

METHODOLOGY REPORT

Arterial Congestion and Delay Ranking Methodology

From a Composite Severity Index to Estimated Annual Vehicle-Hours of Delay (EAVHD)

Author

Soheil Sameti

Date

August 2026

Companion methodology to the CBI Multi-Corridor Technical Report

Contents

1 Purpose of the Analysis	3
2 Unit of Analysis: TMC Segments to Named Arterials	3
3 Free-Flow Baseline	4
3.1 Regional free-flow hour	4
3.2 Segment-specific free-flow speed	4
4 Congestion Intensity: Speed Reduction	4
5 Recurrence	5
6 Why the Original Severity Index Was Reconsidered	5
7 Estimated Annual Vehicle-Hours of Delay (EAVHD)	6
8 Excess Travel Time	6
9 The Missing Hourly-Volume Problem	7
10 Hourly Distribution Factor (HDF)	7
11 Directional Split	7
12 Calculating EAVHD	8
13 Total EAVHD Versus EAVHD Density	8
14 Final Ranking Variable	9
15 Congestion Operating Patterns (COPs)	9
16 Why COP Is Not Used as an HDF	10
17 HDF Sensitivity Testing	10
18 Road-Name Quality Control	11
19 Short-Corridor Denominator Problem	11
20 Delay-Concentration Quality Control	11
21 Proposed Corridor Eligibility Rule	12
22 GA-11 and GA-205 Diagnostic Example	12
22.1 GA-11	12
22.2 GA-205	13
23 Remaining Recurrence Audit	13

24 What the Methodology Can and Cannot Claim	13
25 Workflow Summary	14
26 Recommended Public Interpretation	14
27 Current Assessment and Remaining Work	15
References and Data Sources	16

1. Purpose of the Analysis

The purpose of this analysis is to identify and rank arterial roads in the Atlanta region that experience the most significant recurring congestion and estimated traveler delay.

The analysis uses INRIX/NPMRDS-style probe data. The available database does **not** contain individual vehicle trajectories or observed hourly traffic counts. Instead, it contains:

- 5-minute speed observations for roadway TMC segments;
- travel time;
- historical average speed;
- reference speed;
- TMC roadway metadata;
- segment length; and
- AADT (Average Annual Daily Traffic).

Because of these limitations, the methodology is built around what probe data measure well: **when traffic slows, how much it slows, how persistent the condition is, and how much additional travel time results**. AADT is then used as an exposure measure to estimate how many vehicles are affected by that delay.

The final product is intended as a **regional arterial congestion screening and prioritization tool**. It is not a signal-timing model, an HCM intersection analysis, or a substitute for observed turning-movement counts.

2. Unit of Analysis: TMC Segments to Named Arterials

The original INRIX/NPMRDS data are organized by individual TMC roadway segments. Individual TMCs are often too small to represent a meaningful arterial corridor, and the same road can appear under several text variants or route-number combinations.

The workflow therefore has two levels:

1. Calculate congestion and delay measures at the **individual TMC level**.
2. Aggregate TMCs into a curated **named arterial corridor** before ranking.

A curated alias registry consolidates fragmented road names and co-signed route descriptions into one canonical road name. This cleanup has already corrected duplicate-corridor cases such as:

- N Druid Hills Rd / N Druid Hills Rd NE;
- Camp Creek Pkwy / Camp Creek Pkwy SW; and
- Windward Pkwy / Windward Pkwy W.

This step matters because duplicate road names can split mileage, EAVHD, congestion pattern, and rank across multiple rows.

The final ranking represents **named arterial corridors**, not individual intersections. The “Worst Point” field is a geographic reference identifying the most problematic location associated with that corridor; it is not the unit being ranked.

3. Free-Flow Baseline

A central challenge is defining what counts as uncongested or free-flow operation.

If free-flow speed were calculated independently for every segment using that segment’s own best observed conditions, a chronically congested segment could establish an artificially low baseline and therefore appear less abnormal. To avoid that problem, the analysis first identifies a common **regional free-flow hour**.

3.1. Regional free-flow hour

Average speed is calculated by hour of day across all arterial-classified segments, using the arterial functional-class subset. The hour with the highest average arterial speed is selected once for the region.

In the current dataset:

$$\text{Regional free-flow hour} = 4:00 \text{ AM}$$

3.2. Segment-specific free-flow speed

The common hour determines *when* free flow is measured, but each segment retains its own baseline speed. For TMC i :

$$V_{\text{FF},i} = V_{i,4\text{AM}}$$

Thus, a road that normally operates near 35 mph is not assigned the same free-flow speed as a road that normally operates near 55 mph. They simply share the same regionally selected reference period.

4. Congestion Intensity: Speed Reduction

Congestion intensity measures how severely each segment slows during the main commuting periods.

The current peak windows are:

$$AM = 7:00\text{--}9:00$$

$$PM = 16:00\text{--}18:00$$

For each TMC, the minimum average speed within the AM or PM peak window is compared with its own free-flow speed.

The speed ratio is:

$$\boxed{SpeedRatio_i = \frac{V_{peak,i}}{V_{FF,i}}}$$

The reported speed reduction is:

$$\boxed{SpeedReduction_i = 1 - SpeedRatio_i}$$

For example, if free-flow speed is 40 mph and the relevant peak speed is 20 mph:

$$SpeedRatio = \frac{20}{40} = 0.50$$

$$SpeedReduction = 1 - 0.50 = 50\%$$

A larger speed reduction therefore represents more intense congestion.

5. Recurrence

Intensity alone cannot distinguish a corridor that is severely congested only on a few unusual days from one that experiences similar congestion nearly every weekday.

A segment is considered to experience significant recurring peak congestion when its peak-period speed falls below 70% of its free-flow baseline:

$$\boxed{V_{peak} < 0.70V_{FF}}$$

Recurrence is the percentage of analyzed weekdays on which that condition occurs.

At the road level, recurrence should be mileage weighted so that very short TMCs do not receive the same influence as long segments:

$$\boxed{R_{road} = \frac{\sum_i L_i R_i}{\sum_i L_i}}$$

Many current Top-20 corridors have recurrence values near 100%. This may legitimately indicate that the worst corridors are congested on nearly every analyzed weekday, but recurrence should also be reviewed across the full set of roads to confirm that the measure is not saturating region-wide.

6. Why the Original Severity Index Was Reconsidered

An earlier version of the methodology ranked roads using a composite Severity Index based on queue mile-hours, recurrence, speed reduction, and AADT:

$$Severity \propto QueueDensity \times Recurrence \times SpeedReduction \times AADT$$

This approach was useful for screening, but it had two major weaknesses:

1. The final score was abstract and had no direct physical interpretation.
2. Summed queue burden favored long roads because more mileage created more opportunities to accumulate queue mile-hours.

The first structural fix was to normalize queue burden by analyzed road mileage. The methodology then moved to a more interpretable impact measure: **Estimated Annual Vehicle-Hours of Delay (EAVHD)**.

The old Severity Index is now retained only as an experimental comparison field and **does not determine rank**.

7. Estimated Annual Vehicle-Hours of Delay (EAVHD)

The principal delay measure is:

$$\boxed{\text{Estimated Annual Vehicle-Hours of Delay (EAVHD)}}$$

The construction is **inspired by the general exposure-weighted excess-travel-time logic used in FHWA performance measures such as PHED**, but it is not the official FHWA PHED metric. The CBI implementation uses a segment-specific free-flow baseline and the available INRIX/NPMRDS data structure.

The idea is straightforward:

Estimate the additional travel time experienced under observed conditions relative to free flow, estimate the number of vehicles exposed to those conditions, and accumulate that delay across the analysis period.

8. Excess Travel Time

For segment i of length L_i :

$$TT_{obs,i} = \frac{L_i}{V_{obs,i}}$$

and:

$$TT_{FF,i} = \frac{L_i}{V_{FF,i}}$$

Excess travel time is therefore:

$$\boxed{ETT_{i,h,d} = L_i \times \max\left(0, \frac{1}{V_{obs,i,h,d}} - \frac{1}{V_{FF,i}}\right)}$$

where h is hour of day and d is analyzed weekday.

The $\max(0, \cdot)$ term prevents speeds above the free-flow baseline from creating negative delay.

This is a **continuous delay measure**. It differs intentionally from recurrence:

- **Recurrence** uses the 70% threshold to describe how frequently severe congestion occurs.
- **EAVHD** measures excess travel time whenever observed travel time exceeds the free-flow benchmark.

9. The Missing Hourly-Volume Problem

INRIX/NPMRDS probe data provide detailed temporal speed and travel-time observations, but the current database does not contain observed hourly traffic volume.

The metadata include AADT:

$$AADT = \text{Average Annual Daily Traffic}$$

AADT does not identify how many vehicles travel in a particular hour. Therefore, hourly traffic exposure must currently be estimated.

10. Hourly Distribution Factor (HDF)

Estimated directional hourly volume is:

$$\hat{V}_{i,h} = AADT_i \times HDF_h \times D$$

where:

- $AADT_i$ is the AADT of TMC i ;
- HDF_h is the modeled fraction of daily traffic occurring in hour h ; and
- D is the directional split.

The current HDF comes from the Delaware Department of Transportation Traffic Monitoring Program, *Hourly Distribution of AADT for Traffic Groups* (Average Weekday Traffic, 2002).

The available table contained seven populated traffic-pattern groups plus one empty group. Because the available source did not include a sufficiently clear group-to-facility legend to confidently select one column as the correct Atlanta urban-arterial profile, the primary curve uses the average of the seven populated groups.

This is explicitly treated as a **modeled assumption**, not as an observed Atlanta hourly volume profile.

11. Directional Split

The probe TMCs are directional, but the current database does not provide hourly directional traffic counts. The analysis therefore uses:

$$D = 0.50$$

This assumes that half of the daily roadway traffic belongs to each direction.

The 50/50 factor is a transparent neutral approximation, not a claim that every Atlanta arterial is directionally balanced. It should be replaced if reliable directional count information becomes available.

12. Calculating EAVHD

For each TMC, hour, and analyzed weekday:

$$Delay_{i,h,d} = ETT_{i,h,d} \times \hat{V}_{i,h}$$

This quantity is measured in vehicle-hours of excess travel time.

The segment total is:

$$EAVHD_i = \sum_d \sum_h ETT_{i,h,d} \hat{V}_{i,h}$$

The road total is then:

$$EAVHD_{road} = \sum_{i \in road} EAVHD_i$$

Each TMC should use its **own segment-level AADT** in the EAVHD calculation. Median AADT at the road level is only a descriptive statistic shown in the output table.

13. Total EAVHD Versus EAVHD Density

Total EAVHD measures overall estimated traveler burden, but raw totals naturally increase with corridor length. A 40-mile road has many more miles over which delay can accumulate than a 2-mile road.

Therefore, the ranking uses delay density:

$$EAVHD_{density} = \frac{EAVHD_{road}}{AnalyzedRoadMiles}$$

with units:

$$\text{estimated vehicle-hours of delay per mile}$$

This removes the direct accumulation advantage associated with longer corridors and expresses delay burden on a per-mile basis.

The two outputs answer different questions:

Total EAVHD

How much total estimated traveler delay does this road generate?

EAVHD per mile

How concentrated is the estimated delay burden along this road?

14. Final Ranking Variable

The ranking is explicitly:

$$\boxed{Rank = Descending(EAVHD_{density})}$$

It is **not** ranked by the experimental Severity Index. This is why the Severity Index column does not appear in descending order.

For public reporting, the Severity Index should eventually be removed from the main ranking table or moved to a separate validation/comparison table.

15. Congestion Operating Patterns (COPs)

A separate analysis describes **what type of congestion pattern** each road exhibits. These categories are called Congestion Operating Patterns (COPs).

For each TMC, a normalized 24-hour speed profile is created:

$$\boxed{R_{i,h} = \frac{V_{i,h}}{V_{FF,i}}}$$

This allows roads with different absolute speeds to be compared on the basis of temporal congestion shape.

K-means and hierarchical clustering were tested for multiple values of k . The best tested result was:

$$\boxed{\text{K-means, } k = 4, \text{ silhouette} = 0.293}$$

Other tested results included:

Method	k	Silhouette
K-means	4	0.293
Hierarchical	4	0.280
K-means	5	0.275
Hierarchical	5	0.206
K-means	6	0.267
Hierarchical	6	0.216

The selected four-cluster solution produced approximately:

- 270 mostly uncongested segments;
- 1,000 PM-dominant segments;
- 902 irregular/mixed-pattern segments; and
- 664 all-day urban congestion segments.

Because the silhouette score is modest, COPs should be interpreted as **descriptive archetypes**, not sharply separated natural regimes.

Examples in the current ranking include:

- COP-1 - All-Day Urban;
- COP-2 - PM-Dominant; and
- COP-5 - Irregular/Mixed.

COP is a diagnostic field only and does not affect EAVHD or rank.

16. Why COP Is Not Used as an HDF

The separation between COP and HDF is deliberate.

INRIX probe data tell us how **speed** changes throughout the day. They do not tell us how **traffic volume** is distributed throughout the day.

A strong PM speed decline does not imply that a known percentage of AADT occurs in the PM. Near capacity, speed may fall while actual throughput also falls.

Therefore:

$$\text{Speed profile} \neq \text{Volume profile}$$

The Atlanta COP classification is not used to manufacture a traffic-volume distribution.

17. HDF Sensitivity Testing

Because the HDF is an acknowledged modeling assumption, the entire EAVHD-density ranking was recalculated under alternative published hourly-distribution shapes.

The results were highly stable:

Scenario	Spearman rank correlation	Top-20 overlap
Flatter published curve	0.999	20/20
More sharply peaked curve	1.000	19/20

These results do **not** prove that the absolute EAVHD values are exact. They demonstrate that the **relative ranking is highly insensitive to the particular reasonable published HDF curve tested**.

Accordingly, confidence should be stated separately:

Relative ranking: robust under tested HDF alternatives

Absolute EAVHD magnitude: modeled / estimated

18. Road-Name Quality Control

Road-name normalization is a substantive part of the methodology because alias fragmentation can distort total mileage, EAVHD, delay density, and rank.

After alias cleanup, for example:

- N Druid Hills Road became one approximately 8.0-mile analyzed corridor rather than two separate ranked entries.
- Camp Creek Parkway became one approximately 9.7-mile corridor rather than occupying two positions in the Top 20.

Canonical road grouping should therefore be reviewed whenever duplicate or near-duplicate road names appear in the rankings.

19. Short-Corridor Denominator Problem

Length normalization fixes the bias toward very long roads, but it introduces the opposite potential artifact:

$$Density = \frac{EAVHD}{Miles}$$

can become unstable when the denominator is extremely small.

GA-205 exposed this issue. It contains only two TMCs totaling approximately 0.538 miles, with total EAVHD of roughly 29,782 vehicle-hours. Dividing by the small denominator produces approximately 55,380 vehicle-hours per mile, enough to place the road near the Top 20.

By contrast, GA-11 contains 15 TMCs across approximately 8.06 miles and produces roughly 446,396 vehicle-hours, or about 55,410 vehicle-hours per mile. The two densities are nearly identical, but the amount of corridor evidence supporting them is very different.

20. Delay-Concentration Quality Control

A useful internal QC statistic is the share of corridor EAVHD contributed by the single largest TMC:

$$MaxTMCShare = \frac{\max(EAVHD_i)}{EAVHD_{road}}$$

For GA-11, the largest TMC contributes only about 16.7% of corridor EAVHD, indicating that the burden is distributed across many segments.

For GA-205, one TMC contributes about 82.6% of the total, indicating that the result is dominated by one localized bottleneck rather than broad corridor-wide evidence.

A preliminary QC flag can therefore be defined as:

$$\boxed{MaxTMCShare > 0.60}$$

This should initially be used as a **review flag**, not an automatic exclusion rule.

21. Proposed Corridor Eligibility Rule

Rather than manually deleting suspicious roads, an objective coverage rule should determine whether a named road has sufficient evidence to be ranked as a corridor.

The proposed rule is:

$$\boxed{Miles \geq 0.5 \quad \text{AND} \quad (Miles \geq 1.0 \text{ OR } N_{TMC} \geq 4)}$$

This rule is designed to preserve genuinely short but well-represented urban corridors while excluding tiny fragments supported by only one or two TMCs.

Examples:

- A 0.7-mile downtown corridor represented by six TMCs can remain eligible.
- A 5-mile road represented by three long TMCs can remain eligible.
- A 0.54-mile road represented by only two TMCs, such as the current GA-205 case, would not qualify for the corridor ranking.

This is preferable to manually removing roads because they do not “look right.”

22. GA-11 and GA-205 Diagnostic Example

The detailed QC review of GA-11 and GA-205 illustrates why support size and delay concentration matter.

22.1. GA-11

GA-11 contains 15 TMCs covering approximately 8.06 miles. Several substantial TMCs independently show severe speed deterioration, with multiple speed ratios in roughly the 0.40–0.51 range. Multiple segments independently contribute tens of thousands of estimated vehicle-hours.

Its largest segment contributes only about 16.7% of total corridor EAVHD. This supports interpretation as a **distributed corridor-level congestion problem**.

22.2. GA-205

GA-205 contains only two TMCs totaling approximately 0.54 mile. One direction is relatively mild while the other is severely degraded, and approximately 82.6% of total corridor EAVHD comes from a single short TMC.

The congestion may be real, but the available coverage is more consistent with a **localized bottleneck** than with a robust named-corridor ranking. Under the proposed eligibility rule, GA-205 would be excluded from the corridor ranking without any manual exception.

23. Remaining Recurrence Audit

One specific QC issue remains to be checked. A GA-205 TMC had an aggregate peak speed ratio around 0.797 while its reported recurrence was 100%.

These statistics are not necessarily impossible because they may use different temporal aggregations, but the implementation should be verified to ensure recurrence is exactly:

$\frac{\text{valid weekdays whose defined peak statistic is below } 0.70V_{FF}}{\text{all valid analyzed weekdays}}$
--

and not accidentally based on whether *any* individual 5-minute observation within a peak period crossed the threshold.

The recurrence distribution should also be summarized across all named arterials, not only the Top 20.

24. What the Methodology Can and Cannot Claim

The analysis can reasonably claim to use observed INRIX probe data to characterize:

- speed deterioration;
- temporal congestion patterns;
- recurrence;
- excess travel time;
- spatial concentration of congestion; and
- relative arterial delay burden.

It can also estimate traffic exposure using segment-level AADT.

However, EAVHD should **not** be described as directly measured vehicle-hours of delay because hourly and directional traffic volumes are modeled rather than observed.

The word **Estimated** is therefore essential. Absolute values depend partly on the chosen HDF and the current $D = 0.50$ assumption.

25. Workflow Summary

The complete process can be summarized as follows:

1. **Observe traffic conditions.** Use INRIX/NPMRDS probe speeds for arterial TMCs.
2. **Establish a common free-flow period.** Identify the hour when the regional arterial system collectively operates fastest; currently 4 AM.
3. **Establish individual segment baselines.** Use each TMC's own speed at that regional free-flow hour.
4. **Measure congestion intensity.** Compare peak-period speeds with each segment's free-flow baseline.
5. **Measure recurrence.** Determine how frequently weekday peak speeds fall below 70% of free flow.
6. **Calculate excess travel time.** Compute observed travel time minus free-flow travel time for each segment and hour.
7. **Estimate hourly traffic exposure.** Convert segment AADT into estimated directional hourly volume using HDF and $D = 0.50$.
8. **Calculate EAVHD.** Multiply excess travel time by estimated traffic exposure and accumulate across weekdays.
9. **Build named corridors.** Combine TMCs under canonical road names.
10. **Measure total impact.** Sum segment EAVHD to obtain road-level total EAVHD.
11. **Normalize for road length.** Divide total EAVHD by analyzed road miles.
12. **Apply corridor coverage QC.** Exclude tiny, poorly represented fragments from the named-corridor ranking.
13. **Rank eligible roads.** Sort from highest to lowest EAVHD per mile.
14. **Describe congestion type.** Assign a descriptive COP based on normalized 24-hour speed shape.
15. **Test robustness.** Re-run the ranking under alternate published HDF curves and compare ranks using Spearman correlation and Top-20 overlap.

26. Recommended Public Interpretation

The final ranking should be described as:

A ranking of Atlanta-region arterial corridors by estimated annual vehicle-delay burden per analyzed roadway mile.

It should not be described as a ranking of the worst intersections, because the unit of analysis is the named arterial corridor.

Both total and density measures should be retained:

$$EAVHD_{total}$$

for **total regional impact**, and

$$EAVHD_{density}$$

for **congestion concentration and ranking**.

27. Current Assessment and Remaining Work

The methodology has evolved from an abstract composite severity score into a more interpretable exposure-weighted delay framework. Its strongest components are:

- INRIX-derived speed and travel-time observations;
- a consistent region-wide free-flow reference hour with segment-specific baseline speeds;
- road-name canonicalization;
- segment-level EAVHD aggregation;
- length normalization; and
- strong rank stability under alternate published HDF assumptions.

The largest remaining uncertainty is not the relative ranking itself; the sensitivity testing indicates that ranking is highly stable under the HDF alternatives tested. The larger uncertainty is the **absolute magnitude of EAVHD**, because hourly and directional traffic volumes are estimated rather than directly observed.

The next quality-control steps are therefore:

1. implement the corridor eligibility rule;
2. retain TMC count and maximum-TMC EAVHD share as QC fields;
3. verify the recurrence calculation and full-network recurrence distribution;
4. continue canonical road-name cleanup; and
5. eventually replace the current HDF and directional assumptions if reliable Atlanta/Georgia continuous-count data become available.

Once these checks are complete, the methodology is defensible as a **regional arterial congestion screening and prioritization framework** based primarily on INRIX/NPMRDS probe observations and segment-level AADT.

References and Data Sources

- INRIX/NPMRDS-style probe readings and TMC metadata used in the CBI arterial analysis database.
- Delaware Department of Transportation, Traffic Monitoring Program. *Hourly Distribution of AADT for Traffic Groups*, Average Weekday Traffic, 2002. Used as the current modeled hourly traffic-distribution input and tested against alternate published curves.
- Federal Highway Administration performance-measure methodology, including PHED concepts based on excess travel time and estimated traffic exposure. The CBI EAVHD measure is an adaptation of the general construction, not the federal PHED metric itself.