Managing matchday demand: flow, incidents and response with SCOOT® 8AI

Study by Chris Kettell, Chief Technologist, TRL Software

Paper presentation: Chris Small, Transport for Greater Manchester and Wendy Davidson, TRL Software

Abstract

This paper presents a live evaluation of incident detection within SCOOT® 8AI in Manchester. The work forms part of an ongoing collaboration between Transport for Greater Manchester (TfGM) and TRL Software - and addresses a difficult practical problem: how can an incident detection be evaluated when there is no complete incident log or manually confirmed ground truth?

A matchday event on 28 January 2026 provides a real-world case study. Police escorted supporters from Manchester city centre towards the Etihad Stadium, crossing the southern part of the Ancoats region and creating a recognisable disturbance in the traffic data. We use that event to examine the temporal and spatial signature of disruption, then consider the wider behaviour of the incident detection.

The result is not a claim of measured incident accuracy. It is a practical method for reducing noisy detector output into a smaller set of physically credible candidate incidents for operator attention, supported by evidence that the strongest filtered events follow recognisable traffic patterns.

Introduction and deployment context

TfGM manages traffic signals across Greater Manchester and works with a network in which disruption can arise from planned events, short-lived obstructions, collisions, breakdowns and recurrent congestion. The SCOOT® 8AI work discussed here combines flow forecasting, network validation, incident detection and public transport analysis within TRL UTC.

For an operator, the first challenge is often recognition. The longer an incident continues without intervention, the longer the network may take to recover. That relationship between disruption and recovery motivates the work in this paper: if an emerging problem can be identified earlier and with sufficient confidence, the operator has more opportunity to investigate it and decide whether intervention is required.

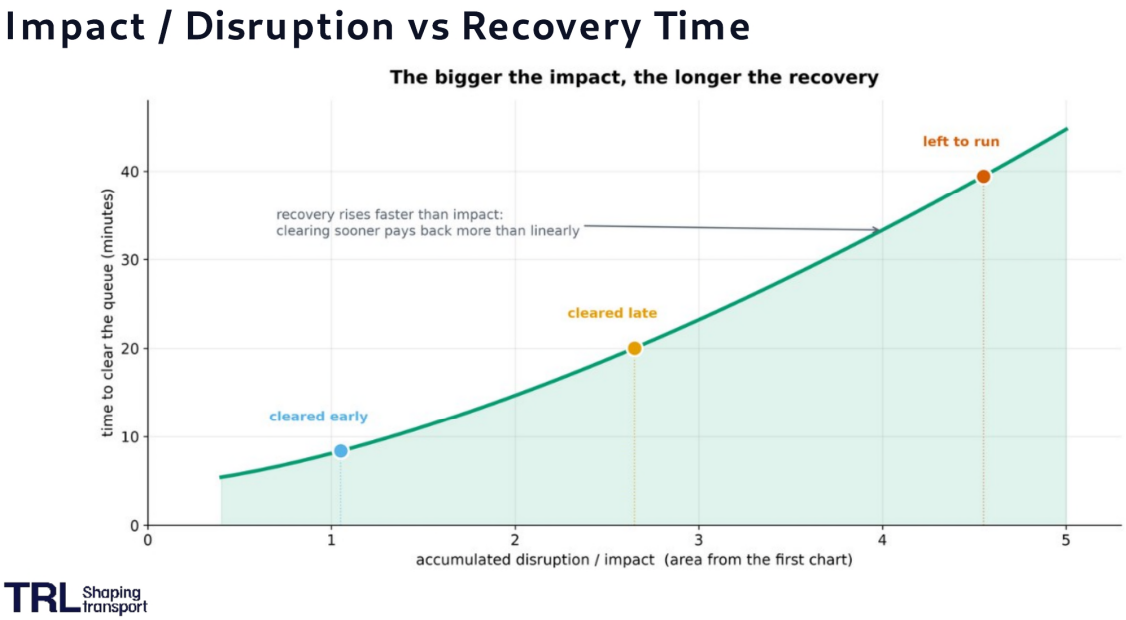


Figure 1. Relationship between disruption duration and recovery time.

Matchday case study: 28 January 2026

On 28th January 2026 Manchester City played Galatasaray at the Etihad Stadium. Before kick-off, supporters were police escorted from Manchester city centre towards the stadium. The walking route crossed the southern part of the Ancoats SCOOT® region and affected a number of links. Because the event was known operationally and produced a visible traffic response, it provides a useful case through which to examine how an incident-like disturbance develops in the data.

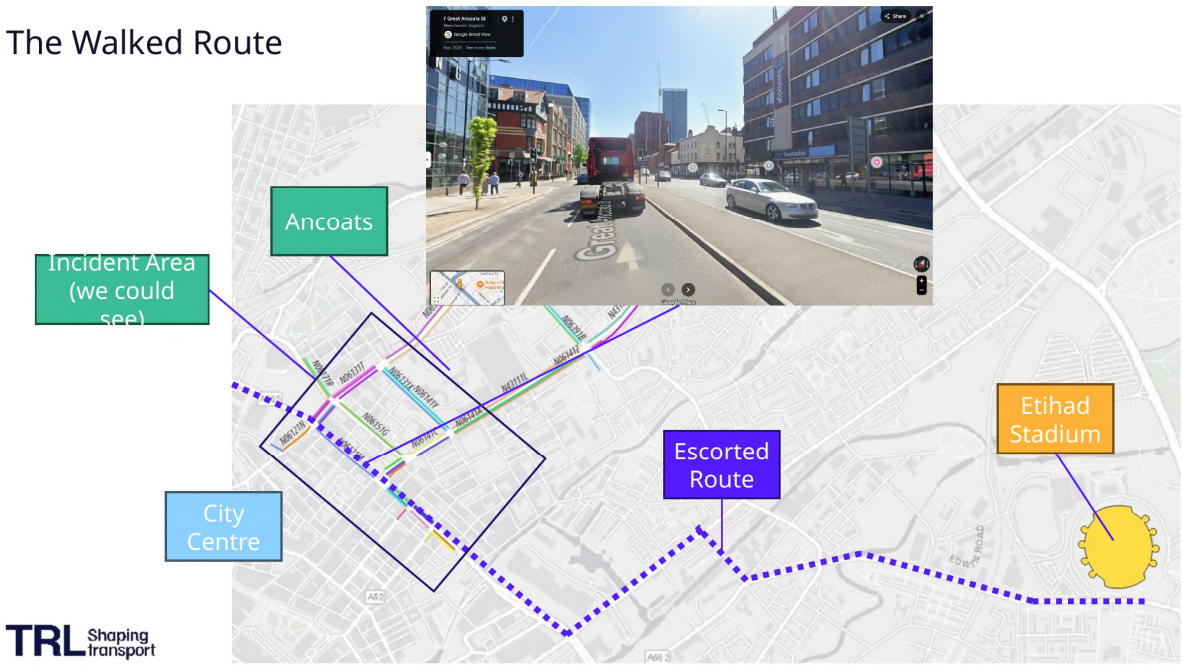


Figure 2. Police-escorted walking route from Manchester city centre through Ancoats to the Etihad Stadium.

Ancoats matchday case: link reference key

Link	Node	Location / description
N06121L	N06121	Swan Street / Great Ancoats Street
N06191N	N06191	Return arm towards N06161
N06171B	N06171	N06171
N06191P	N06191	Great Ancoats Street / Old Mill Street
N06161R	N06161	Old Mill Street
N06131T	N06131	N06131
N06151F	N06151	Old Mill Street / New Islington
N06181J	N06181	N06181

Figure 2a. Link and location reference key for Figures 2 and 3.

Event timeline

The first sustained disturbance appears at 18:43:13 on N06121L at Swan Street / Great Ancoats Street and lasts for approximately two cycles. A further short sustained event follows on N06191N at 18:52-18:53. At 19:08 the main disruption becomes clear: N06191P remains affected for 17 cycles, approximately 23 minutes, and N06161R shows a shorter three-cycle event nested within the same period. A second later wave begins from around 19:27, followed by activity at N06151F and a single-cycle dip at N06181J. The full source timeline, including the unconfirmed single-cycle drops, is retained in Figure 3.

Timeline

Time	Link	Node / location	Duration	Evidence
18:43:13 – 18:45	N06121L	N06121 (Swan St / Gt Ancoats St)	2 cycles (~2 min)	Sustained
18:52 18:53	N06191N	N06191 (return arm to N06161)	2 cycles (~2 min)	Sustained
19:07	N06171B	N06171	1 cycle	Single-cycle dip only (99% drop, unconfirmed)
19:08 - 19:30	N06191P	N06191 (Gt Ancoats St / Old Mill St)	17 cycles (~23 min)	Sustained, longest and clearest signature
19:12 – 19:15	N06161R	N06161 (Old Mill St)	3 cycles (~3 min)	Sustained, nested inside N06191P's window
19:27	N06131T	N06131	1 cycle	Single-cycle dip only (81% drop, unconfirmed)
19:27 – 19:3	N06191N	N06191 (return arm to N06161)	8 cycles (~10 min)	Sustained, second, longer collapse on the same link
19:36 19:38	N06151F	N06151 (Old Mill St / New Islington)	3 cycles (~3 min)	Sustained
19:37:29	N06181J	N06181	1 cycle	Single-cycle dip only (82% drop)

TRL Shaping transport

Figure 3. Event timeline for the Ancoats matchday case study.

Traffic response

The event can be read in four phases. Between 18:43 and 18:53 there is a light, brief disturbance at the Swan Street end. From 19:08 to 19:31 the disruption becomes sustained, with the clearest and longest signature at N06191 and a shorter event at N06161 within the same window. A second distinct wave then develops from around 19:27 to 19:39. The final phase is recovery through to the 20:00 kick-off.

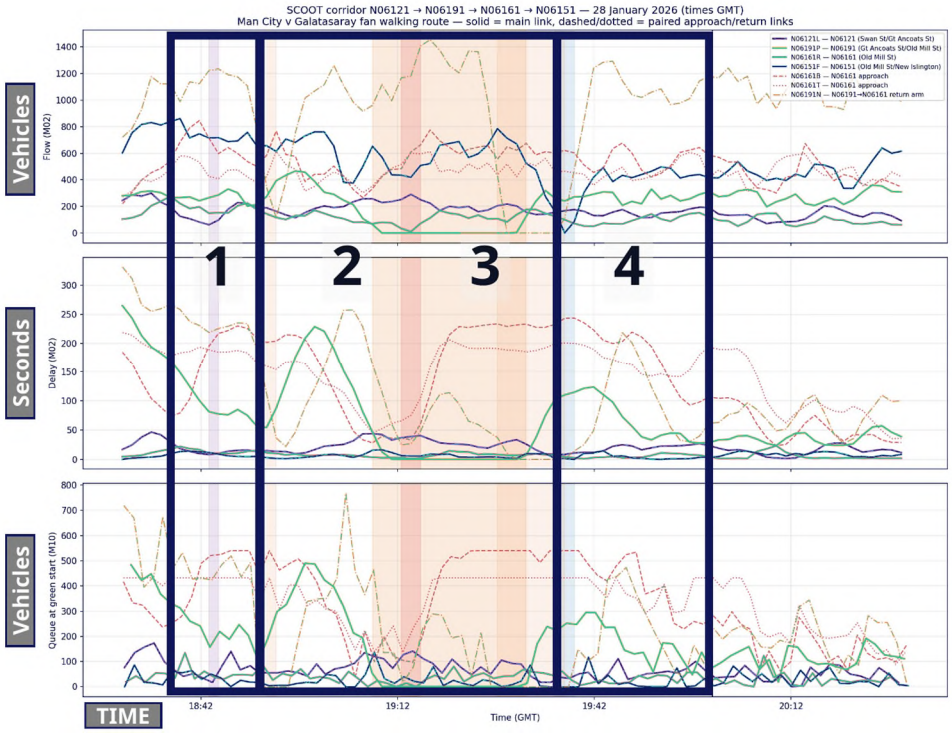


Figure 4. Four phases of the matchday disruption as observed in the incident-detection data.

The traffic-engineering interpretation is link starvation: little or no flow enters the affected link, queues build cycle by cycle, discharge is constrained and the effect begins to spread beyond the initial entry point. This behaviour is consistent with the established SCOOT® congestion model, in which queues can extend backwards into upstream junctions and produce a wider, domino-like loss of capacity across connected links.

Incident detection without ground truth

The matchday case is helpful because the cause of disruption was known and the resulting traffic response could be observed. Most operational incidents do not come with that level of confirmation. The wider evaluation therefore has to answer a harder question: whether the detector output contains credible traffic structure even when no complete incident log exists.

In normal operation, disruption can range from a very short obstruction lasting seconds or minutes to a collision, breakdown, utility response or major event lasting much longer. Impact and duration therefore have to be considered together: even a short interruption can create a significant queue where traffic is already close to capacity.

The central validation problem is the absence of a complete truth set. There is no comprehensive incident log, collision record or manually confirmed reference against which every link-cycle can be labelled as incident or non-incident. At the same time, the default detector operating point is deliberately sensitive and fires frequently, particularly on links affected by recurrent congestion.



Figure 5. Behaviour of the raw incident detector at its default operating point.

Across the analysed data, 42% of all link-cycles are flagged as incidents. Of 113 links, 26 fire in more than 90% of cycles, and these effectively always-on links account for 53% of all firings. Those figures are not plausible as a direct measure of genuine incidents. They show that the detector is identifying difficult or abnormal traffic states but is also conflating repeated congestion with one-off disruption. For that reason, neither the raw binary flag nor severity is sufficient as the operational decision axis. Without ground truth, conventional precision and recall cannot be measured directly.

Label-free validation using traffic physics

The absence of labels changes the validation question. Rather than asking whether every individual alert is correct, we ask whether the detector output contains physical structure that would be difficult for random noise to reproduce. A real blockage should create a recognisable spatial and temporal fingerprint in the surrounding traffic.

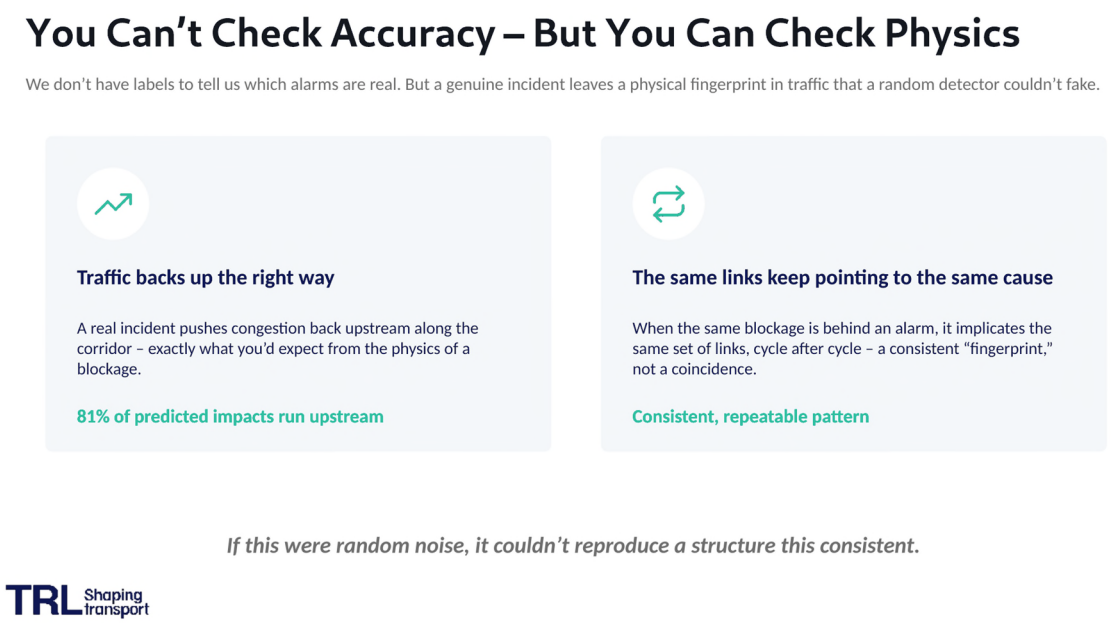


Figure 6. Physical checks used when conventional accuracy cannot be measured directly.

The first check is direction. When a junction or link becomes blocked, congestion should propagate upstream against the direction of travel. In the analysed cascades, 81% of predicted impacts are upstream of the suspected root. The second check is repeatability. If successive alarms are being driven by the same underlying blockage, they should repeatedly implicate the same connected links rather than a different set of links on every cycle.

Propagation coherence

Each cycle produces a predicted cascade: the set of links thought to be affected by a suspected root incident. To test whether those cascades are stable, successive predictions for the same root are compared using Jaccard overlap:

$J(A,B) = |A \cap B| / |A \cup B|$

The numerator is the number of links shared by the two predictions and the denominator is the total number of distinct links appearing in either prediction. A Jaccard value of 1.0 therefore represents identical impact sets, while 0 represents no common links.

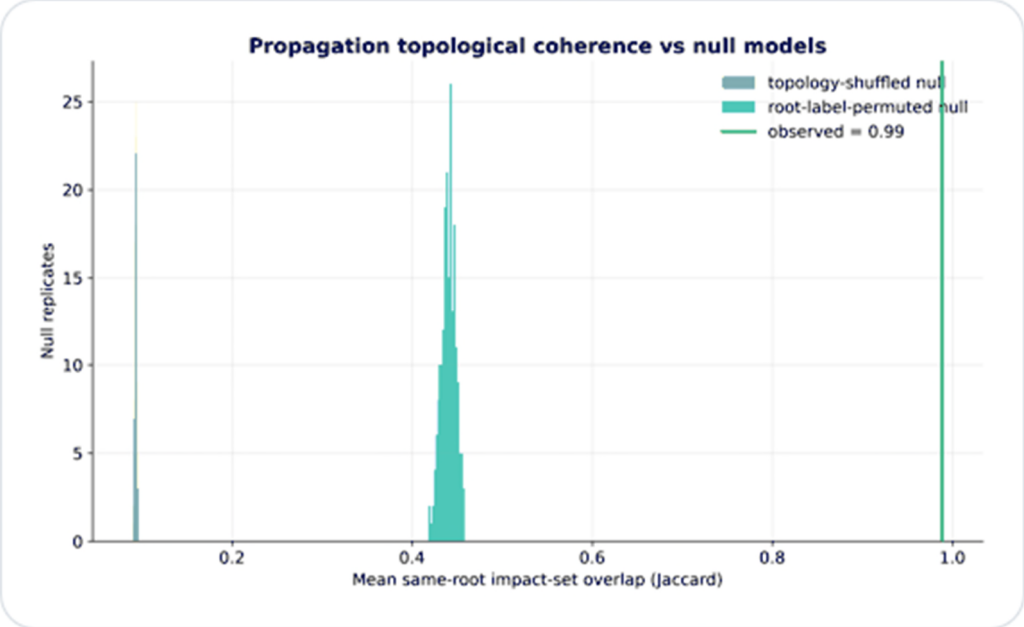


Figure 7. Same-root propagation overlap compared with root-label-permuted and topology-shuffled null models.

The observed same-root overlap is 0.99. Two null models are used as negative controls. In the topology-shuffled null, the road-network structure is rewired so that predicted overlaps can no longer follow the real corridors; the overlap falls to approximately 0.09. In the root-label-permuted null, each cascade is kept intact but reassigned to a different root; the overlap is approximately 0.44. The observed 0.99 result is reported as $p<0.005$ against both nulls.

The root-label-permutation test is deliberately the harder comparison. It preserves each individual cascade and breaks only the association between that cascade and the junction from which it originated. If the 0.99 overlap were simply an artefact of repeatedly recomputing the same incident, this null would be expected to reproduce much more of the observed coherence. It does not. Together with the 81% upstream result, this supports the interpretation that the predicted cascades are tied to specific corridor topology and behave in a way that is consistent with real queue propagation.

One-off events versus recurrent congestion

Propagation alone does not solve the whole problem. The detector must also separate one-off disruption from the broad, repeated congestion patterns that occur on the same links day after day. A genuine one-off blockage should appear, grow through connected links and then drain as the obstruction clears. Recurrent congestion, by contrast, produces a similar broad pattern at similar times on multiple days.

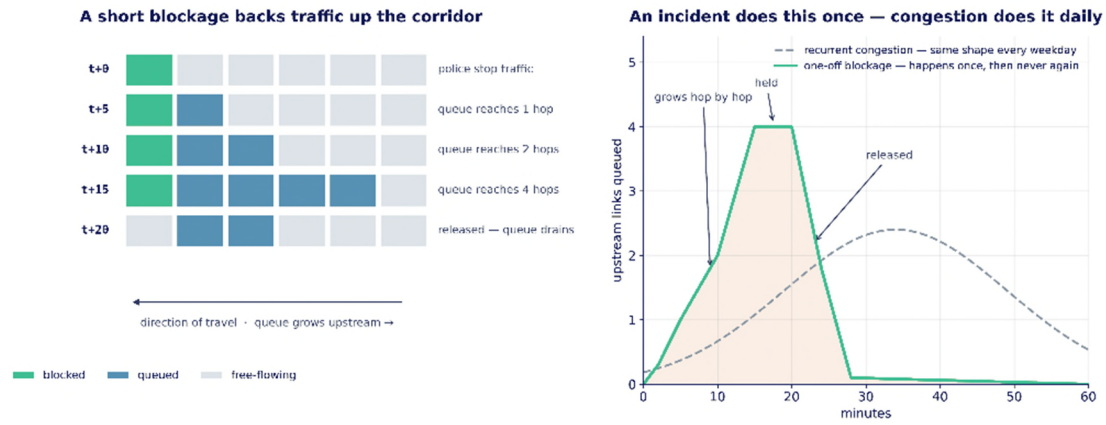


Figure 8. Conceptual difference between a one-off blockage and recurrent congestion.

This distinction provides two further filtering tests. Persistence removes isolated one-cycle fluctuations, while non-recurrence removes links that repeatedly exhibit the same congested behaviour. Coherent propagation is then used as the spatial check. None of these tests requires a complete external incident log.

Building an operational filter

The operating point developed from the analysis therefore combines persistence, non-recurrence and coherent propagation. Severity is deliberately not used as the principal decision axis. Medium-severity firings are concentrated in the most congested cells, indicating that severity is acting as a proxy for congestion and risks reintroducing the very confound the filter is intended to remove.



Figure 9. Three-stage incident filter and resulting reduction in candidate alarms.

The consolidated three-stage result starts with 8,367 raw alarms, retains 1,598 after the persistence stage and reduces the set to 134 after non-recurrence and coherent propagation. This is described as a reduction of roughly 98% achieved from the physical structure of the data, without using an incident ground truth to tune the filter. An earlier intermediate count identified 4,318 persistent episodes before later consolidation rules were applied. The final operational funnel reported here therefore uses the consolidated sequence: 8,367 raw alarms, 1,598 after persistence and 134 final candidates.

Results and validation findings

At its default operating point the detector over-fires, largely because recurrent congestion is repeatedly classified as incident-like behaviour. The strongest validation result comes from coherent propagation: successive cascades associated with the same root show an overlap of 0.99, compared with 0.44 for a root-label-permuted null and 0.09 for a topology-shuffled null, with $p<0.005$ against both. The analysis also retains weak, negative and confounded findings, because these define the limits of what the current evidence can support.

These findings are deliberately presented separately rather than collapsed into a single accuracy claim, because the method is strongest when it distinguishes between useful evidence, weak evidence and confounded evidence.

- Propagation coherence is the strongest result: 0.99 observed same-root overlap compared with null values of 0.44 and 0.09, with $p<0.005$.
- The discriminant funnel is operationally useful: the consolidated filter reduces 8,367 raw alarms to 134 candidates.
- Persistence alone is weak as a predictor. The reported AUC is 0.52 with $p=0.005$. Because an AUC of 0.5 is equivalent to chance and 1.0 is perfect discrimination, 0.52 represents only a very small effect.
- Multi-signal corroboration is described as operational to strong where independent feeds agree, although this result must be interpreted alongside the confounding test described below.
- Day-to-day stability is also operationally useful, but a correlation of 0.99 across days can be produced by recurrent congestion repeating consistently. Stability by itself is therefore not evidence of incident accuracy.

The persistence result illustrates the difference between statistical significance and operational value. With approximately 232,000 evaluations, a very small effect can still produce $p=0.005$. The numerical significance does not change the fact that AUC 0.52 is only marginally above chance.

The confounding analysis provides an equally important negative result. A corroboration signal initially appeared enriched by 23.6 times, but after controlling for the confound the value fell to 0.42. The apparent effect therefore did not survive the control. Retaining this result is important: it demonstrates that the validation framework can reject an apparently strong finding when the evidence does not support it.

Operational implications

The purpose of incident detection is not to create another data feed. It is to provide actionable information early enough for an operator to investigate and respond. Incident detection therefore sits alongside flow forecasting, which estimates what the next 5, 10 and 15 minutes should look like. Taken together, the two capabilities can help identify that current conditions have departed from expectation, show where the effect is propagating and indicate how the situation may develop.

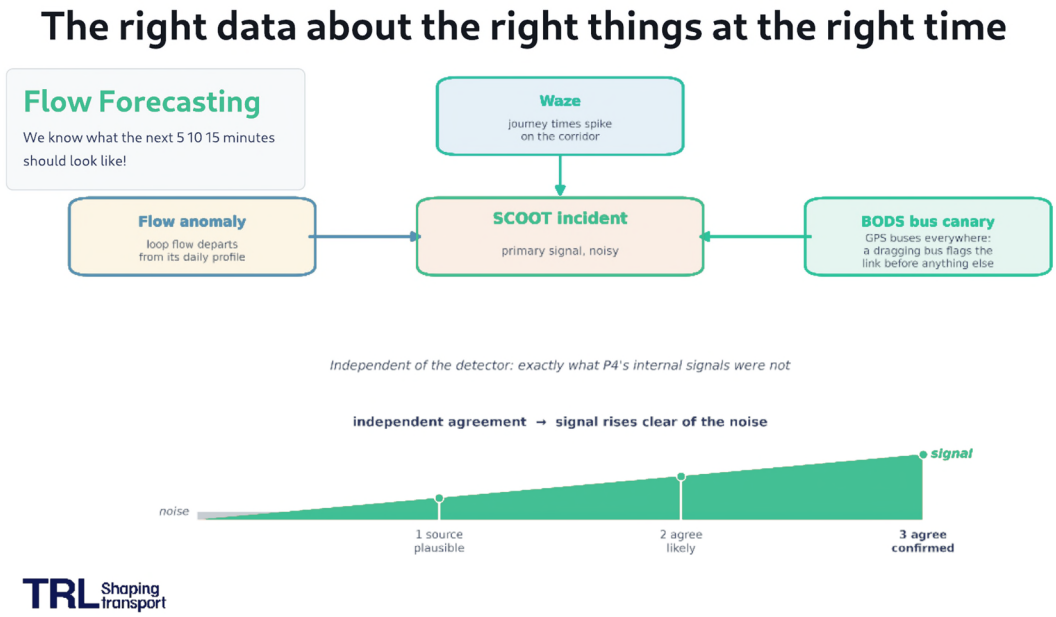


Figure 10. Relationship between incident information and short-horizon flow forecasting.

Without filtering, the detector produces too many alerts to support practical control-room use. With the proposed filter, the output becomes a smaller set of candidate events that can be checked against CCTV, site intelligence, bus-operator reports or local operational knowledge.

For TfGM, an incident can be verified using the evidence available to the control room, including cameras, site information and bus-operator intelligence. Once confirmed, the control room can disseminate information about the disruption to bus and tram operators, other road-network partners and the travelling public. If an obstruction is confirmed, it can be removed or passed to the relevant party and traffic-signal timings can be amended, manually or automatically, where appropriate to mitigate the effect. This is particularly valuable for short-lived or unplanned events that may occur outside direct CCTV coverage and would otherwise be recognised only after congestion has become established.

Discussion and limitations

The work makes three practical contributions. First, it provides a label-free validation approach based on explicit negative controls rather than treating the raw incident flag as truth. Second, it identifies an operational filter based on persistence, non-recurrence and

coherent propagation. Third, it defines a clear boundary around the current evidence: true precision and recall still require ground truth. The current results show that the filtered output has strong physical structure and can reduce an impractically noisy detector to a manageable candidate set, but they do not prove the accuracy of every alert.

Conclusions

The 28 January 2026 matchday case provides a demonstration of how SCOOT® 8 AI incident detection can identify a real operational disruption. Because the cause of disruption was known and the resulting traffic response could be observed, it provides a clear reference point for how an incident-like event appears in the data. The wider operational value lies in applying this capability to the harder and more common situation where the cause is not already known. The wider dataset also shows why the raw incident flag cannot simply be passed to an operator, recurrent congestion causes persistent over-firing and makes conventional accuracy impossible to establish without labels.

The strongest evidence is that the predicted impacts behave like real traffic disruption. They mainly propagate upstream, and repeated detections from the same root location affect the same corridor rather than appearing randomly across the network. This gives confidence that SCOOT® 8 AI incident detection is identifying physically credible disruption patterns. Other results are more limited: persistence alone is weak, stability can be caused by recurrent congestion, and one apparent corroboration result disappears after controlling for a confound. These limitations do not undermine the matchday finding; they define how the wider operational evidence should be interpreted.

Knowing there is a problem is the first step to starting to solve it. The contribution of this work is to make that first step more structured: not by claiming accuracy that cannot yet be measured, but by using persistence, non-recurrence, network topology and traffic physics to identify the cases that are most credible and most useful for operational attention. The aim is to expose those candidate incidents to operators in a way that supports effective intervention as quickly as possible, before the impact has time to spread further through the network.