

Bandwidth, Representation Loss, and Temporal Drift in Quantum Kernel Fraud Detection: A Pre-Specified Falsification Study

Mohamed Faisal Sindhi

sindhi@usf.edu

University of South Florida <https://orcid.org/0009-0000-1348-0566>

Hammaz Buksh

University of South Florida

Research Article

Keywords: quantum kernels, fraud detection, temporal distribution shift, kernel bandwidth, representation loss, falsification

Posted Date: October 13th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11322513/v1>

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

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Bandwidth, Representation Loss, and Temporal Drift in Quantum Kernel Fraud Detection: A Pre-Specified Falsification Study

Mohamed Faisal Sindhi and Hammaz Buksh

School of Information Systems

University of South Florida

Tampa, Florida, USA

Corresponding author: sindhi@usf.edu

Abstract

Quantum kernel results are often reported without holding the input representation, training sample, tuning budget and temporal evaluation protocol fixed against a matched classical kernel. This pre-specified, protocol-locked falsification study uses the Elliptic++ transaction tables (203,769 transactions, 49 time steps) under a locked temporal split (training steps 1–24, validation 25–29, test 30–49 in 4 locked windows). The primary metric is pooled Recall@5%. A frozen full-dimensional gradient-boosting model reached 0.381, and every model, including this one, collapsed in the final window. Graph structure and selective reranking were tested first and falsified. Compressing 182 features to 4, 6 and 8 principal components retained 34%, 43% and 51% of the variance and 64%, 50% and 33% of the full-dimensional model’s pooled Recall@5%. In the primary, default-bandwidth analysis, an entangling fidelity quantum kernel was below the strongest matched classical kernel at 4, 6 and 8 qubits (pooled Recall@5% differences -0.078 , -0.203 and -0.146). In a secondary, post hoc bandwidth correction, explicit scale tuning raised the entangling kernel’s Recall@5% by 0.070, 0.121 and 0.109; the corrected kernel remained below the strongest matched classical kernel at 6 and 8 qubits and was statistically unresolved at 4 qubits. The selected scale lay at the lower boundary of the tested grid, and the apparent effect of entanglement depended on bandwidth. No repeatable predictive benefit of the quantum kernel over matched classical kernels was observed in the tested setting; the findings do not extend beyond this dataset, protocol, representation, feature-map family, sample size and bandwidth range.

Keywords: quantum kernels; fraud detection; temporal distribution shift; kernel bandwidth; representation loss; falsification

1 Introduction

Claims that quantum kernels compete with classical kernels depend on choices that are easy to leave unmatched: the compressed input representation, the training sample, the tuning budget, the kernel bandwidth, and, for financial data, the evaluation protocol under temporal distribution shift. Benchmarking practice for quantum machine learning has been questioned on exactly such grounds [1], and kernel bandwidth and kernel concentration are known determinants of quantum-kernel behaviour [2–4].

This study asks whether quantum feature-space geometry provides repeatable predictive value beyond matched classical alternatives when both receive identical low-dimensional representations under a leakage-aware temporal protocol. The evidence is organised around these mechanisms:

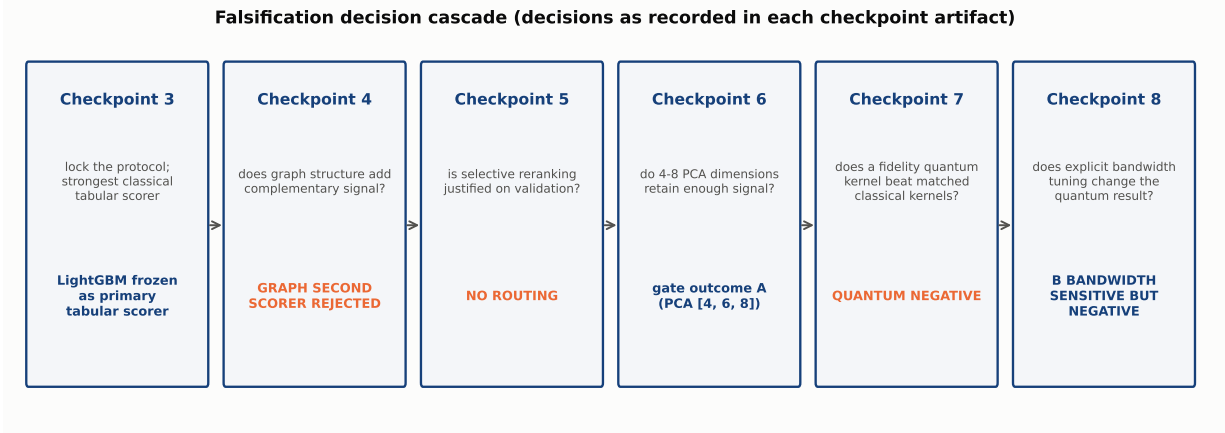


Figure 1: Falsification decision cascade: the question, decision and consequence of each stage, from the classical baseline to the secondary bandwidth correction.

temporal drift, representation loss, kernel concentration and bandwidth, and matched quantum–classical performance. Graph structure and selective reranking were examined first as falsification steps that kept unsupported complexity out of the final direct kernel audit (Figure 1).

The design separates a *primary* analysis (the default-bandwidth quantum kernel, confirmatory under locked protocols) from a *secondary* analysis (an explicit bandwidth correction added after review, post hoc relative to the primary result). All protocols were written to hash-recorded files before the corresponding stage ran; there is no public registry entry, so the term “pre-specified” refers to repository-internal, timestamped locking. The conclusions are restricted to the Elliptic++ transaction tables, the tested temporal protocol, PCA dimensions 4, 6 and 8, the tested fidelity feature maps, the tested bandwidth range and a training sample of 2,000 rows.

2 Related Work

Datasets. Elliptic++ extends the Elliptic Bitcoin transaction data [5] with transaction and actor tables [6].

Quantum kernels. Quantum feature maps define kernels through state overlaps [7, 8], and supervised quantum models can be analysed as kernel methods [9]. The role of data [10], bandwidth [2, 3] and exponential concentration [4] in quantum kernels has been studied, and benchmarking against classical models has been discussed critically [1]. Software documentation for the feature maps and fidelity kernels used here is cited as vendor documentation [11–13].

Classical tools and statistics. Baselines use gradient boosting [14, 15], scikit-learn [16], GraphSAGE [17], probability calibration [18], centered kernel alignment [19], trustworthiness [20] and a moving-block bootstrap [21]. This section states only what each cited source is about; positioning claims relative to prior work await a systematic literature review.

Table 1: Dataset periods and temporal protocol. Illicit share is the share of labeled rows.

Period	Steps	Labeled rows	Illicit rows	Illicit share	Role
training	1–24	23,606	2,219	9.4%	fit
validation	25–29	2,775	652	23.5%	selection
test window 30–34	30–34	3,513	591	16.8%	locked evaluation window
test window 35–39	35–39	5,486	447	8.1%	locked evaluation window
test window 40–44	40–44	7,458	515	6.9%	locked evaluation window
test window 45–49	45–49	3,726	121	3.2%	locked evaluation window

3 Dataset and Leakage Audit

The primary experiment uses the transaction tables only: 203,769 transactions over 49 time steps with 234,355 directed edges and 182 predictive features after excluding the transaction identifier and the time step. About 23% of transactions are labeled, and the labeled fraction varies from 11% to 43% across steps; every metric is therefore conditional on being labeled, and unknown rows are never treated as licit. Actor tables are excluded because their features are cumulative across steps. Several properties of the data limit leakage claims. First, the shipped local and aggregate features were standardised with statistics from all time steps, which cannot be undone; “leakage-aware” therefore refers to the evaluation protocol, not to the raw features. Second, 965 transactions have all augmented features missing, and all labeled ones among them (154 labeled, 0 illicit) are licit; no missingness indicator is used and the rows are kept. Third, the time of labeling and the real duration of a time step are undocumented. Fourth, redistribution terms are ambiguous, so raw data and row-level derived outputs are not released (Data and Code Availability).

4 Temporal Evaluation Protocol

Models are fitted on steps 1–24, selected on steps 25–29, and evaluated once on steps 30–49, reported in 4 locked windows (30–34, 35–39, 40–44, 45–49); Table 1 lists the periods. The primary metric is pooled Recall@5%: at each step the highest-scoring 5% of labeled rows form the alert list, true positives and positives are summed over steps, and the ratio is reported. The achievable ceiling is $\min(1, K/P)$ for list size K and P positives; equal scores are ordered by a seeded permutation. Each stochastic component uses 5 locked seeds, and the quantum comparison uses 5 independent training samples. Validation Recall@5% saturates for strong models, so selection uses Recall@5% within its binomial standard error, then PR-AUC, then Brier score after leave-one-validation-step-out Platt calibration. Uncertainty for paired comparisons uses a moving-block bootstrap over test time steps (block length 3, 10,000 resamples, 95% percentile interval) with step-level true positives averaged over the 5 samples. No multiplicity adjustment is applied. The interpretation rule for the quantum comparison requires a result in at least 2 of 3 dimensions, and differences smaller than the paired minimum detectable difference (about 2.4 points) are near the resolution of this test. Protocols, search spaces and manifests were hashed before use, and test labels were loaded only after the relevant manifests verified.

5 Classical Baselines and Temporal Failure

Under the locked protocol, the frozen full-dimensional LightGBM model reached pooled test Recall@5% of 0.381 and PR-AUC of 0.861; CatBoost reached 0.384 and 0.868, and logistic regression

Table 2: Classical baselines, pooled test metrics (primary protocol; mean of the locked seeds).

Model	Recall@5%	PR-AUC	Recall@2%	Recall@10%	Precision@5%	Brier	ECE
Logistic regression	0.156	0.229	0.047	0.395	0.256	0.208	0.232
LightGBM	0.381	0.861	0.169	0.639	0.625	0.021	0.014
CatBoost	0.384	0.868	0.169	0.644	0.631	0.021	0.012

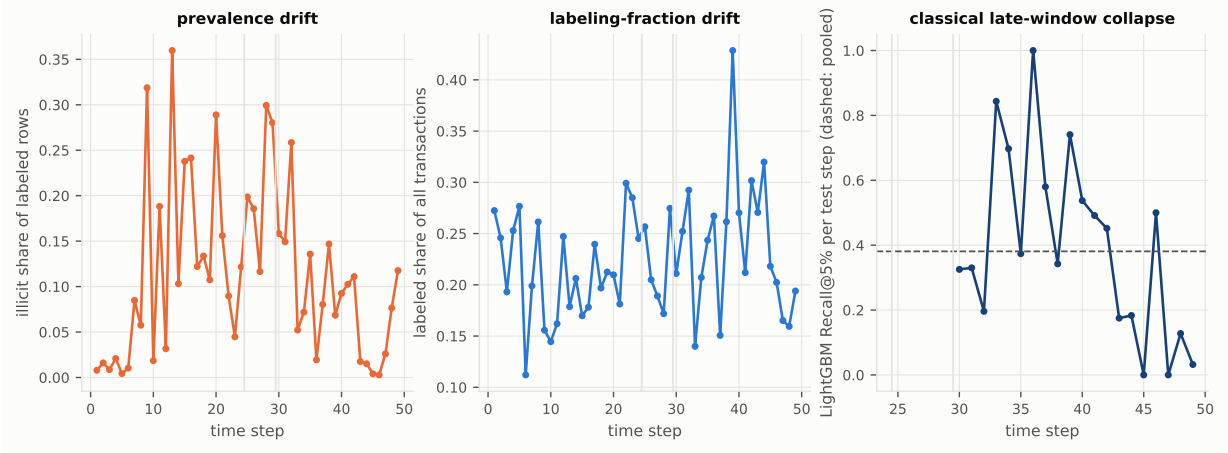


Figure 2: Temporal drift: illicit share of labeled rows (left), labeled share of transactions (centre) and LightGBM Recall@5% per test step (right; dashed line: pooled value). Vertical lines separate training, validation and test.

0.156 and 0.229 (Table 2). LightGBM was frozen as the practical reference and is not a primary comparator for the quantum kernel. Performance was not stable across time: LightGBM Recall@5% was 0.295, 0.497, 0.454 and 0.061 in the 4 windows while the labeled illicit share fell from 16.8% to 3.2% (Figure 2; training 9.4%, validation 23.5%). The late-window collapse appears in every model family examined below, so it is treated as dataset drift and not as a model-specific failure.

6 Falsification of Graph Complementarity and Selective Routing

Graph. On per-step transaction graphs, GraphSAGE reached pooled Recall@5% of 0.342 against 0.370 for a matched edge-free MLP (difference -0.028 , 95% block-bootstrap interval -0.063 to -0.007), and randomised edges gave 0.344 (a single seed). The union of the GraphSAGE and LightGBM alert lists recalled 0.503 of positives, whereas LightGBM read to the same list length recalled 0.527; the graph model was therefore rejected as a second scorer.

Routing. On validation, LightGBM’s top-5% list had precision 1.0, so a perfect reranker of any routed subset had oracle Recall@5% headroom of 0.0 percentage points, and 0 of 37 candidate routing rules passed the locked criteria. Validation is saturated, so this shows an absence of headroom on validation, not in later steps. Both results are retained as negative results; they removed graph and routing components from the kernel audit rather than adding them.

Table 3: Representation loss for PCA at the quantum dimensions. Drift is the mean Wasserstein distance from training in window 45–49 (training standard-deviation units). Test columns are evaluation-only and follow validation selection.

Dim.	Variance	NN Jaccard	Trustworth.	Drift (PCA / 182-D)	Validation-selected model	Test Recall@5%	Fraction of LightGBM
4	34%	0.13	0.81	1.20 / 0.38	MLP	0.245	64%
6	43%	0.20	0.89	1.08 / 0.38	Nyström + linear	0.189	50%
8	51%	0.25	0.91	1.10 / 0.38	MLP	0.124	33%

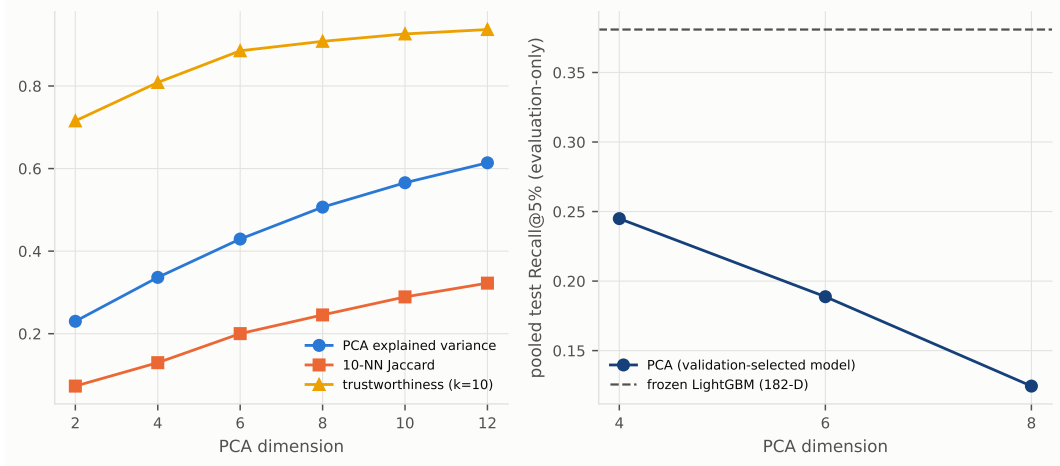


Figure 3: Representation loss for PCA: explained variance, neighbourhood preservation and trustworthiness (left) and test Recall@5% of the validation-selected model against the frozen full-dimensional LightGBM (right; evaluation-only).

7 Representation-Loss Audit

PCA was fitted on training steps only and frozen. At 4, 6 and 8 dimensions it retained 34%, 43% and 51% of the variance, with 10-nearest-neighbour Jaccard overlap of 0.13, 0.20 and 0.25 and trustworthiness of 0.81, 0.89 and 0.91 (Table 3, Figure 3). Compressed coordinates drifted more than the original features: the mean Wasserstein distance from training in window 45–49 was 1.20 training standard deviations for 4 dimensions against 0.38 for the full 182-dimensional space. The validation-selected classical models on these representations reached 64%, 50% and 33% of LightGBM’s pooled test Recall@5%. Compression therefore removed most of the full-dimensional signal on test, and quantum and classical kernels compete within this weak-signal regime. A minimal supervised feature selection performed unexpectedly well on test, but this was observed only after test unblinding; it is a test-informed secondary observation, is excluded from the primary quantum analysis and from every headline claim, and needs an independent dataset or a new specification before it can be assessed.

8 Quantum Kernel Protocol

The quantum model is a fidelity kernel $K(x, y) = |\langle \phi(x) | \phi(y) \rangle|^2$ evaluated by exact ideal statevector simulation, with a support-vector classifier on the precomputed kernel; no hardware, noise model or variational circuit was used. The entangling map with linear ZZ-type blocks (Q_E) and the separable map without entangling gates (Q_S) were compared, each with 1 or 2 repetitions chosen

on validation and qubit count equal to the PCA dimension (4, 6, 8); no coordinate was padded or dropped. Frozen PCA coordinates were scaled using training rows only and mapped to angles in $[-\pi, \pi]$, and the scaling was selected by the matched classical RBF kernel so that it was not tuned for the quantum kernel. Quantum and classical models received identical rows, identical scaled coordinates and identical temporally stratified natural-prevalence training samples of 2,000 rows (5 samples; the tuning sample held 180 positives). Matched classical controls (RBF and polynomial SVMs, Nyström and random Fourier feature maps with a linear classifier, a linear model and a small MLP) used the same inputs and presampled search spaces. The strongest matched classical kernel at each dimension was fixed by validation before test scoring. The default-bandwidth quantum search used 16 fits per dimension against 96 for the matched classical models. The batched simulation was verified against Qiskit state vectors and the fidelity-kernel implementation [11–13]. The primary analysis fixed the Pauli-rotation factor at its verified default of 2.0; classical kernels, in contrast, tuned a bandwidth parameter.

9 Bandwidth Sensitivity and Kernel Geometry

Because the primary quantum kernel had no tuned bandwidth, a secondary, post hoc correction was specified before any candidate kernel was computed and was never used to replace the primary result. The only new hyperparameter was a multiplier s on the verified Pauli-rotation factor, $\alpha = 2.0s$, applied to every Pauli rotation, with 5 values from 0.125 to 2; $s = 1$ reproduces the primary kernel exactly. Scale and the regularisation constant (4 values from 0.1 to 100) were selected on validation with the same hierarchy as before, with 40 fits per dimension against 96 for the matched classical models. The entangling kernel selected $s = 0.125$ ($\alpha = 0.25$), the smallest value of the grid, in every dimension; the optimum may therefore lie below the tested range. The grid was deliberately not extended: extending it after observing test results would add researcher degrees of freedom, and this is recorded as a limitation and a replication question. Bandwidth correction reduced concentration of the 8-qubit entangling kernel: the effective rank over sample size fell from 0.90 to 0.25 and the share of similarities below 0.01 fell from 0.84 to 0.68 (Figure 4). Centered alignment with the matched RBF kernel rose from 0.24, 0.26 and 0.32 at the default scale to 0.77, 0.82 and 0.82 at the selected scale for 4, 6 and 8 qubits, so the improved kernel was geometrically closer to a classical RBF kernel. This is an association across the default and selected settings of each dimension; it is consistent with geometric distinctiveness not being sufficient for predictive value in the tested setting, and it does not establish a causal link or any statement about quantum geometry in general.

10 Results

10.1 Primary analysis: default bandwidth

At default bandwidth the entangling kernel Q_E reached pooled test Recall@5% of 0.039, 0.046 and 0.054 at 4, 6 and 8 qubits, against 0.118, 0.249 and 0.200 for the strongest matched classical kernels (Table 4, Figure 5). The paired differences were -0.078 (95% interval -0.127 to -0.050), -0.203 (-0.275 to -0.153) and -0.146 (-0.196 to -0.091), all intervals lying below zero; the entangling kernel was below the strongest matched classical kernel in all 3 dimensions. The separable control Q_S reached 0.089, 0.164 and 0.124, and at default bandwidth Q_E minus Q_S was -0.049 (-0.103 to -0.005), -0.118 (-0.200 to -0.032) and -0.070 (-0.118 to -0.008). These statements describe the default-bandwidth kernel only.

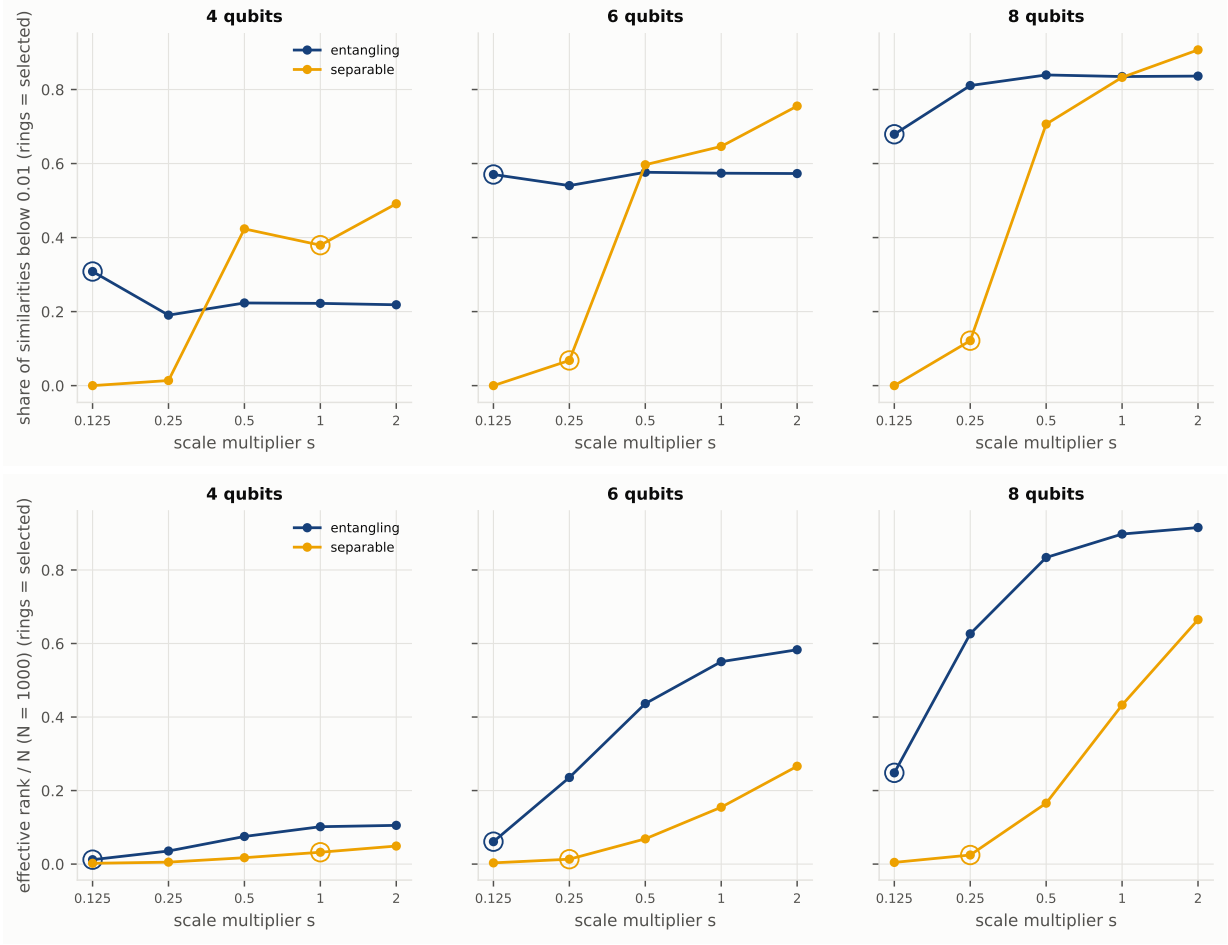


Figure 4: Secondary analysis: share of kernel similarities below 0.01 (top) and effective rank over sample size (bottom) against the scale multiplier s for the entangling and separable kernels; rings mark the selected scale.

10.2 Secondary analysis: bandwidth correction

This subsection reports the secondary, post hoc bandwidth correction. The corrected entangling kernel reached 0.109, 0.167 and 0.163 (Table 5, Figure 6), improving on the default-bandwidth kernel by 0.070 (0.054 to 0.100), 0.121 (0.066 to 0.180) and 0.109 (0.061 to 0.142). Relative to the strongest matched classical kernel the corrected differences were -0.008 (-0.030 to 0.011), -0.082 (-0.141 to -0.039) and -0.037 (-0.067 to -0.015). At 4 qubits the interval contains zero, so the difference is statistically unresolved; an unresolved difference neither shows that the kernels perform alike nor favours the quantum kernel. At 6 and 8 qubits the intervals lie below zero, so the corrected kernel remained below the strongest matched classical kernel in this setting. The apparent effect of entanglement depended strongly on bandwidth: corrected Q_E minus corrected Q_S was 0.021 (-0.008 to 0.058) and 0.011 (-0.033 to 0.077), both unresolved, and 0.074 (0.032 to 0.146) at 8 qubits; the 8-qubit comparison is an unadjusted secondary comparison and is not generalised.

Table 4: Primary (confirmatory) analysis at default bandwidth: pooled test Recall@5% of the entangling kernel Q_E and the separable control Q_S versus the strongest matched classical kernel; intervals are paired moving-block bootstrap intervals over test steps (no multiplicity adjustment).

Qubits	Q_E	Q_S	Strongest matched kernel (Recall@5%)	$Q_E - \text{matched}$ [95% interval]	$Q_E - Q_S$ [95% interval]
4	0.039	0.089	RBF SVM (0.118)	-0.078 [-0.127, -0.050]	-0.049 [-0.103, -0.005]
6	0.046	0.164	Nyström + linear (0.249)	-0.203 [-0.275, -0.153]	-0.118 [-0.200, -0.032]
8	0.054	0.124	Nyström + linear (0.200)	-0.146 [-0.196, -0.091]	-0.070 [-0.118, -0.008]

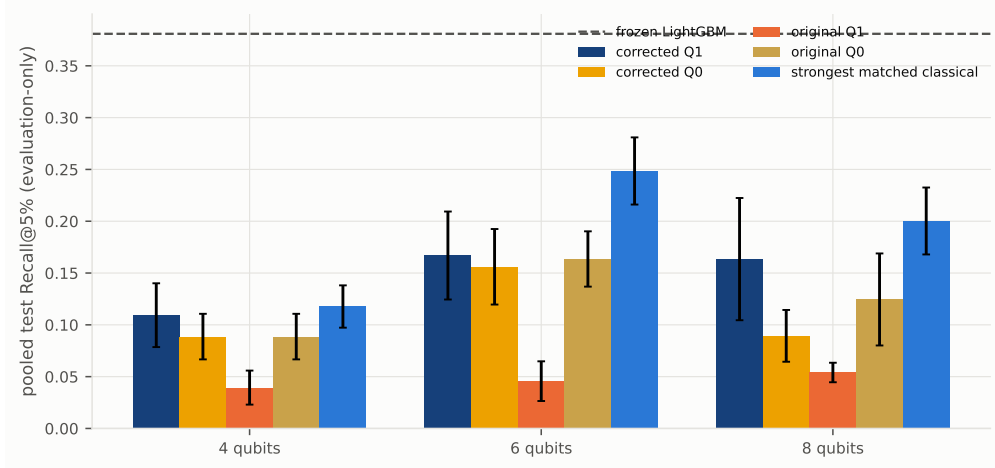


Figure 5: Primary analysis: pooled test Recall@5% of the entangling kernel Q_E and separable control Q_S (default and secondary bandwidth-corrected variants as labelled) against the strongest matched classical kernel and LightGBM (error bars: standard deviation over 5 samples; evaluation-only).

10.3 Performance by locked window and resources

Every model was weak in the final window (Table 6, Figure 7): full-dimensional LightGBM reached 0.061 in window 45–49 against 0.295 in window 30–34, so the collapse is not specific to compressed or quantum models. Under ideal simulation the quantum models were not cheaper in the quantities measured here (Table 7, Figure 8): test scoring of the default-bandwidth kernels required 363,654,720 exact kernel evaluations in 13.9 s, the corrected kernels 210,341,196 in 11.0 s, and the matched classical models took 13.5 s; these are simulation costs on a single machine, not hardware costs.

Table 5: Secondary (post hoc) bandwidth correction: the entangling kernel Q_E at the selected scale s versus the default-bandwidth kernel, the corrected separable control Q_S and the strongest matched classical kernel (paired moving-block bootstrap intervals; no multiplicity adjustment). The selected scale is the grid minimum in every dimension.

Qubits	s	Default Q_E	Corrected Q_E	Corrected Q_S	Matched	Corrected Q_E - matched	Corrected - default Q_E	Corrected Q_E - Q_S	Reading
4	0.125	0.039	0.109	0.089	0.118	-0.008 [-0.030, 0.011]	0.070 [0.054, 0.100]	0.021 [-0.008, 0.058]	unresolved
6	0.125	0.046	0.167	0.156	0.249	-0.082 [-0.141, -0.039]	0.121 [0.066, 0.180]	0.011 [-0.033, 0.077]	below matched (this setting)
8	0.125	0.054	0.163	0.089	0.200	-0.037 [-0.067, -0.015]	0.109 [0.061, 0.142]	0.074 [0.032, 0.146]	below matched (this setting)

Table 6: Recall@5% by locked test window for full-dimensional LightGBM and the 4, 6 and 8-dimensional models (default and corrected quantum kernels, matched classical kernels, validation-selected PCA models).

Dim.	Model	30-34	35-39	40-44	45-49
4	default Q_E	0.059	0.026	0.019	0.079
4	corrected Q_E	0.132	0.138	0.068	0.069
4	corrected Q_S	0.130	0.116	0.023	0.069
4	matched RBF SVM	0.146	0.154	0.067	0.063
4	best PCA model (validation-selected)	0.243	0.310	0.232	0.069
6	default Q_E	0.047	0.041	0.044	0.061
6	corrected Q_E	0.206	0.227	0.092	0.074
6	corrected Q_S	0.215	0.124	0.147	0.025
6	matched Nyström + linear	0.240	0.311	0.252	0.045
6	best PCA model (validation-selected)	0.240	0.168	0.177	0.068
8	default Q_E	0.061	0.066	0.032	0.071
8	corrected Q_E	0.215	0.172	0.118	0.073
8	corrected Q_S	0.179	0.022	0.050	0.064
8	matched Nyström + linear	0.227	0.227	0.181	0.053
8	best PCA model (validation-selected)	0.246	0.052	0.071	0.030
full	LightGBM	0.295	0.497	0.454	0.061

Table 7: Resource comparison from the run ledgers (ideal simulation on a single CPU; wall and CPU time summed over all models and samples; circuit sizes after generic-basis decomposition are characterisation only).

Phase	Wall time (s)	CPU time (s)	Kernel evaluations / circuit size
matched classical final fits	8.4	10.5	—
matched classical test scoring	13.5	29.6	—
default-bandwidth quantum test scoring	13.9	54.8	363,654,720
bandwidth-corrected quantum test scoring	11.0	37.5	210,341,196
default-bandwidth quantum final fits (span)	36.8	58.2	—
bandwidth-corrected quantum final fits (span)	36.1	57.5	—
selected Q_E circuit, 4 qubits: depth / two-qubit gates	—	—	11 / 6
selected Q_E circuit, 6 qubits: depth / two-qubit gates	—	—	17 / 10
selected Q_E circuit, 8 qubits: depth / two-qubit gates	—	—	31 / 28

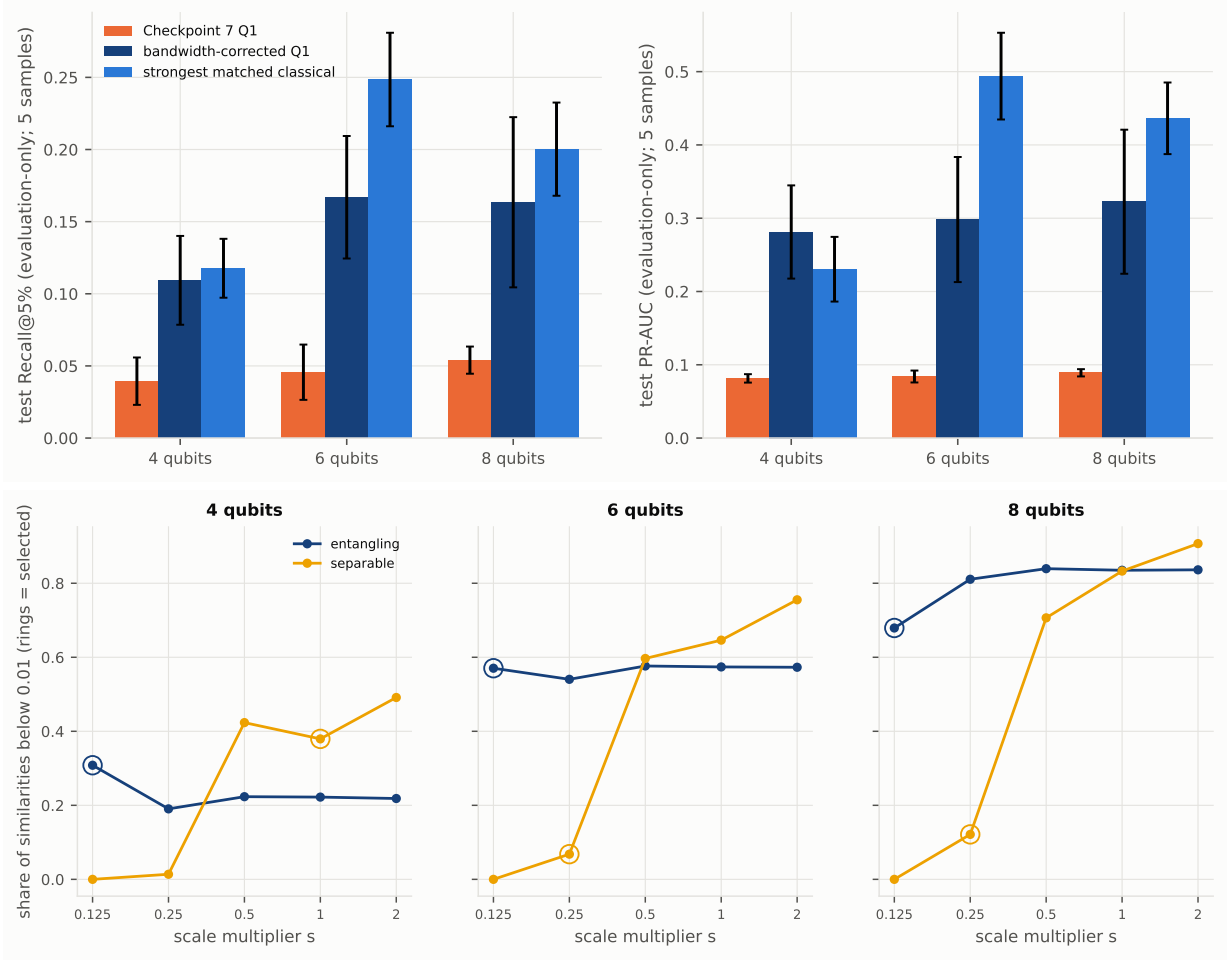


Figure 6: Secondary analysis: pooled test Recall@5% of the default-bandwidth and bandwidth-corrected entangling kernels and the strongest matched classical kernel (top) with the associated concentration (bottom).

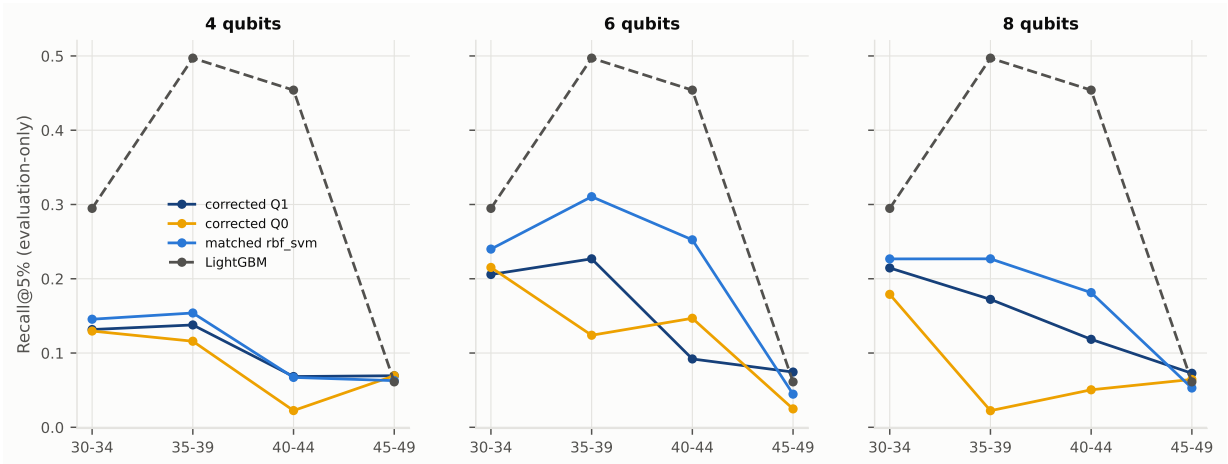


Figure 7: Recall@5% by locked test window for the corrected and separable kernels (secondary), the strongest matched classical kernel and LightGBM at each qubit count.

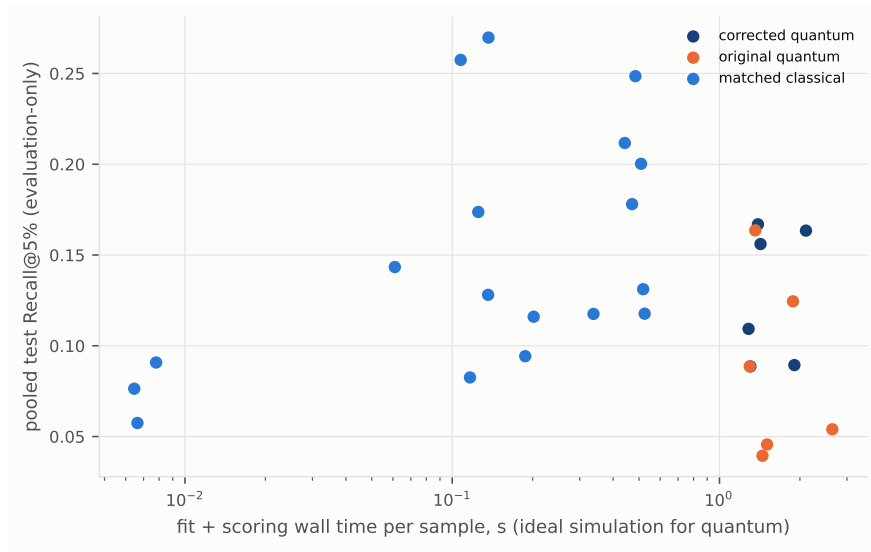


Figure 8: Test Recall@5% against fit-plus-scoring wall time per sample for the quantum kernels (ideal simulation) and matched classical models.

11 Discussion

Mechanisms. *Temporal drift* affected every model family: the labeled illicit share fell from 16.8% to 3.2% and Recall@5% of the strongest classical model fell to 0.061 in the final window. *Representation loss* removed most of the full-dimensional signal and amplified feature drift. *Concentration and bandwidth* explained most of the primary quantum deficit: explicit bandwidth tuning, a degree of freedom the classical kernels already had, substantially improved the entangling kernel. *Matched performance* still favoured the classical kernels at 6 and 8 qubits and was unresolved at 4 qubits.

Geometry. The default-bandwidth kernel was geometrically distinct from the classical kernels but performed worst, and the corrected kernel became more RBF-like while improving; within this setting, geometric distinctiveness was not sufficient for predictive value.

Entanglement. The sign of the entangling-versus-separable difference changed with bandwidth, so no general statement about entanglement is supported.

Hardware. The mechanical hardware gate returned `PASS_CONDITION_C`, driven only by finite-shot ranking instability (top-5% overlap with the exact ranking of 0.60, 0.96 and 0.88 at 1024 shots on a 500-row subset); the scientific decision was `DEFER` because the ideal model was already below the matched classical kernel. No quantum hardware was run, and no statement about hardware performance is made.

What this does and does not show. The evidence supports the absence of a repeatable predictive benefit of the tested quantum kernels over matched classical kernels in the tested setting. It does not show that quantum kernels cannot help on other data, representations, feature maps, bandwidths or sample sizes.

12 Limitations

Threats to internal validity. (i) Validation prevalence (23.5%) differs from test prevalence and saturates Recall@5%, so validation-based selection and the representation feasibility gate were permissive. (ii) The corrected entangling scale was selected at the grid boundary and the correction is secondary and post hoc relative to the primary result. (iii) PCA compression removed most of the full-dimensional signal. (iv) The shipped features were standardised with all time steps, and the missingness pattern aligns with the licit label. (v) Only a non-random share of transactions is labeled. (vi) Quantum and kernel-SVM models used 2,000-row samples while LightGBM used all labeled rows; matched controls used identical samples. (vii) Test scores vary across the 5 samples and the final window has few positives; intervals are wide relative to differences of the order of the paired minimum detectable difference (2.4 points), and secondary comparisons are not multiplicity-adjusted. (viii) A test-informed observation exists in the record and is excluded from all conclusions. (ix) Quantum results are ideal simulations, with finite shots modelled by a binomial draw. (x) The feature-map family is limited to separable and linear-entanglement ZZ-type maps with few repetitions.

Threats to external validity. Results come from a single dataset with ambiguous redistribution terms, undocumented label timing and no external replication; actor features and graph models were excluded from the final audit; and nothing is claimed about quantum hardware.

13 Conclusion

In the primary, default-bandwidth analysis, the entangling fidelity kernel was below the strongest matched classical kernel at 4, 6 and 8 qubits. In the secondary bandwidth correction, explicit scale tuning substantially improved the kernel and made its geometry more RBF-like, but the corrected kernel remained below the strongest matched classical kernel at 6 and 8 qubits and was statistically unresolved at 4 qubits; the selected scale lay at the grid boundary. Temporal drift affected every model family, and compression to 4–8 dimensions removed most of the full-dimensional signal. No repeatable predictive benefit of the quantum kernel over matched classical kernels was observed in the tested setting. Reported quantum-kernel performance should therefore be accompanied by bandwidth tuning, matched classical controls on identical inputs, and time-ordered evaluation. Extending the bandwidth range, other datasets and other representations are future hypotheses, not findings.

Data and Code Availability

Code, configurations, protocol and lock files, frozen-model manifests, hash records, tables and figure inputs derived from aggregate results are available from the corresponding author upon reasonable request. The raw Elliptic++ data and all row-level derived outputs (predictions, kernel caches, model binaries) are not redistributed; the raw file hashes are listed in the supplement so that independent copies can be verified.

Acknowledgments

The authors acknowledge the developers and maintainers of the open-source software used in this study.

Declarations

Funding: This research received no external funding.

Competing interests: The authors declare no competing interests.

Author contributions: Mohamed Faisal Sindhi and Hammaz Buksh contributed equally to Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Visualization, Writing — original draft, and Writing — review & editing.

Ethics statement: This study used an existing transaction dataset and did not involve human participants, interventions or identifiable private information. No quantum hardware and no data beyond Elliptic++ were used. Software versions are listed in the supplement (Python 3.12.9, Qiskit 2.5.2, Qiskit Machine Learning 0.9.1, scikit-learn 1.9.1).

References

- [1] Joseph Bowles, Shahnawaz Ahmed, and Maria Schuld. Better than classical? the subtle art of benchmarking quantum machine learning models, 2024. preprint (arXiv record).
- [2] Ruslan Shaydulin and Stefan M. Wild. Importance of kernel bandwidth in quantum machine learning. *Physical Review A*, 106(4):042407, 2022. doi: 10.1103/PhysRevA.106.042407. peer-reviewed journal article.

- [3] Abdulkadir Canatar, Evan Peters, Cengiz Pehlevan, Stefan M. Wild, and Ruslan Shaydulin. Bandwidth enables generalization in quantum kernel models, 2022. preprint (arXiv record).
- [4] Supanut Thanasilp, Samson Wang, M. Cerezo, and Zoë Holmes. Exponential concentration in quantum kernel methods. *Nature Communications*, 15(1):5200, 2024. doi: 10.1038/s41467-024-49287-w. peer-reviewed journal article.
- [5] Mark Weber, Giacomo Domeniconi, Jie Chen, Daniel Karl I. Weidele, Claudio Bellei, Tom Robinson, and Charles E. Leiserson. Anti-money laundering in bitcoin: Experimenting with graph convolutional networks for financial forensics, 2019. preprint (arXiv record).
- [6] Youssef Elmougy and Ling Liu. Demystifying fraudulent transactions and illicit nodes in the bitcoin network for financial forensics. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 3979–3990. ACM, 2023. doi: 10.1145/3580305.3599803. peer-reviewed conference paper.
- [7] Vojtěch Havlíček, Antonio D. Córcoles, Kristan Temme, Aram W. Harrow, Abhinav Kandala, Jerry M. Chow, and Jay M. Gambetta. Supervised learning with quantum-enhanced feature spaces. *Nature*, 567(7747):209–212, 2019. doi: 10.1038/s41586-019-0980-2. peer-reviewed journal article.
- [8] Maria Schuld and Nathan Killoran. Quantum machine learning in feature hilbert spaces. *Physical Review Letters*, 122(4):040504, 2019. doi: 10.1103/PhysRevLett.122.040504. peer-reviewed journal article.
- [9] Maria Schuld. Supervised quantum machine learning models are kernel methods, 2021. preprint (arXiv record).
- [10] Hsin-Yuan Huang, Michael Broughton, Masoud Mohseni, Ryan Babbush, Sergio Boixo, Hartmut Neven, and Jarrod R. McClean. Power of data in quantum machine learning. *Nature Communications*, 12(1):2631, 2021. doi: 10.1038/s41467-021-22539-9. peer-reviewed journal article.
- [11] Qiskit documentation: `zz_feature_map` (qiskit.circuit.library). https://quantum.cloud.ibm.com/docs/en/api/qiskit/qiskit.circuit.library.zz_feature_map. vendor/official documentation.
- [12] Qiskit machine learning documentation: `FidelityStatevectorKernel`. https://qiskit-community.github.io/qiskit-machine-learning/stubs/qiskit_machine_learning.kernels.FidelityStatevectorKernel.html. vendor/official documentation.
- [13] Ali Javadi-Abhari, Matthew Treinish, Kevin Krsulich, Christopher J. Wood, Jake Lishman, Julien Gacon, Simon Martiel, Paul D. Nation, Lev S. Bishop, Andrew W. Cross, Blake R. Johnson, and Jay M. Gambetta. Quantum computing with qiskit, 2024. preprint (arXiv record).
- [14] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. https://papers.nips.cc/paper_files/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html, 2017. Advances in Neural Information Processing Systems; publisher or proceedings page (title found on page).

- [15] Liudmila Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin. Catboost: unbiased boosting with categorical features, 2017. preprint (arXiv record).
- [16] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Andreas Müller, Joel Nothman, Gilles Louppe, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine learning in python, 2012. preprint (arXiv record); arXiv lists journal reference: Journal of Machine Learning Research (2011).
- [17] William L. Hamilton, Rex Ying, and Jure Leskovec. Inductive representation learning on large graphs, 2017. preprint (arXiv record).
- [18] Alexandru Niculescu-Mizil and Rich Caruana. Predicting good probabilities with supervised learning. In *Proceedings of the 22nd international conference on Machine learning - ICML '05*, pages 625–632. ACM Press, 2005. doi: 10.1145/1102351.1102430. peer-reviewed conference paper.
- [19] Corinna Cortes, Mehryar Mohri, and Afshin Rostamizadeh. Algorithms for learning kernels based on centered alignment. <https://www.jmlr.org/papers/v13/cortes12a.html>, 2012. Journal of Machine Learning Research; publisher or proceedings page (title found on page).
- [20] Jarkko Venna and Samuel Kaski. Neighborhood preservation in nonlinear projection methods: An experimental study. In *Lecture Notes in Computer Science*, pages 485–491. Springer Berlin Heidelberg, 2001. doi: 10.1007/3-540-44668-0_68. published conference-proceedings chapter (Crossref type book-chapter; peer-review status not asserted).
- [21] Hans R. Kunsch. The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17(3), 1989. doi: 10.1214/aos/1176347265. peer-reviewed journal article.