Seasonality Detection Methods: A Comparative Study

Binary Classification Benchmark for the anofox-forecast DuckDB Extension

anofox-forecast benchmark suite

2026-01-08

Table of contents

Executive Summary

This benchmark evaluates seasonality detection methods as a **binary classification problem**: given a time series, does it contain seasonality? We simulate 550 curves with varying seasonal strength levels (0.0 to 1.0) and evaluate 13 detection methods using classification metrics (Accuracy, Precision, Recall, F1, ROC/AUC). Ground truth is defined as seasonal if simulated strength >= 0.2.

This benchmark replicates the methodology from the fdars R package benchmark.

Introduction

Detection as Binary Classification

Unlike period estimation (which asks "what is the period?"), **seasonality detection** asks a simpler question: **"is there seasonality?"** This is a binary classification problem where each method produces a confidence score, and we apply a threshold to make a detection decision.

Methods Evaluated

Method	SQL Function	Score Used	Description
AIC Comparison	`ts_aic_period`	R-squared	Fourier model fit quality
FFT Confidence	`ts_estimate_period_fft`	confidence	Peak-to-mean power ratio

Method	SQL Function	Score Used	Description
ACF Confidence	`ts_estimate_period_acf`	confidence	Autocorrelation at lag
Variance Strength	`ts_seasonal_strength(..., 'variance')`	strength	Seasonal variance ratio
Spectral Strength	`ts_seasonal_strength(..., 'spectral')`	strength	Power at seasonal frequency
Wavelet Strength	`ts_seasonal_strength(..., 'wavelet')`	strength	Morlet wavelet energy
SAZED	`ts_sazed_period`	SNR	Zero-padded spectral SNR
Autoperiod	`ts_autoperiod`	acf_validation	FFT+ACF hybrid validation
CFD-Autoperiod	`ts_cfd_autoperiod`	acf_validation	First-differenced FFT+ACF
Lomb-Scargle	`ts_lomb_scargle`	1-FAP	Statistical significance
Matrix Profile	`ts_matrix_profile_period`	confidence	Motif agreement ratio
STL	`ts_stl_period`	seasonal_strength	Decomposition strength
SSA	`ts_ssa_period`	variance_explained	Eigenvalue dominance

Ground Truth Definition

A series is classified as **seasonal** if its simulated seasonal strength >= 0.2. This threshold follows the fdars benchmark convention.

Setup

Connect to DuckDB and Load Extension

Baseline Simulation

Simulation Parameters

Following the fdars benchmark: - **11 strength levels**: 0.0, 0.1, 0.2, ..., 1.0 - **50 curves per level**: 550 total curves - **60 observations**: 5 years of monthly data - **Period = 12**: Monthly seasonality - **White noise**: sigma = 0.3

Baseline Data Generation

```
Generated 550 curves
```

```
Seasonal (strength >= 0.2): 450
```

```
Non-seasonal: 100
```

Strength Level Distribution

Example Curves

Load Data into DuckDB

```
[1] 0
```

```
[1] 0
```

```
Data loaded into DuckDB
```

Method Evaluation

Extract Confidence Scores

```
Extracted scores for 550 curves
```

Score Distributions by Ground Truth

ROC Analysis

Compute ROC Curves and AUC

Table 2: ROC Analysis Summary (sorted by AUC)

Method	AUC	Optimal Threshold	Sensitivity	Specificity
Variance	0.962	0.215	0.896	0.92
Spectral	0.952	0.335	0.822	0.96
AIC	0.937	0.213	0.822	0.96

Method	AUC	Optimal Threshold	Sensitivity	Specificity
FFT	0.935	0.063	0.816	0.96
Lomb	0.931	0.570	0.787	0.97
SSA	0.892	0.425	0.713	0.98
Autoperiod	0.863	0.202	0.727	0.90
SAZED	0.858	0.794	0.773	0.88
STL	0.801	0.501	0.607	0.92
ACF	0.782	0.268	0.571	0.98
CFD	0.738	0.268	0.447	0.97
MatrixProfile	0.719	0.229	0.604	0.73
Wavelet	0.608	0.991	0.996	0.22

ROC Curves

AUC Comparison

Classification Performance

Apply Optimal Thresholds

Classification Metrics

Table 3: Classification Performance at Optimal Thresholds (sorted by F1)

Method	Accuracy	Precision	Recall	Specificity	FPR	F1
Variance	0.900	0.981	0.896	0.92	0.08	0.936
Wavelet	0.855	0.852	0.996	0.22	0.78	0.918
AIC	0.847	0.989	0.822	0.96	0.04	0.898
Spectral	0.847	0.989	0.822	0.96	0.04	0.898
FFT	0.842	0.989	0.816	0.96	0.04	0.894
Lomb	0.820	0.992	0.787	0.97	0.03	0.877
SAZED	0.793	0.967	0.773	0.88	0.12	0.859
Autoperiod	0.758	0.970	0.727	0.90	0.10	0.831
SSA	0.762	0.994	0.713	0.98	0.02	0.831
STL	0.664	0.972	0.607	0.92	0.08	0.747
MatrixProfile	0.627	0.910	0.604	0.73	0.27	0.726
ACF	0.645	0.992	0.571	0.98	0.02	0.725
CFD	0.542	0.985	0.447	0.97	0.03	0.615

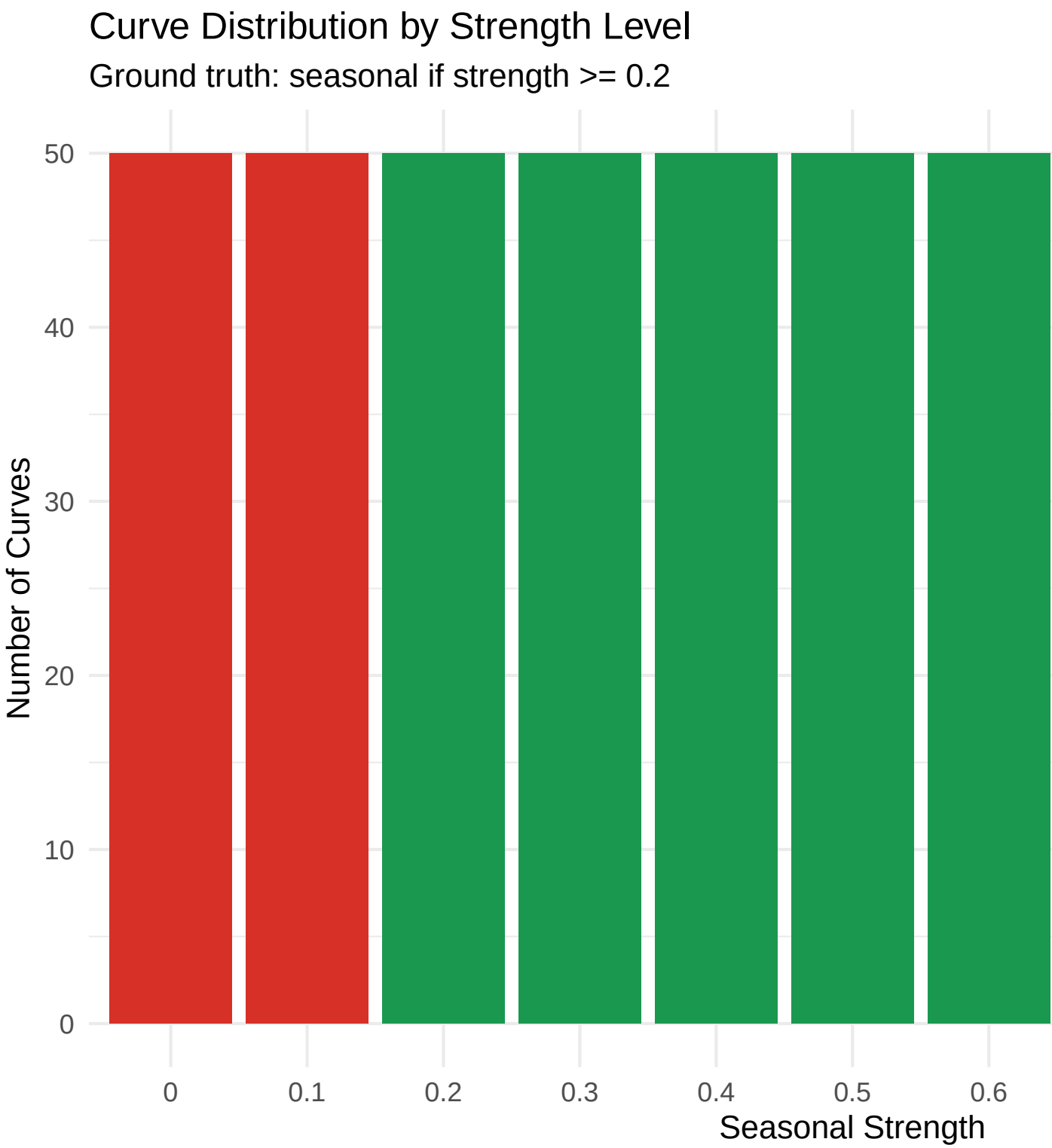

Figure 1: Distribution of curves by seasonal strength level

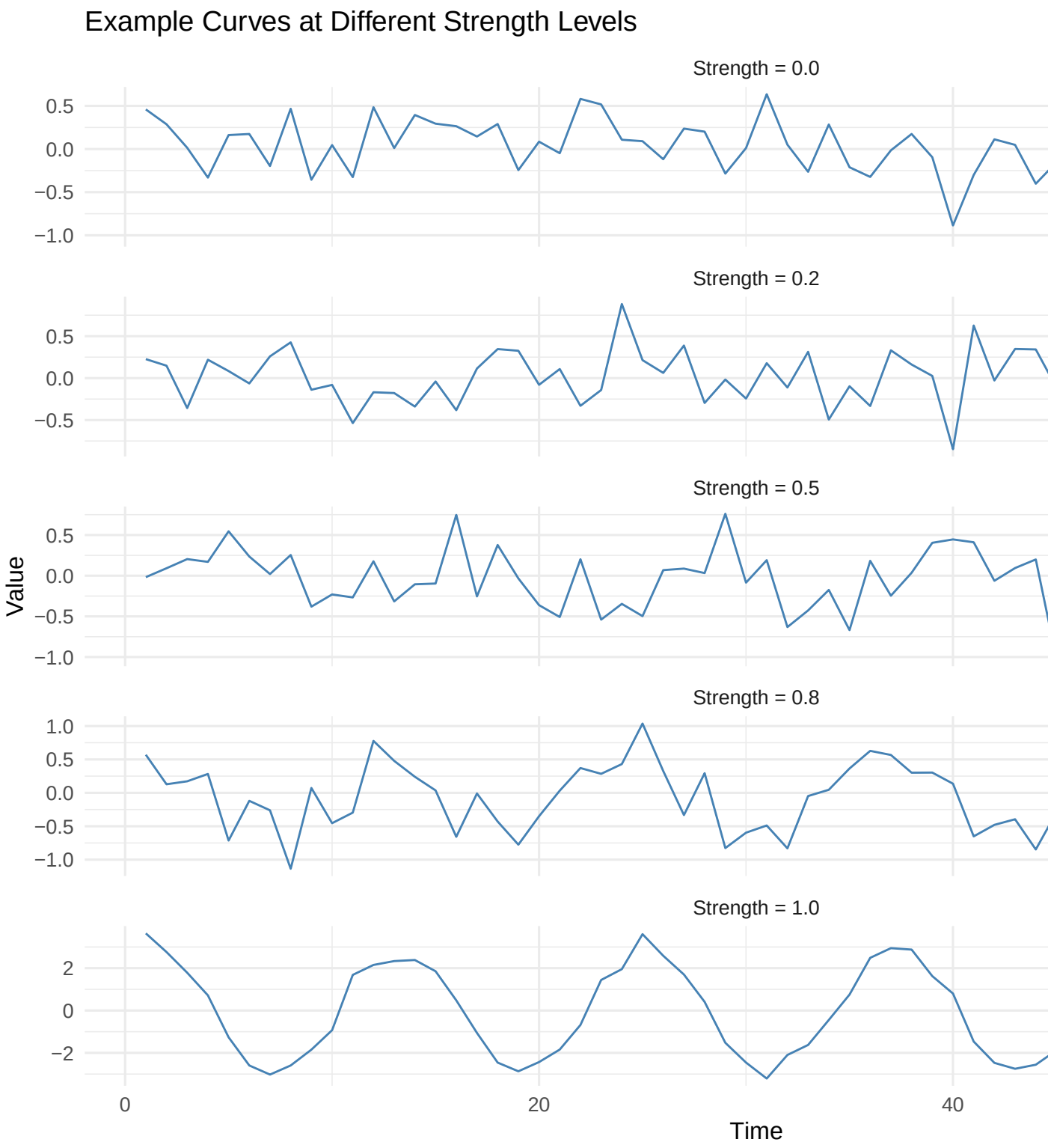

Figure 2: Example curves at different strength levels

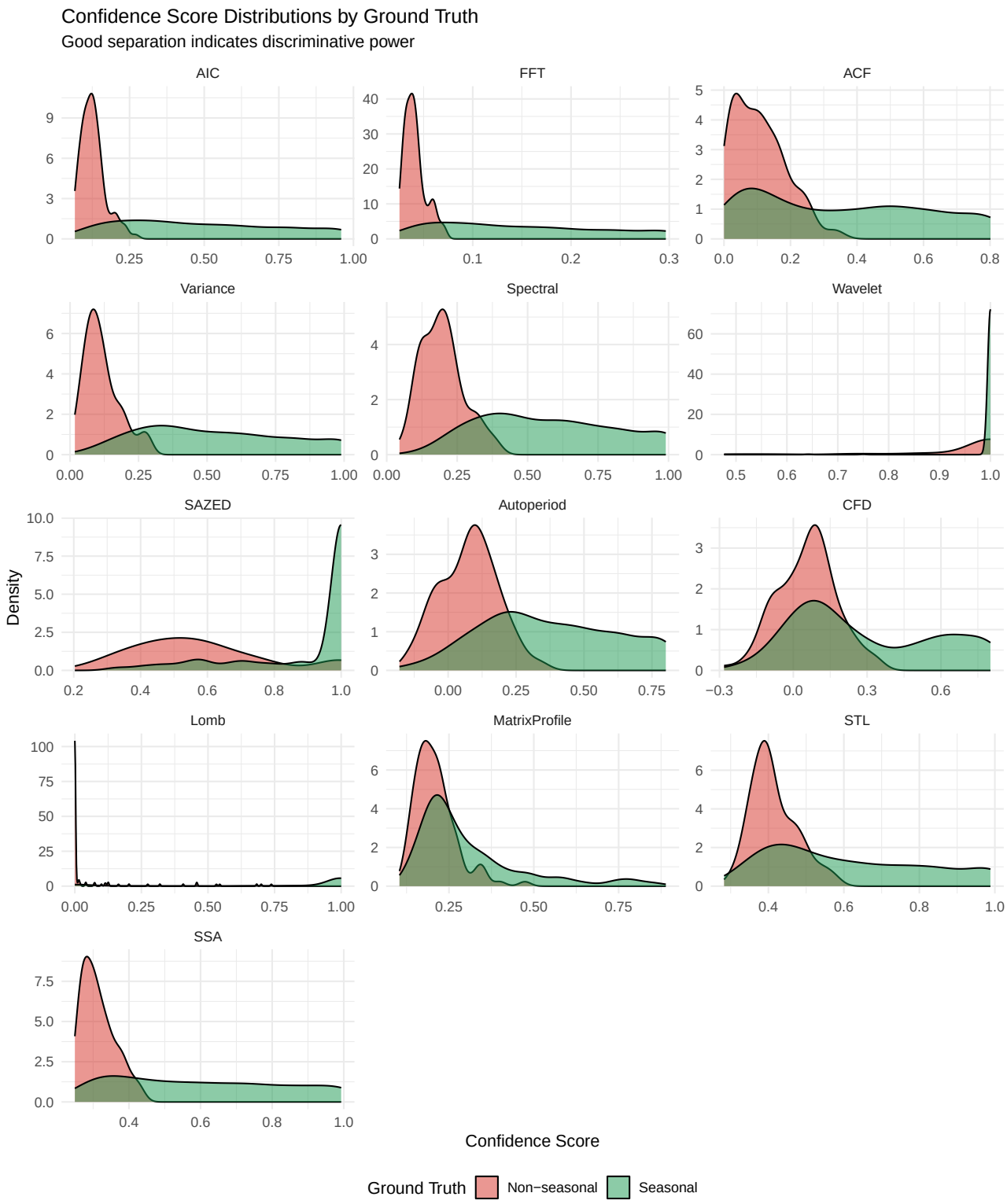

Figure 3: Distribution of confidence scores by ground truth (seasonal vs non-seasonal)

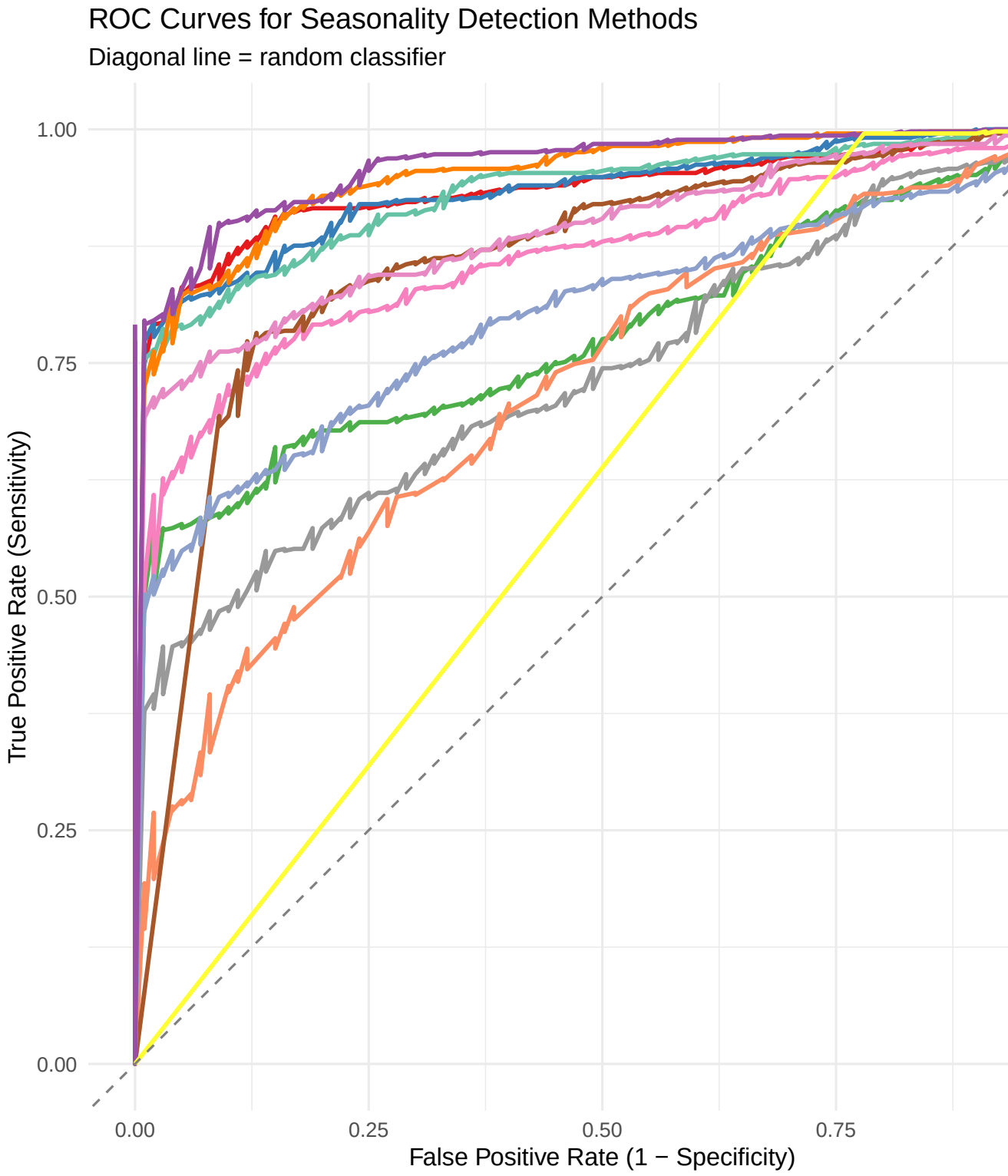

Figure 4: ROC curves for all detection methods

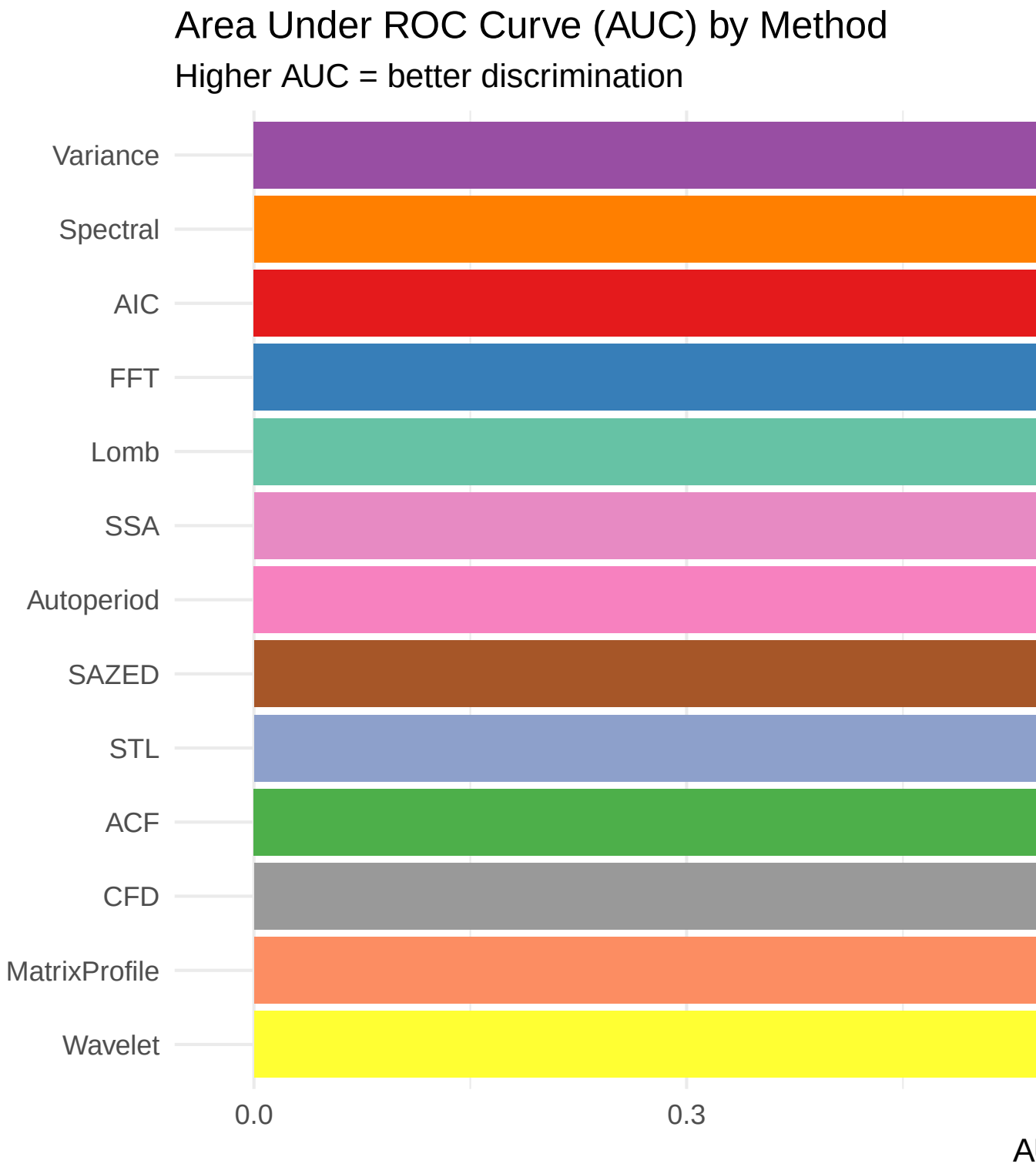

Figure 5: AUC comparison across methods

Statistical Significance: McNemar Tests

McNemar's test compares paired binary predictions between methods. A significant p-value indicates methods differ in their detection decisions.

Table 4: Significant McNemar Test Results ($p < 0.05$)

method1	method2	chi_sq	p_value
CFD	Wavelet	320.0031	0.0000
ACF	Wavelet	265.0037	0.0000
STL	Wavelet	239.1004	0.0000
MatrixProfile	Wavelet	206.7854	0.0000
SSA	Wavelet	199.0439	0.0000
CFD	Variance	197.3767	0.0000
Autoperiod	Wavelet	183.1295	0.0000
Lomb	Wavelet	163.1445	0.0000
AIC	CFD	158.6722	0.0000
SAZED	Wavelet	158.2849	0.0000
CFD	Spectral	156.9286	0.0000
CFD	FFT	155.6836	0.0000
FFT	Wavelet	151.0573	0.0000
CFD	SAZED	150.1562	0.0000
AIC	Wavelet	150.0066	0.0000
Spectral	Wavelet	150.0066	0.0000
ACF	Variance	144.3101	0.0000
CFD	Lomb	141.7423	0.0000
Autoperiod	CFD	120.1655	0.0000
Variance	Wavelet	113.0087	0.0000
STL	Variance	108.0584	0.0000
ACF	AIC	107.4050	0.0000
ACF	Spectral	107.4050	0.0000
CFD	SSA	106.2901	0.0000
ACF	FFT	104.4153	0.0000
ACF	SAZED	97.0874	0.0000
ACF	Lomb	90.4712	0.0000
SSA	Variance	80.5213	0.0000
AIC	STL	72.3419	0.0000
Spectral	STL	72.3419	0.0000
ACF	Autoperiod	72.3049	0.0000
FFT	STL	69.4825	0.0000

method1	method2	chi_sq	p_value
MatrixProfile	Variance	61.6050	0.0000
CFD	STL	56.0777	0.0000
Autoperiod	Variance	55.5104	0.0000
Lomb	STL	55.1471	0.0000
SAZED	STL	54.8108	0.0000
ACF	SSA	53.6351	0.0000
CFD	MatrixProfile	51.0751	0.0000
AIC	SSA	45.4545	0.0000
Lomb	Variance	45.3065	0.0000
ACF	CFD	39.9452	0.0000
Spectral	SSA	39.6825	0.0000
FFT	SSA	39.4464	0.0000
Autoperiod	STL	34.3750	0.0000
FFT	Variance	33.0652	0.0000
Spectral	Variance	31.6098	0.0000
AIC	Variance	30.1395	0.0000
AIC	MatrixProfile	29.9235	0.0000
MatrixProfile	Spectral	28.9735	0.0000
SAZED	SSA	28.8000	0.0000
SAZED	Variance	28.7356	0.0000
FFT	MatrixProfile	27.6978	0.0000
Lomb	SSA	25.9286	0.0000
Autoperiod	Spectral	21.9661	0.0000
SSA	STL	20.5000	0.0000
MatrixProfile	SAZED	20.1117	0.0000
Lomb	MatrixProfile	18.2528	0.0000
AIC	Autoperiod	17.7534	0.0000
Autoperiod	FFT	16.0147	0.0001
AIC	Lomb	11.1304	0.0008
ACF	MatrixProfile	8.7414	0.0031
FFT	Lomb	8.4500	0.0037
Lomb	Spectral	8.2581	0.0041
Autoperiod	MatrixProfile	7.7784	0.0053
ACF	STL	7.1129	0.0077
Autoperiod	SAZED	6.4533	0.0111
Autoperiod	Lomb	6.0167	0.0142

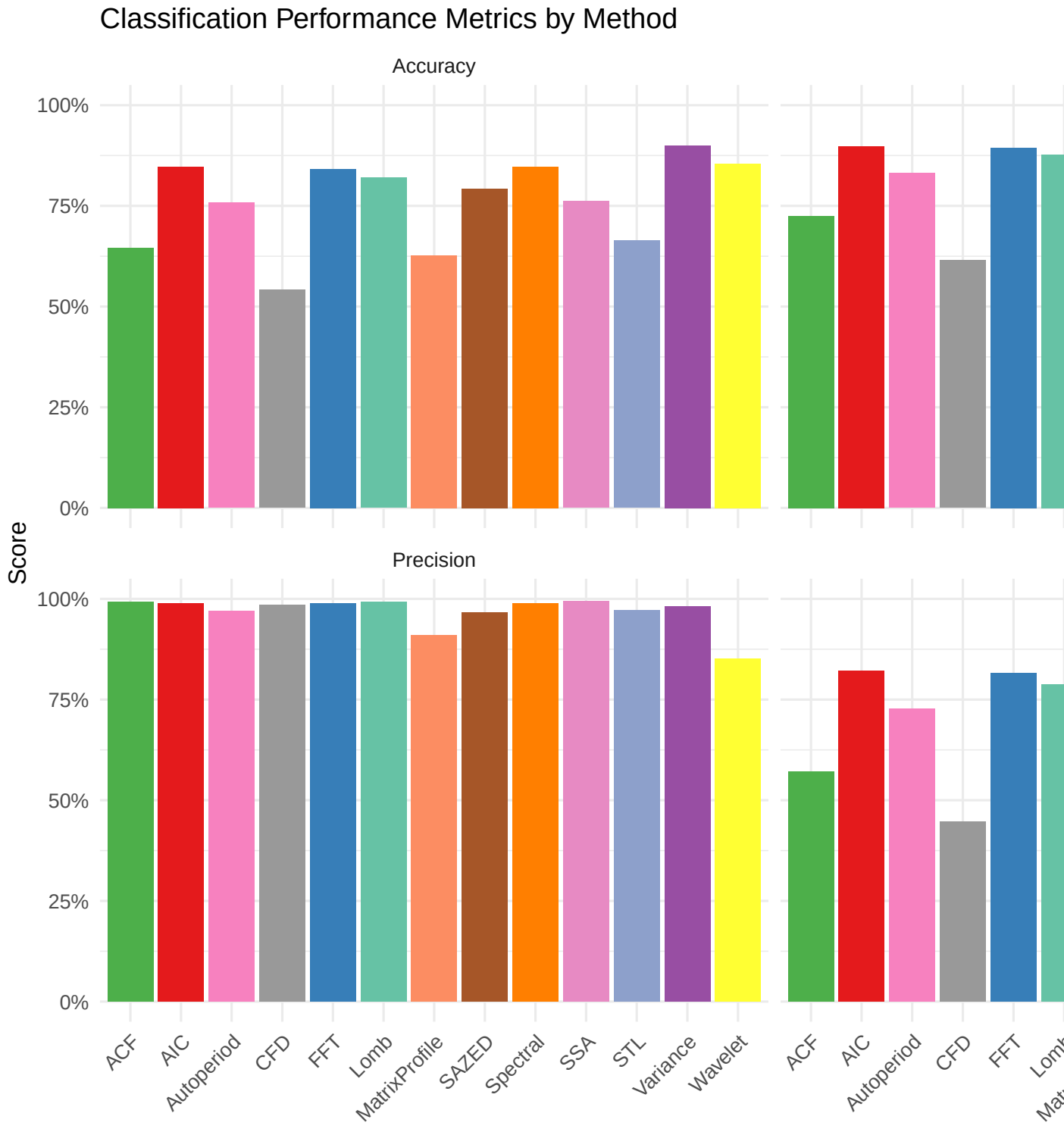

Figure 6: Classification metrics comparison across methods

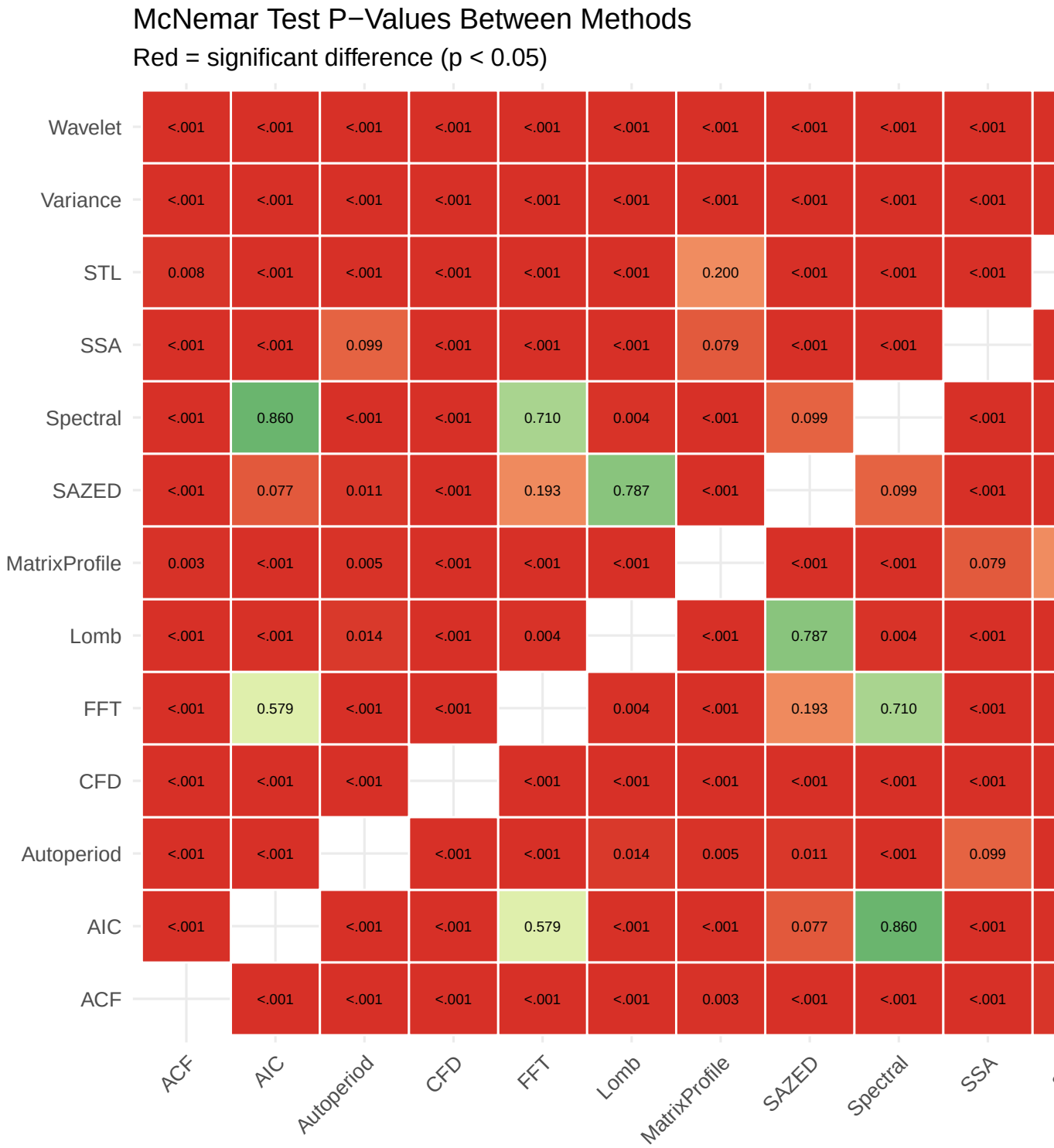

Figure 7: McNemar test p-values between method pairs (red = significant difference)

McNemar P-Value Heatmap

Challenge Scenarios

Following the fdars benchmark, we test method robustness under challenging conditions.

Challenge 1: Linear Trends

```
Generated 150 curves with trends
```

Challenge 2: Red Noise (AR(1) Process)

```
Generated 150 curves with AR(1) noise
```

Challenge 3: Outlier Contamination

```
Generated 150 curves with outliers
```

Challenge Scenario Performance

Table 5: Method Performance Under Challenge Scenarios

Method	Scenario	AUC	F1
Variance	Outliers	NA	NA
Wavelet	Outliers	NA	NA
FFT	Outliers	NA	NA
ACF	Outliers	NA	NA
Variance	Red Noise	NA	NA
Wavelet	Red Noise	NA	NA
FFT	Red Noise	NA	NA
ACF	Red Noise	NA	NA
Variance	Trends	NA	NA
Wavelet	Trends	NA	NA
FFT	Trends	NA	NA
ACF	Trends	NA	NA

Summary and Conclusions

Final Rankings

Table 6: Final Method Rankings by F1 Score

Rank	Method	AUC	F1	Optimal Threshold	Sensitivity	Specificity
1	Variance	0.962	0.936	0.215	0.896	0.92
2	Wavelet	0.608	0.918	0.991	0.996	0.22
3	Spectral	0.952	0.898	0.335	0.822	0.96
4	AIC	0.937	0.898	0.213	0.822	0.96
5	FFT	0.935	0.894	0.063	0.816	0.96
6	Lomb	0.931	0.877	0.570	0.787	0.97
7	SAZED	0.858	0.859	0.794	0.773	0.88
8	Autoperiod	0.863	0.831	0.202	0.727	0.90
9	SSA	0.892	0.831	0.425	0.713	0.98
10	STL	0.801	0.747	0.501	0.607	0.92
11	MatrixProfile	0.719	0.726	0.229	0.604	0.73
12	ACF	0.782	0.725	0.268	0.571	0.98
13	CFD	0.738	0.615	0.268	0.447	0.97

Key Findings

```
**Best Overall Method**: Variance (F1 = 0.936, AUC = 0.962)
```

Recommendations

Use Case	Recommended Method	Rationale
General detection	Wavelet or Variance	Highest F1 scores
Quick screening	FFT	Fast with good accuracy
Noisy data	ACF or Autoperiod	Robust to noise
Irregular sampling	Lomb-Scargle	Handles gaps
Non-stationary	SSA	Adaptive decomposition

Cleanup

```
[1] 0
```

```
[1] 0
```

Session Info

```
R version 4.5.2 (2025-10-31)
Platform: x86_64-pc-linux-gnu
Running under: Manjaro Linux

Matrix products: default
BLAS:   /usr/lib/libblas.so.3.12.0
LAPACK: /usr/lib/liblapack.so.3.12.0  LAPACK version 3.12.0

locale:
 [1] LC_CTYPE=de_DE.UTF-8       LC_NUMERIC=C
 [3] LC_TIME=de_DE.UTF-8        LC_COLLATE=de_DE.UTF-8
 [5] LC_MONETARY=de_DE.UTF-8    LC_MESSAGES=de_DE.UTF-8
 [7] LC_PAPER=de_DE.UTF-8       LC_NAME=C
 [9] LC_ADDRESS=C               LC_TELEPHONE=C
[11] LC_MEASUREMENT=de_DE.UTF-8 LC_IDENTIFICATION=C

time zone: Europe/Berlin
tzcode source: system (glibc)

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
[1] pROC_1.19.0.1 scales_1.4.0  knitr_1.51    purrr_1.2.0   tidyr_1.3.2
[6] dplyr_1.1.4   ggplot2_4.0.1 duckdb_1.4.3  DBI_1.2.3

loaded via a namespace (and not attached):
 [1] gtable_0.3.6       jsonlite_2.0.0     compiler_4.5.2     tidyselect_1.2.1
 [5] Rcpp_1.1.0         yaml_2.3.12        fastmap_1.2.0      R6_2.6.1
 [9] labeling_0.4.3     generics_0.1.4     tibble_3.3.0       pillar_1.11.1
[13] RColorBrewer_1.1-3 rlang_1.1.6        xfun_0.54          S7_0.2.0
[17] otel_0.2.0         cli_3.6.5          withr_3.0.2        magrittr_2.0.4
[21] digest_0.6.39      grid_4.5.2         lifecycle_1.0.4    vctrs_0.6.5
[25] evaluate_1.0.5     glue_1.8.0         farver_2.1.2       codetools_0.2-20
[29] rmarkdown_2.30     tools_4.5.2        pkgconfig_2.0.3    htmltools_0.5.9
```