanya K. Joshi, Quentin Cappart, Louis-Martin Rousseau, and Thomas Laurent. Learning the travelling salesperson problem requires rethinking generalization. *Constraints*, 27:70–98, 2022. doi: 10.1007/s10601-022-09327-y. URL https://doi.org/10.1007/s10601-022-09327-y.

Wouter Kool, Herke van Hoof, and Max Welling. Attention, learn to solve routing problems! In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=ByxBFsRqYm.

Yeong-Dae Kwon, Jinho Choo, Byoungjip Kim, Iljoo Yoon, Youngjune Gwon, and Seungjai Min. POMO: Policy optimization with multiple optima for reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 33, pp. 21188–21198, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/f231f2107df69eab0a3862d50018a9b2-Abstract.html.

Yang Li, Jiale Ma, Wenzheng Pan, Runzhong Wang, Haoyu Geng, Nianzu Yang, and Junchi Yan. Unify ML4TSP: Drawing methodological principles for TSP and beyond from streamlined design space of learning and search. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu

(eds.), *International Conference on Learning Representations*, volume 2025, pp. 65269–65305, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/a3dc6e903082902d8e916bb9fccbfcbc-Paper-Conference.pdf.

S. Lin and B. W. Kernighan. An effective heuristic algorithm for the traveling-salesman problem. *Operations Research*, 21(2):498–516, 1973. doi: 10.1287/opre.21.2.498. URL https://doi.org/10.1287/opre.21.2.498.

Shen Lin. Computer solutions of the traveling salesman problem. *Bell System Technical Journal*, 44 (10):2245–2269, 1965. doi: 10.1002/j.1538-7305.1965.tb04146.x. URL https://doi.org/10.1002/j.1538-7305.1965.tb04146.x.

Jiale Ma, Wenzheng Pan, Yang Li, and Junchi Yan. ML4CO-Bench-101: Benchmark machine learning for classic combinatorial problems on graphs. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.52202/085713-3921. URL https://doi.org/10.52202/085713-3921. Datasets and Benchmarks Track.

A Controlled Training Details

Within each size/distribution block, local AM and POMO use three seeds, 50,000 optimizer steps, batch size 64, and checkpoints and validation every 5,000 steps. Checkpoints are selected by the highest mean reward on the fixed validation set, with ties broken in favor of the earlier checkpoint. Every uniform and clustered validation set contains 512 fixed instances. The held-out evaluation sets were sampled only after selected checkpoint identities and SHA256 values were fixed. The models are size- and distribution-matched and trained independently; this is not one model evaluated zero-shot across sizes or distributions.

B Anchor Implementations

B.1 TSP20 Exact Float64 Held–Karp Anchor

Following Held & Karp (1962), the subset dynamic program stores the minimum path cost $D(S, j)$ from city 0 through subset S ending at j and closes the cycle after all non-start cities are visited. It uses the common float64 distance matrix, deterministic predecessor handling, and no approximate pruning. Brute-force preflight and independent tour recomputation passed. All 2,048 uniform and all 1,024 clustered TSP20 held-out instances received valid exact float64 point optima under the stored-coordinate convention; the blocks remain separate.

B.2 TSP50 Certified Float64 Interval

For each instance, integer weights are $w_{ij} = \lfloor 10^6 d_{ij} + 0.5 \rfloor$. Concorde solves this explicit integer objective exactly (Applegate et al., 2006). Per-edge quantization error yields the deterministic interval in Section 3. All 1,024 uniform and all 1,024 clustered exact integer solves and certificates passed. Uniform mean lower and upper bounds are 5.687865479 and 5.687890494; clustered means are 2.654208947 and 2.654233925. The corresponding mean interval widths are approximately 2.50×10^{-5} in both blocks. These remain separate certified intervals and are never replaced by a float64 point estimate.

C Four-Block and Clustered Evidence

The four evaluation blocks are analyzed separately and are never pooled across size or distribution. TSP20 uses exact float64 Held–Karp point optima. TSP50 uses certified float64 optimum intervals propagated from exact scaled-integer Concorde solves.

The clustered TSP50 AM mean spread is 0.140510 and the POMO mean spread is 0.000708. These are descriptions of variation across three seeds for the selected local models and do not characterize either model family. The clustered paired analysis uses 5,000 deterministic instance-level resam-

Table 2: Four-block summary under the pre-specified evaluation protocol. Anchor values are shown first; remaining rows report point gaps to exact float64 optima at TSP20 and certified relative-gap intervals at TSP50. Blocks are not pooled.

TSP20 role	Uniform	Clustered
Anchor value	3.837092	1.813967
Selected local AM model	7.2787%	10.9629%
Selected local POMO model	0.7767%	1.0596%
A2C example model	1.7992%	–
Fixed-start NN	16.9072%	17.8436%
JAX 2-opt	1.4851%	1.2656%
Bounded sampled 3-opt	1.3712%	1.1683%
TSP50 role	**Uniform**	**Clustered**
Anchor value	[5.687865, 5.687890]	[2.654209, 2.654234]
Selected local AM model	[21.6243%, 21.6248%]	[30.4763%, 30.4775%]
Selected local POMO model	[4.6413%, 4.6417%]	[7.1463%, 7.1474%]
A2C example model	–	–
Fixed-start NN	[22.7846%, 22.7852%]	[25.0155%, 25.0167%]
JAX 2-opt	[3.5561%, 3.5565%]	[3.2141%, 3.2151%]
Bounded sampled 3-opt	[3.5511%, 3.5516%]	[3.2043%, 3.2053%]

Table 3: Detailed clustered primary quality results. All non-anchor rows have validity and objective-alignment rate 1.000. In the TSP50 anchor row, the repeated interval gives objective bounds, not a relative gap.

n	Artifact/control	Mean length or anchor interval	Anchor-relative gap	Held-out instances
20	Anchor	1.813967	0.0000%	1024
20	Primary local AM	2.011807	10.9629%	1024
20	Primary local POMO	1.833264	1.0596%	1024
20	Fixed-start NN	2.136542	17.8436%	1024
20	JAX 2-opt	1.836558	1.2656%	1024
20	Bounded sampled 3-opt	1.834891	1.1683%	1024
50	Anchor	[2.654209, 2.654234]	[2.654209, 2.654234]	1024
50	Primary local AM	3.461267	[30.4763%, 30.4775%]	1024
50	Primary local POMO	2.844036	[7.1463%, 7.1474%]	1024
50	Fixed-start NN	3.316574	[25.0155%, 25.0167%]	1024
50	JAX 2-opt	2.739098	[3.2141%, 3.2151%]	1024
50	Bounded sampled 3-opt	2.738836	[3.2043%, 3.2053%]	1024

ples, base seed 20260926, and tie tolerance 10^{-9}. These intervals are descriptive rather than formal significance tests. Uniform detailed tables remain in the subsequent appendices.

Table 4: Clustered learned-model variation across three seeds. Primary identities were fixed by validation before the held-out evaluation sets were sampled.

n	Family	Seed / selected checkpoint	Mean length	Median length	Anchor-relative gap	Primary
20	AM	AM seed 0, step 45k	2.012109	2.010471	10.9847%	no
20	AM	AM seed 1, step 50k	2.011807	2.019427	10.9629%	yes
20	AM	AM seed 2, step 45k	2.012907	2.018337	10.9401%	no
20	POMO	POMO seed 0, step 45k	1.832191	1.830411	0.9934%	no
20	POMO	POMO seed 1, step 50k	1.833264	1.831525	1.0596%	yes
20	POMO	POMO seed 2, step 50k	1.833002	1.831301	1.0378%	no
50	AM	AM seed 0, step 45k	3.461267	3.465017	[30.4763%, 30.4775%]	yes
50	AM	AM seed 1, step 50k	3.601777	3.592910	[35.7710%, 35.7723%]	no
50	AM	AM seed 2, step 40k	3.514997	3.511528	[32.4299%, 32.4312%]	no
50	POMO	POMO seed 0, step 50k	2.844036	2.850571	[7.1463%, 7.1474%]	yes
50	POMO	POMO seed 1, step 50k	2.844169	2.849821	[7.1481%, 7.1491%]	no
50	POMO	POMO seed 2, step 50k	2.844744	2.858643	[7.1633%, 7.1644%]	no

Table 5: Selected clustered descriptive paired differences (method A minus method B). Intervals use 5,000 deterministic paired resamples and are not formal significance tests.

n	Comparison	Mean difference	95% interval	A shorter	Tie / B shorter
20	Primary local AM − primary local POMO	+0.178544	[+0.170897, +0.185991]	17	10 / 997
20	Primary local POMO − fixed-start NN	-0.303278	[-0.313208, -0.293292]	1010	1 / 13
20	Primary local POMO − JAX 2-opt	-0.003294	[-0.005687, -0.000944]	375	178 / 471
20	Primary local POMO − bounded sampled 3-opt	-0.001627	[-0.003985, +0.000678]	359	184 / 481
20	JAX 2-opt − fixed-start NN	-0.299983	[-0.310391, -0.289686]	1012	12 / 0
20	Bounded sampled 3-opt − fixed-start NN	-0.301650	[-0.311674, -0.291415]	1012	12 / 0
50	Primary local AM − primary local POMO	+0.617231	[+0.601897, +0.632266]	0	0 / 1024
50	Primary local POMO − fixed-start NN	-0.472538	[-0.487751, -0.457630]	1009	0 / 15
50	Primary local POMO − JAX 2-opt	+0.104938	[+0.098870, +0.110782]	135	0 / 889
50	Primary local POMO − bounded sampled 3-opt	+0.105200	[+0.099191, +0.111127]	135	0 / 889
50	JAX 2-opt − fixed-start NN	-0.577476	[-0.591567, -0.562890]	1024	0 / 0
50	Bounded sampled 3-opt − fixed-start NN	-0.577738	[-0.592329, -0.562756]	1024	0 / 0

D UNIFORM LEARNED VARIATION ACROSS THREE SEEDS

The detailed uniform seed results are secondary to the validation-based primary checkpoint selection, which was completed before the held-out evaluation sets were sampled. The clustered three-seed variation table appears in the preceding four-block appendix.

Table 6: Variation across three seeds for locally trained checkpoints on the fresh held-out evaluation set. Selection labels refer to validation, not held-out evaluation results.

Checkpoint	Mean	Median	Mean relative gap (%)	Exact hits
AM seed 0	4.129765	4.126616	7.6185	60
AM seed 1 (selected)	4.116253	4.114234	7.2787	73
AM seed 2	4.130785	4.130344	7.6762	76
POMO seed 0	3.868908	3.884486	0.8205	846
POMO seed 1 (selected)	3.867252	3.880819	0.7767	881
POMO seed 2	3.867388	3.881221	0.7802	872

Table 7: TSP50 variation across three seeds. Primary checkpoints remain the validation-selected AM seed 0 and POMO seed 0; held-out evaluation was performed only after selection.

Checkpoint	Mean	Median	Mean certified gap (%)
AM selected seed 0	6.915965	6.921938	[21.6243, 21.6248]
AM selected seed 1	6.972276	6.959699	[22.5887, 22.5892]
AM selected seed 2	6.937367	6.941405	[21.9842, 21.9847]
POMO selected seed 0	5.951547	5.953725	[4.6413, 4.6417]
POMO selected seed 1	5.965124	5.965972	[4.8800, 4.8804]
POMO selected seed 2	5.956923	5.971256	[4.7321, 4.7325]

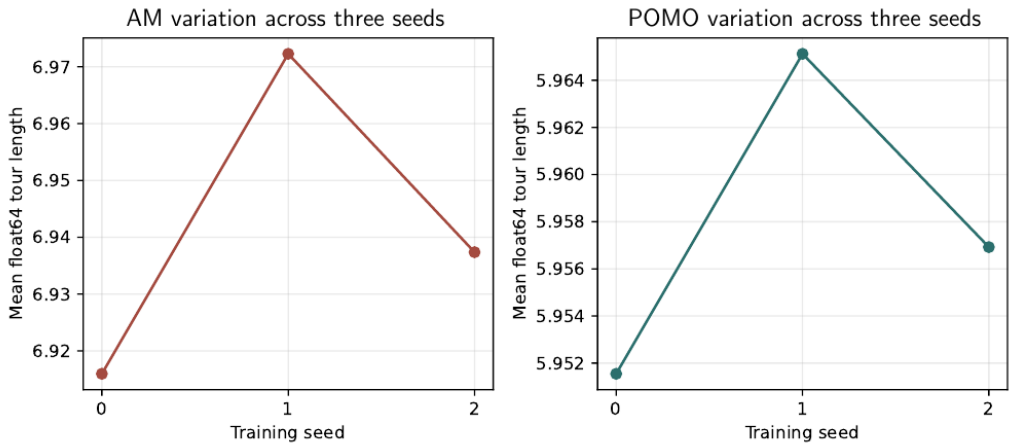


Figure 3: TSP50 AM and POMO variation across three seeds on held-out instances. Primary checkpoints were fixed by validation before the held-out evaluation sets were sampled.

The TSP50 mean spread is 0.056311 for AM and 0.013576 for POMO. We report this as variation across three seeds, not as a universal robustness result.

E Uniform TSP20 Selected Descriptive Paired Comparisons

The selected contrasts are supplementary to the exact-gap scale. They use method A minus method B and are descriptive summaries of the same coordinate-indexed instances, not formal hypothesis tests.

Table 8: Selected descriptive paired differences on the same 2,048 instances. Each value is method A minus method B; negative values indicate shorter tours for A. Intervals are paired-bootstrap summaries, not formal significance tests.

Comparison	Mean difference	95% bootstrap interval
AM selected − A2C example-agent	+0.209776	[+0.200586, +0.219108]
AM selected − JAX 2-opt	+0.221412	[+0.212314, +0.230935]
AM selected − bounded sampled 3-opt	+0.225893	[+0.216533, +0.235220]
POMO selected − A2C example-agent	-0.039225	[-0.043331, -0.035196]
JAX 2-opt − A2C example-agent	-0.011636	[-0.017241, -0.006093]
Bounded sampled 3-opt − A2C example-agent	-0.016117	[-0.021684, -0.010703]

F UNIFORM TSP20 FULL DESCRIPTIVE PAIRED COMPARISONS

Table 9: Full descriptive paired summary. Differences are A minus B; W/T/L counts are from A's perspective. Intervals summarize 5,000 paired instance resamples and are not formal significance tests.

Comparison	Mean difference	95% interval	W / T / L
A2C example-agent − Exact Held–Karp	+0.069385	[+0.065180, +0.073660]	0 / 512 / 1536
AM seed 1 − Bounded sampled 3-opt	+0.225893	[+0.216533, +0.235220]	172 / 57 / 1819
AM seed 1 − JAX 2-opt	+0.221412	[+0.212314, +0.230935]	188 / 56 / 1804
AM seed 1 − Fixed-start NN	-0.370566	[-0.388942, -0.351273]	1632 / 2 / 414
AM seed 1 − POMO seed 1	+0.249001	[+0.240636, +0.257818]	32 / 95 / 1921
AM seed 0 − Exact Held–Karp	+0.292673	[+0.283926, +0.301794]	0 / 60 / 1988
AM seed 0 − A2C example-agent	+0.223288	[+0.214083, +0.232623]	196 / 53 / 1799
AM seed 0 − POMO seed 0	+0.260857	[+0.252605, +0.269081]	37 / 66 / 1945
AM seed 0 − AM seed 1	+0.013512	[+0.002272, +0.024800]	973 / 20 / 1055
AM seed 0 − AM seed 2	-0.001019	[-0.012408, +0.010494]	1029 / 14 / 1005
AM seed 1 − Exact Held–Karp	+0.279161	[+0.270330, +0.287960]	0 / 73 / 1975
AM seed 1 − A2C example-agent	+0.209776	[+0.200586, +0.219108]	219 / 71 / 1758
AM seed 2 − Exact Held–Karp	+0.293693	[+0.284421, +0.302743]	0 / 76 / 1972
AM seed 2 − A2C example-agent	+0.224307	[+0.215103, +0.233924]	215 / 68 / 1765
AM seed 2 − POMO seed 2	+0.263397	[+0.254696, +0.272154]	22 / 93 / 1933
AM seed 2 − AM seed 1	+0.014531	[+0.003149, +0.026024]	951 / 71 / 1026
Fixed-start NN − Exact Held–Karp	+0.649728	[+0.633446, +0.666778]	0 / 19 / 2029

Table 9 (continued).

Comparison	Mean difference	95% interval	W / T / L
JAX 2-opt − A2C example-agent	-0.011636	[-0.017241, -0.006093]	1077 / 334 / 637
JAX 2-opt − Exact Held–Karp	+0.057749	[+0.053872, +0.061725]	0 / 955 / 1093
Bounded sampled 3-opt − A2C example-agent	-0.016117	[-0.021684, -0.010703]	1097 / 344 / 607
Bounded sampled 3-opt − Exact Held–Karp	+0.053268	[+0.049572, +0.056822]	0 / 978 / 1070
Fixed-start NN − A2C example-agent	+0.580342	[+0.563281, +0.597590]	72 / 11 / 1965
POMO seed 1 − Bounded sampled 3-opt	-0.023108	[-0.026978, -0.019173]	811 / 557 / 680
POMO seed 1 − JAX 2-opt	-0.027588	[-0.031859, -0.023508]	842 / 543 / 663
POMO seed 1 − Fixed-start NN	-0.619567	[-0.636358, -0.602676]	2009 / 13 / 26
POMO seed 0 − Exact Held–Karp	+0.031816	[+0.029770, +0.033921]	0 / 846 / 1202
POMO seed 0 − A2C example-agent	-0.037569	[-0.041888, -0.033285]	1157 / 451 / 440
POMO seed 0 − POMO seed 1	+0.001656	[-0.000192, +0.003541]	613 / 761 / 674
POMO seed 0 − POMO seed 2	+0.001520	[-0.000473, +0.003518]	621 / 753 / 674
POMO seed 1 − Exact Held–Karp	+0.030160	[+0.028136, +0.032146]	0 / 881 / 1167
POMO seed 1 − A2C example-agent	-0.039225	[-0.043331, -0.035196]	1174 / 483 / 391
POMO seed 2 − Exact Held–Karp	+0.030296	[+0.028359, +0.032306]	0 / 872 / 1176
POMO seed 2 − A2C example-agent	-0.039090	[-0.043401, -0.035153]	1177 / 475 / 396
POMO seed 2 − POMO seed 1	+0.000135	[-0.001369, +0.001691]	457 / 1117 / 474

G UNIFORM TSP50 DESCRIPTIVE PAIRED COMPARISONS

Table 10: TSP50 descriptive paired differences. Values are A minus B; negative means A is shorter. Intervals use 5,000 deterministic paired resamples. W/T/L refers to A.

Comparison	Mean difference	95% descriptive interval	W/T/L
Primary local AM − primary local POMO	0.964418	[0.940320, 0.987272]	1/0/1023
Primary local AM − fixed-start NN	-0.068671	[-0.104558, -0.032831]	558/0/466
Primary local AM − JAX 2-opt	1.026157	[1.001818, 1.051168]	2/0/1022
Primary local AM − bounded sampled 3-opt	1.026434	[1.002068, 1.051334]	2/0/1022
Primary local POMO − fixed-start NN	-1.033089	[-1.062505, -1.004056]	1013/0/11
Primary local POMO − JAX 2-opt	0.061739	[0.049920, 0.073532]	350/0/674
Primary local POMO − bounded sampled 3-opt	0.062016	[0.050485, 0.074019]	350/0/674
JAX 2-opt − fixed-start NN	-1.094828	[-1.123428, -1.066132]	1024/0/0
Bounded sampled 3-opt − fixed-start NN	-1.095105	[-1.122816, -1.067940]	1024/0/0
JAX 2-opt − bounded sampled 3-opt	0.000277	[0.000033, 0.000589]	0/1018/6
AM seed 0 − AM seed 1	-0.056311	[-0.086397, -0.026586]	562/0/462
AM seed 0 − AM seed 2	-0.021401	[-0.047924, 0.005080]	525/0/499
AM seed 1 − AM seed 2	0.034909	[0.004529, 0.062940]	475/0/549
POMO seed 0 − POMO seed 1	-0.013576	[-0.022681, -0.004393]	557/1/466
POMO seed 0 − POMO seed 2	-0.005375	[-0.012247, 0.001686]	510/8/506
POMO seed 1 − POMO seed 2	0.008201	[-0.000628, 0.017110]	483/3/538

We report 16 paired comparisons. Intervals use 5,000 deterministic paired resamples with base seed 20260924 and tie tolerance 10^{-9}. They are descriptive summaries of this held-out evaluation block, not formal significance tests.

H CORRECTNESS AND OBJECTIVE AUDITS

Every reported uniform and clustered learned, classical, and anchor output is a valid permutation and agrees with independent objective recomputation; learned invalid-action counts are zero and every 2-opt output passes the all-pairs local-optimum audit. For the clustered blocks specifically, all 10,240 TSP20 and 9,216 TSP50 reported method-instance evaluations pass these checks, including 12,288 selected learned tours and 6,144 classical outputs. All 61,440 TSP20 and 153,600 TSP50 clustered POMO inherent-start tours are valid and aligned. All 1,024 clustered TSP20 exact anchors and all 1,024 clustered TSP50 certificates pass.