New approaches for accounting for confounding factors when analysing collision data to predict collision hotspots and evaluate road safety schemes

Dr. Neil Thorpe and Dr. Lee Fawcett; Paulo Humanes, PTV Group

School of Engineering and School of Mathematics, Statistics and Physics, Newcastle University (UNEW); UK

Presentation to the 6th IRTAD Conference 'Better Road Safety Data for Better Safety Outcomes'

10th-12th October, 2017 Marrakech, Morocco

School of Civil Engineering & Geosciences

School of Mathematics & Statistics

Newcastle University

Contents

1. **Confounding factors:**
   - What are they?
   - Why should we account for them?
   - How can we account for them?
2. **Overview of the methodology developed**
   - Data from multiple time-periods
   - Global and site-specific trends
   - Variance-inflation factor
   - Bayesian posterior predictive distribution; model validation
3. **Application in available software programs**
   - RAPTOR (UNEW)
   - VISUM Safety (PTV)
4. **Benefits of the Approach**
   - Scheme Evaluation
   - Hotspot prediction

Perceived safety problem at a location(s) on the road network (hotspot identification) → Implement intervention strategy → Monitor impact (scheme evaluation)

COLLISION DATA

…but prone to confounding factors (e.g. regression-to-mean; trend)

1. Confounding factors

"Any factor that may lead to confounding…e.g. to effects that may erroneously be mixed up with the effects of a road safety measure" (Elvik; 2004 p. 1032)

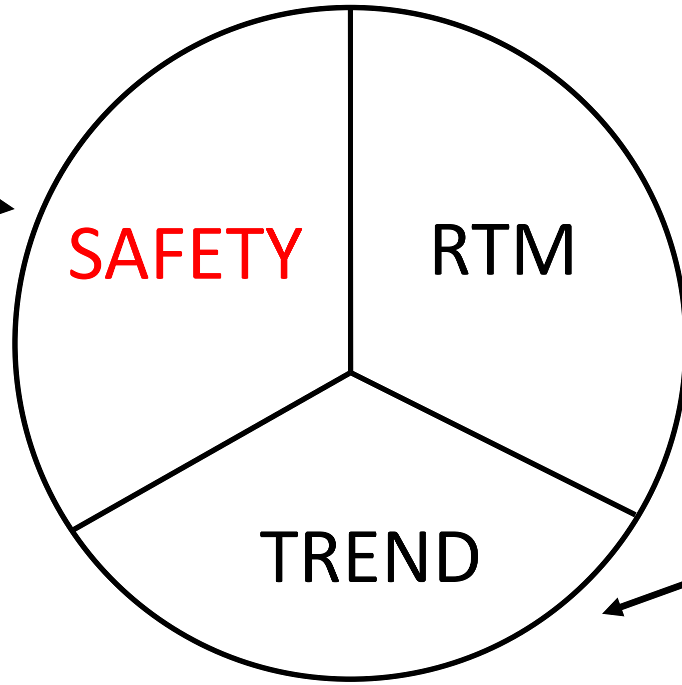
- **Regression-to-the mean** (the tendency for unusually high or low counts to be followed by values closer to the underlying mean)

- **General trends** in collisions/casualties (for example due to changes in vehicle safety and driver education)

Remove confounding effects from our analyses

Total observed frequency (100%)

The changing level of site safety

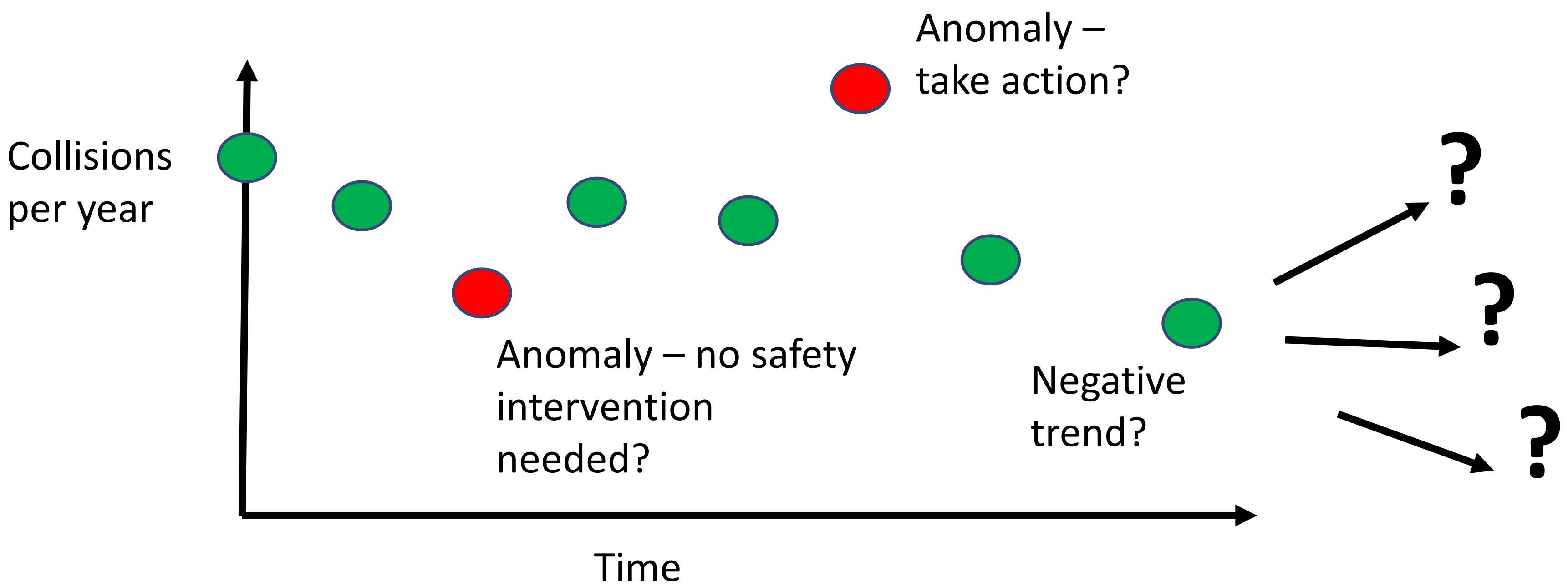


Confounding factors

- Scheme evaluation
- Hotspot prediction

Variation over time and between sites

Problems for scheme evaluation and hotspot prediction



Why are confounding factors a problem?

Cause ‘noise’ in the collision count (time series) data

For hotspot identification:

- False positives: identifying and treating sites as hotspots when they are not – collision rate would have reduced anyway; an issue of ‘wasted’ resources
- False negatives: not treating a genuinely unsafe site; impact for future collision rates

For scheme evaluation:

- Believing that our schemes are being more effective than they actually are – value for money issues and ‘misguided’ future decisions

Accounting for RTM and Trend

RTM

- Ignore it – assume it doesn't exist
- Bayesian techniques (Empirical or Full)
  - Not widely accessible to practitioners

Trend

- Ignore it
- Network-wide and site-specific trends
- Relative influence of more recent observations and observations further back in time

2. Overview of the methodology

Key functions:

- Hotspot prediction (Fawcett *et al*., 2017)
- Scheme evaluation (Fawcett and Thorpe, 2012, 2013)

RTM

- Combines what we observe at a site with a state-of-the art model-based estimate of safety
- Natural extension to classic methods (e.g. Empirical Bayes) to account for observations across multiple time periods (hotspot)
- Variations in historical data to inform predictions of future counts (hotspot)
- Crash modification factors to account for discrepancies between APM and observed accident counts caused by missing data (hotspot)

Trend

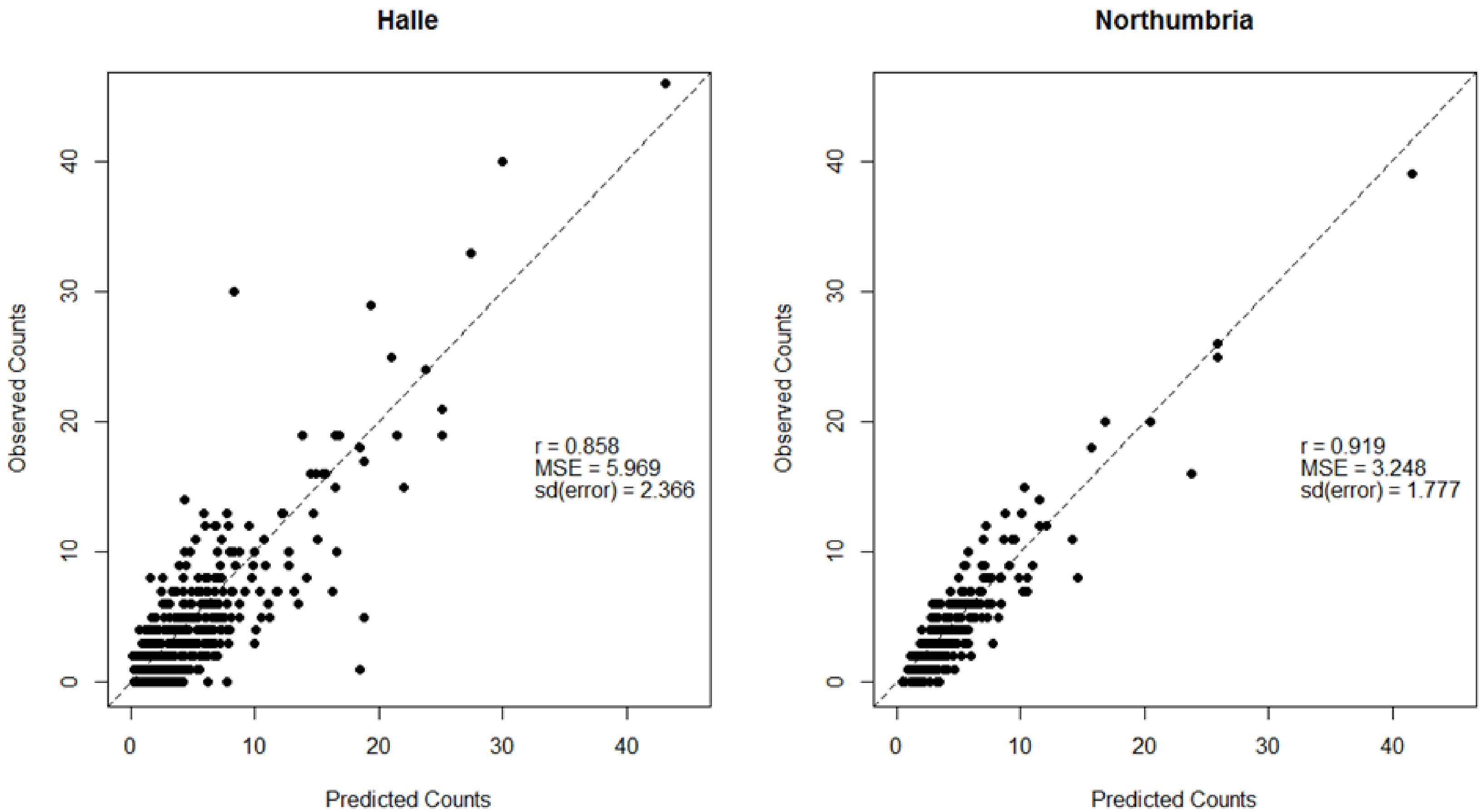
- Simple multiplicative factor applied to accident prediction model based on historic records or include time as a covariate in the model (Scheme evaluation)
- Variance inflation (predictions rely more heavily on more recent observations) (hotspot)
- Allows for statistically significant site-specific deviations to offset globally-observed trend when predicting future collision counts (hotspot)

Data requirements

- Hotspot prediction **and** scheme evaluation
  - **Dependent variable:** Collision/casualty counts in discrete time periods (e.g. months, quarters or years) for each site
  - **Independent variables:** Static site data (e.g. speed limit; road type; road class, urban/rural); dynamic site data (e.g. flow; average speeds) for each time period

- Scheme evaluation **only**
  - The same but for a reference pool of sites to construct the accident prediction model

Validation: how good are the hotspot predictions?



3. Application in available software

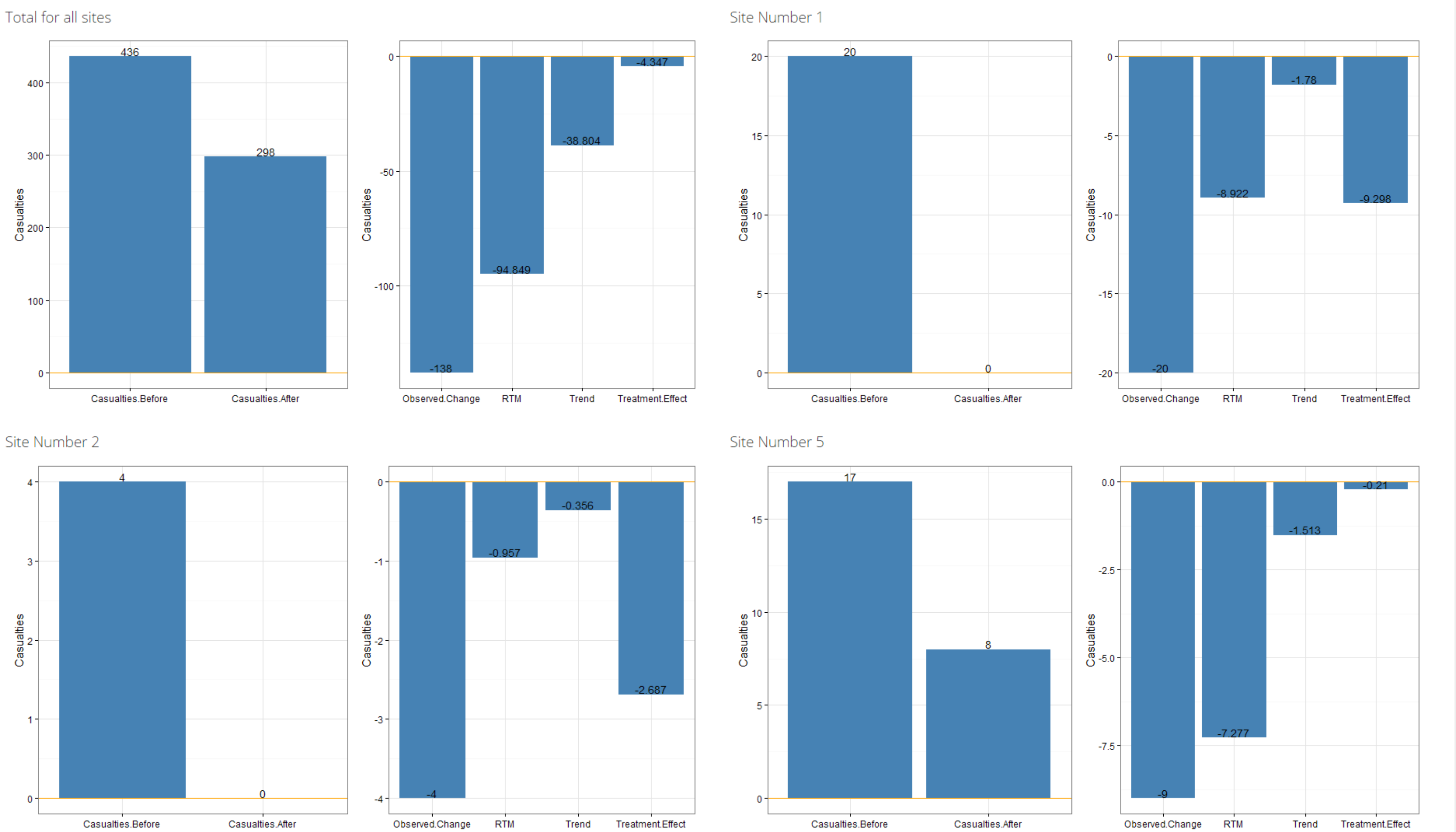
RAPTOR

- Hosted on UNEW servers; web-based
- Logins/passwords freely available
- Supports hotspot prediction and scheme evaluation

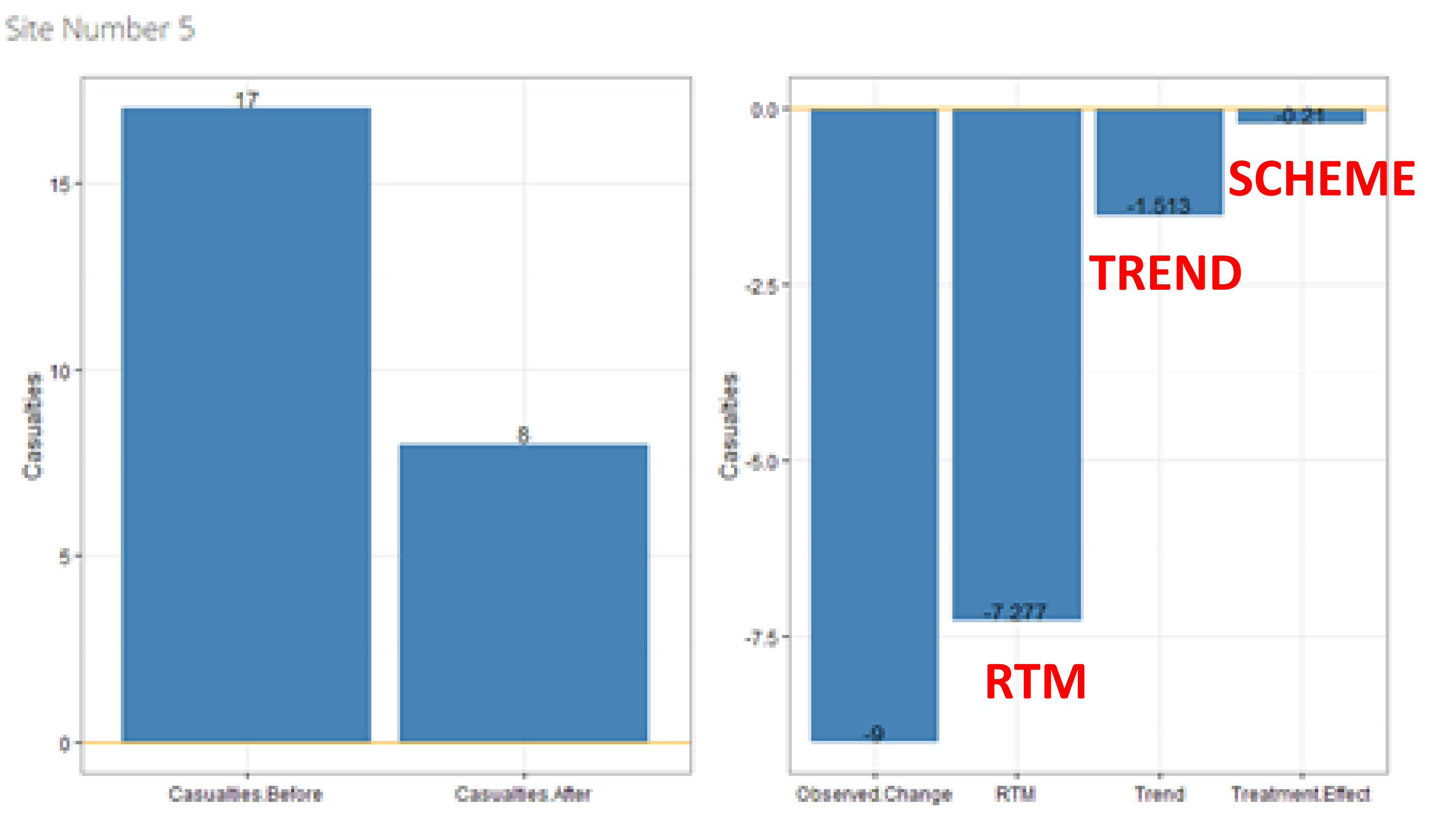
VISUM Safety

- Available from PTV Group under licence
- Supports hotspot prediction only
- Allows mapping of future collision sites
- Linked to strategic transport model VISUM for scenario testing

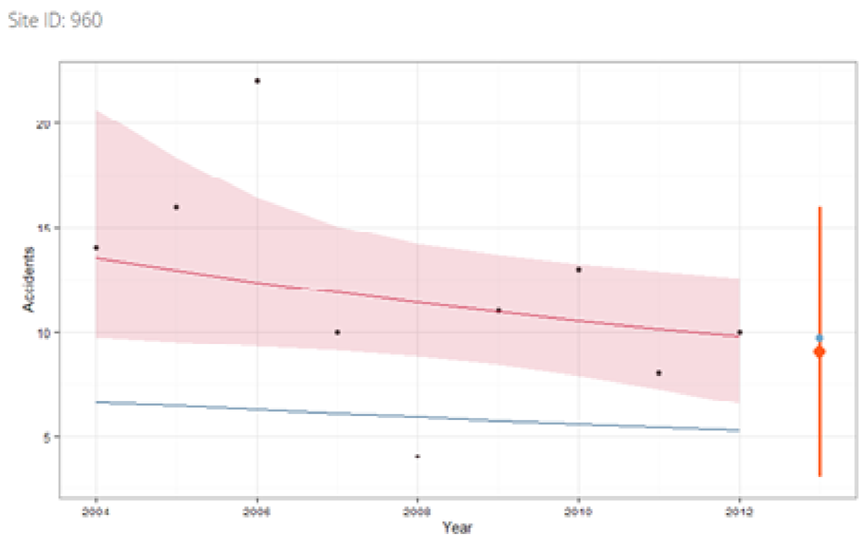
RAPTOR: Scheme Evaluation

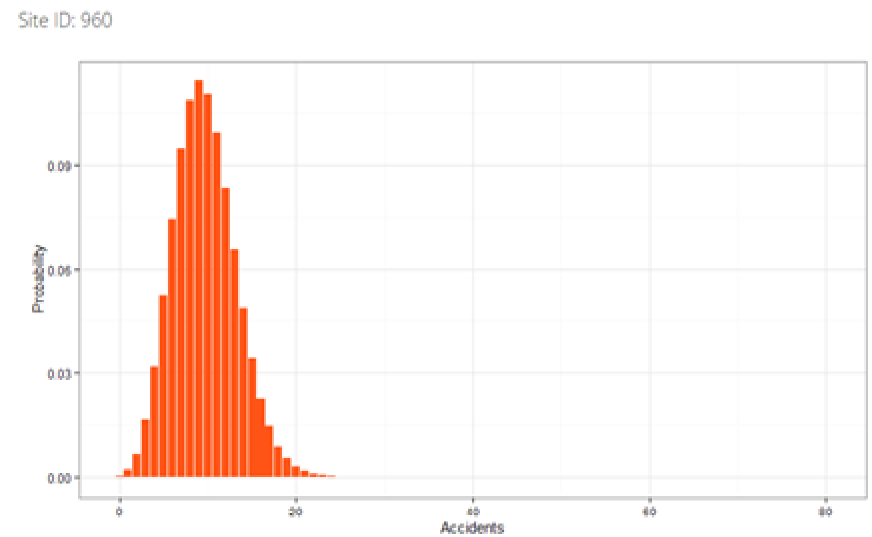


Site-by-site breakdown: Site 5



RAPTOR: Hotspot Prediction





Results

Summary Tables

Predicted number of accidents | APM Output | Site Warnings

Site Warnings

Set Warning thresholds

Warnings can be triggered for sites that are predicted to exceed a specified number of accidents with the selected probability or higher.

Number of accidents

1 — 5 — 30

Minimum % probability of having 5 or more accidents next year needed to trigger a warning

10 — 50 — 100

Sites with warnings

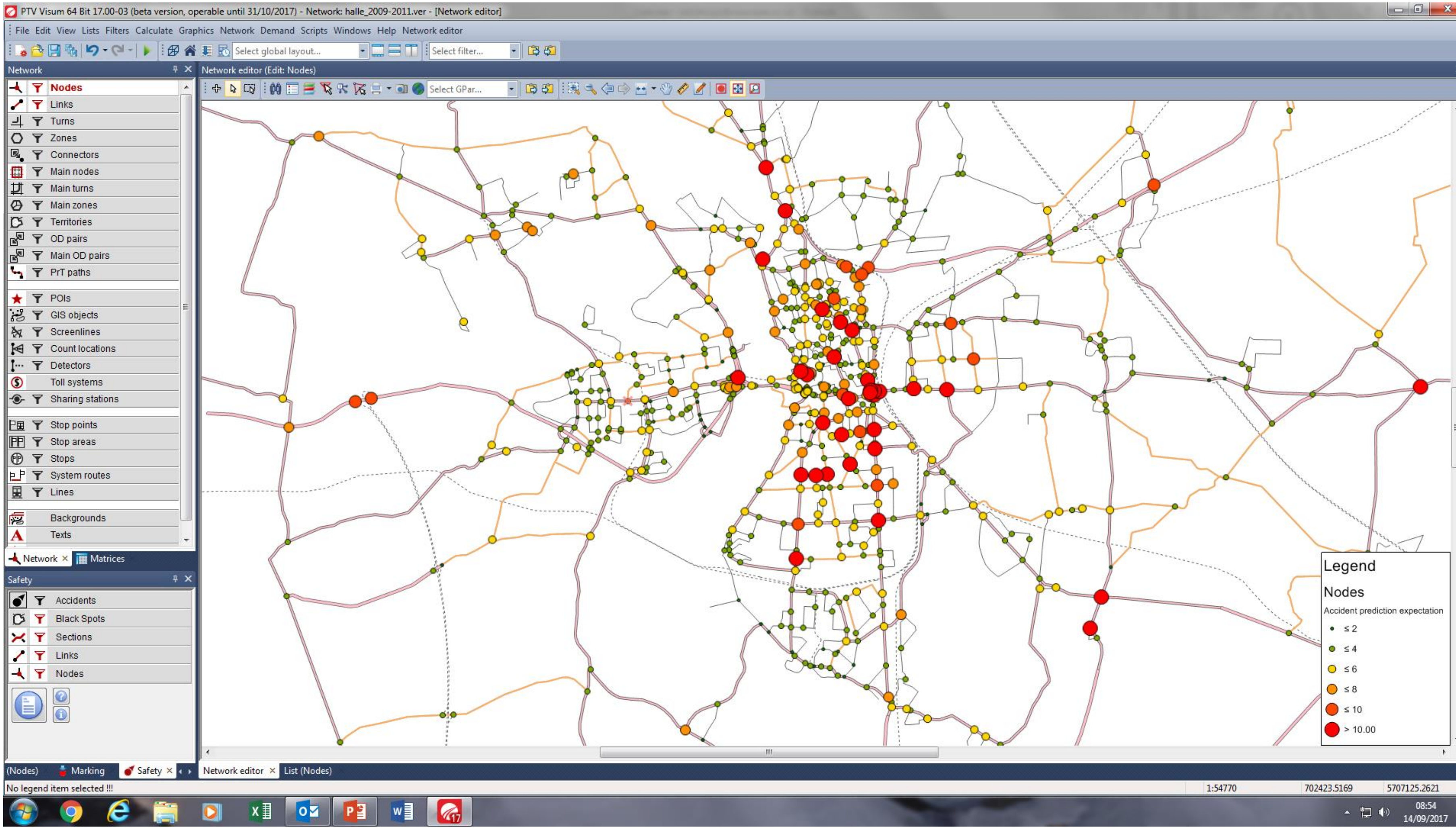
Showing sites that are predicted to have 5 accidents next year with 50% probability or higher.

There are 4 sites with warnings.

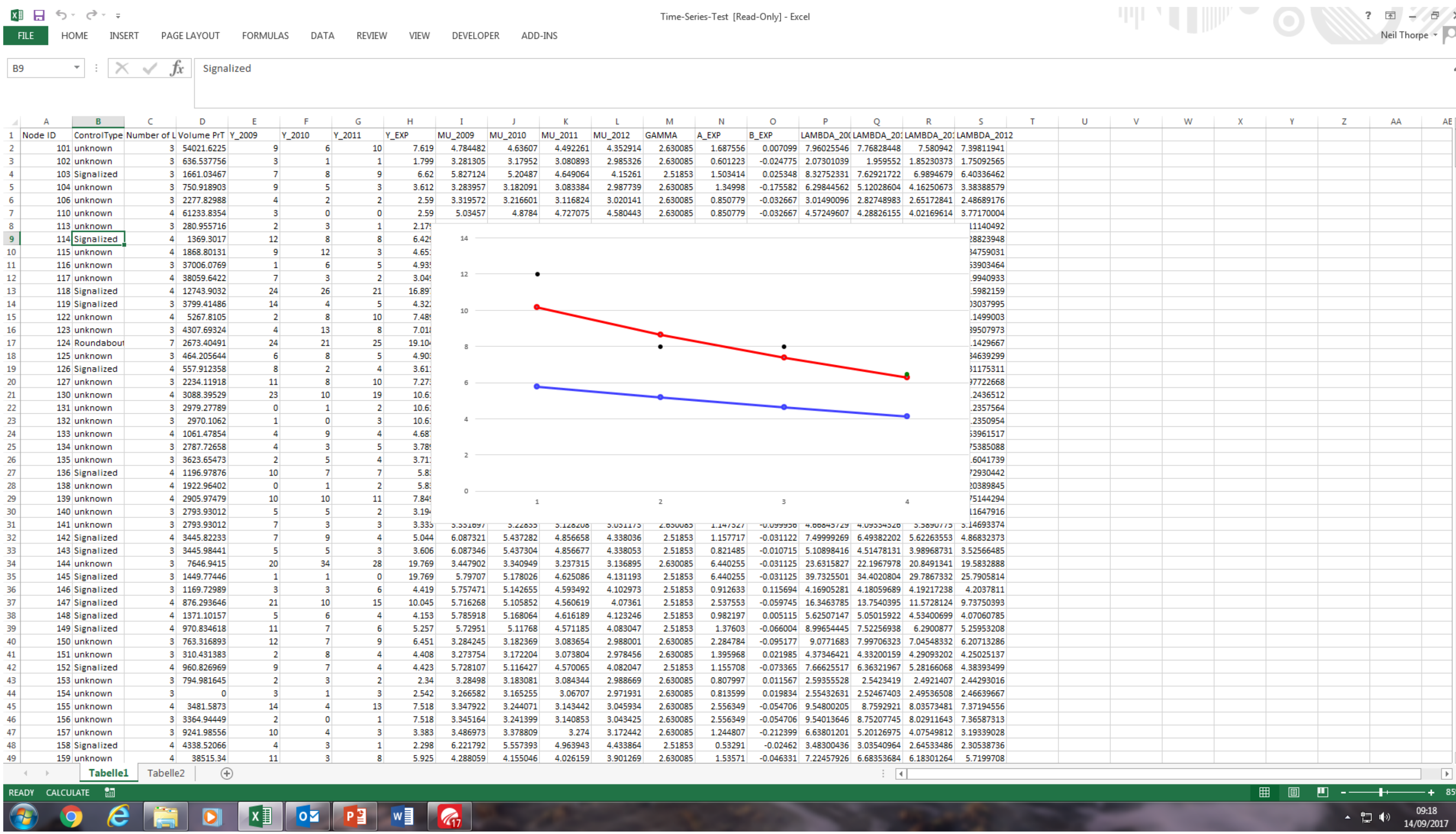
Table | List

Site	Accidents expected with 50% probability	Probability of exceeding 5 or more accidents
960	9	94.2%
1145	8	87.9%
965	7	78.5%
103	6	70.8%

VISUM Safety: Current clusters

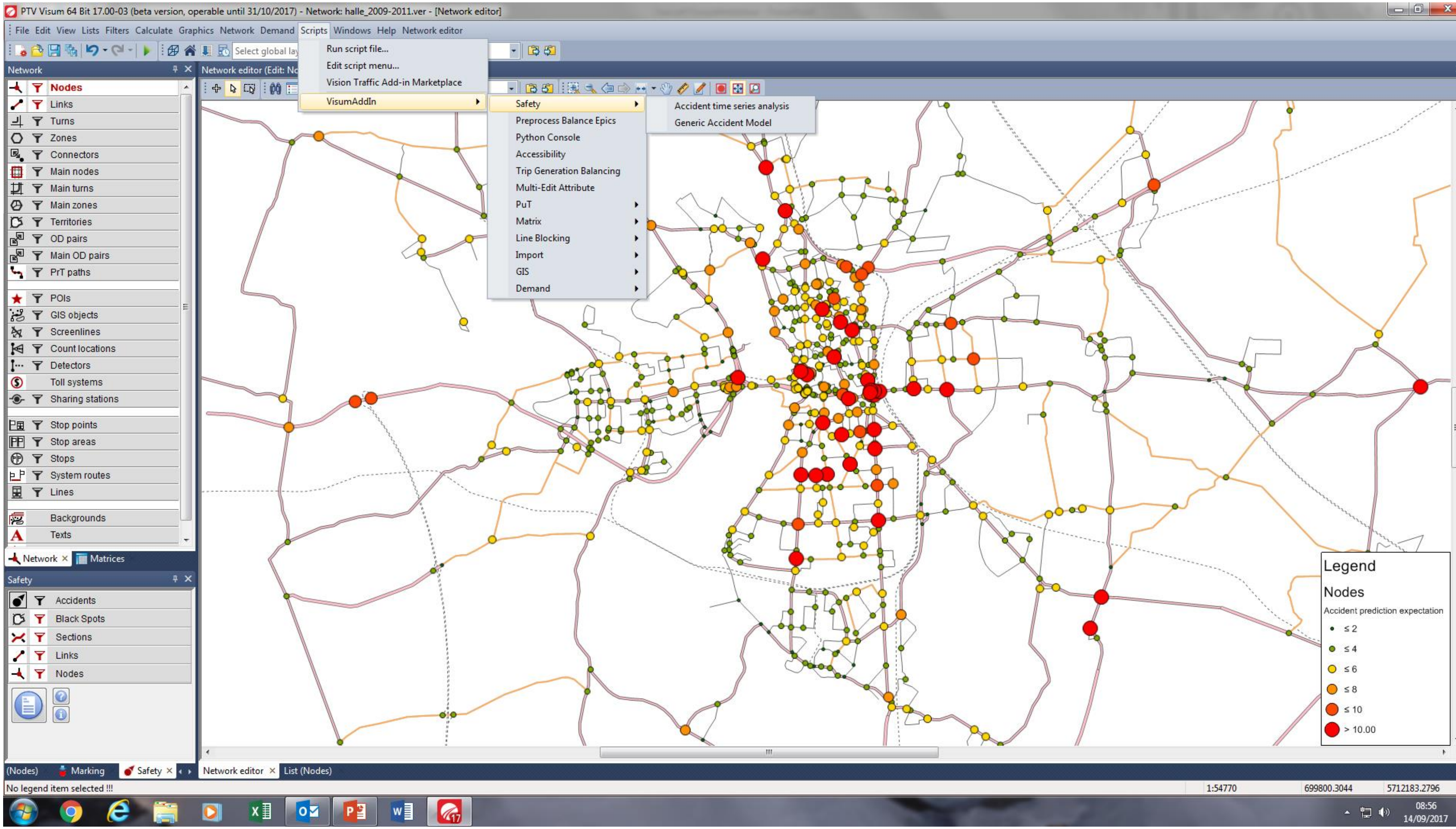


VISUM Safety: Output and Analysis



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Node ID	ControlType	Number of L	Volume PrT	Y_2009	Y_2010	Y_2011	Y_EXP	MU_2009	MU_2010	MU_2011	MU_2012	GAMMA	A_EXP	B_EXP	LAMBDA_20(	LAMBDA_201	LAMBDA_201	LAMBDA_2012
2	101	unknown	3	54021.6225	9	6	10	7.619	4.784482	4.63607	4.492261	4.352914	2.630085	1.687556	0.007099	7.96025546	7.76828448	7.580942	7.39811941
3	102	unknown	3	636.537756	3	1	1	1.799	3.281305	3.17952	3.080893	2.985326	2.630085	0.601223	-0.024775	2.07301039	1.959552	1.85230373	1.75092565
4	103	Signalized	3	1661.03467	7	8	9	6.62	5.827124	5.20487	4.649064	4.15261	2.51853	1.503414	0.025348	8.32752331	7.62921722	6.9894679	6.40336462
5	104	unknown	3	750.918903	9	5	3	3.612	3.283957	3.182091	3.083384	2.987739	2.630085	1.34998	-0.175582	6.29844562	5.12028604	4.16250673	3.38388579
6	106	unknown	3	2277.82988	4	2	2	2.59	3.319572	3.216601	3.116824	3.020141	2.630085	0.850779	-0.032667	3.01490096	2.82748983	2.65172841	2.48689176
7	110	unknown	4	61233.8354	3	0	0	2.59	5.03457	4.8784	4.727075	4.580443	2.630085	0.850779	-0.032667	4.57249607	4.28826155	4.02169614	3.77170004
8	113	unknown	3	280.955716	2	3	1	2.17											11140492
9	114	Signalized	4	1369.3017	12	8	8	6.42											28823948
10	115	unknown	4	1868.80131	9	12	3	4.65											34759031
11	116	unknown	3	37006.0769	1	6	5	4.93											53903464
12	117	unknown	4	38059.6422	7	3	2	3.04											.9940933
13	118	Signalized	4	12743.9032	24	26	21	16.89											.5982159
14	119	Signalized	3	3799.41486	14	4	5	4.32											03037995
15	122	unknown	4	5267.8105	2	8	10	7.48											.1499003
16	123	unknown	3	4307.69324	4	13	8	7.01											39507973
17	124	Roundabout	7	2673.40491	24	21	25	19.10											.1429667
18	125	unknown	3	464.205644	6	8	5	4.90											34639299
19	126	Signalized	4	557.912358	8	2	4	3.61											31175311
20	127	unknown	3	2234.11918	11	8	10	7.27											97722668
21	130	unknown	4	3088.39529	23	10	19	10.6											.2436512
22	131	unknown	3	2979.27789	0	1	2	10.6											.2357564
23	132	unknown	3	2970.1062	1	0	3	10.6											.2350954
24	133	unknown	4	1061.47854	4	9	4	4.68											53961517
25	134	unknown	3	2787.72658	4	3	5	3.78											75385088
26	135	unknown	3	3623.65473	2	5	4	3.71											.6041739
27	136	Signalized	4	1196.97876	10	7	7	5.8											72930442
28	138	unknown	4	1922.96402	0	1	2	5.8											20389845
29	139	unknown	4	2905.97479	10	10	11	7.84											75144294
30	140	unknown	3	2793.93012	5	5	2	3.19											11647916
31	141	unknown	3	2793.93012	7	3	3	3.335	3.331697	3.22833	3.128208	3.031173	2.630085	1.147327	-0.099936	4.66845729	4.09334326	3.5890775	3.14693374
32	142	Signalized	4	3445.82233	7	9	4	5.044	6.087321	5.437282	4.856658	4.338036	2.51853	1.157717	-0.031122	7.49999269	6.49382202	5.62263553	4.86832373
33	143	Signalized	3	3445.98441	5	5	3	3.606	6.087346	5.437304	4.856677	4.338053	2.51853	0.821485	-0.010715	5.10898416	4.51478131	3.98968731	3.52566485
34	144	unknown	3	7646.9415	20	34	28	19.769	3.447902	3.340949	3.237315	3.136895	2.630085	6.440255	-0.031125	23.6315827	22.1967978	20.8491341	19.5832888
35	145	Signalized	3	1449.77446	1	1	0	19.769	5.79707	5.178026	4.625086	4.131193	2.51853	6.440255	-0.031125	39.7325501	34.4020804	29.7867332	25.7905814
36	146	Signalized	3	1169.72989	3	3	6	4.419	5.757471	5.142655	4.593492	4.102973	2.51853	0.912633	0.115694	4.16905281	4.18059689	4.19217238	4.2037811
37	147	Signalized	4	876.293646	21	10	15	10.045	5.716268	5.105852	4.560619	4.07361	2.51853	2.537553	-0.059745	16.3463785	13.7540395	11.5728124	9.73750393
38	148	Signalized	4	1371.10157	5	6	4	4.153	5.785918	5.168064	4.616189	4.123246	2.51853	0.982197	0.005115	5.62507147	5.05015922	4.53400699	4.07060785
39	149	Signalized	4	970.834618	11	7	6	5.257	5.72951	5.11768	4.571185	4.083047	2.51853	1.37603	-0.066004	8.99654445	7.52256938	6.2900877	5.25953208
40	150	unknown	3	763.316893	12	7	9	6.451	3.284245	3.182369	3.083654	2.988001	2.630085	2.284784	-0.095177	9.0771683	7.99706323	7.04548332	6.20713286
41	151	unknown	3	310.431383	2	8	4	4.408	3.273754	3.172204	3.073804	2.978456	2.630085	1.395968	0.021985	4.37346421	4.33200159	4.29093202	4.25025137
42	152	Signalized	4	960.826969	9	7	4	4.423	5.728107	5.116427	4.570065	4.082047	2.51853	1.155708	-0.073365	7.66625517	6.36321967	5.28166068	4.38393499
43	153	unknown	3	794.981645	2	3	2	2.34	3.28498	3.183081	3.084344	2.988669	2.630085	0.807997	0.011567	2.59355528	2.5423419	2.4921407	2.44293016
44	154	unknown	3	0	3	1	3	2.542	3.266582	3.165255	3.06707	2.971931	2.630085	0.813599	0.019834	2.55432631	2.52467403	2.49536508	2.46639667
45	155	unknown	4	3481.5873	14	4	13	7.518	3.347922	3.244071	3.143442	3.045934	2.630085	2.556349	-0.054706	9.54800205	8.7592921	8.03573481	7.37194556
46	156	unknown	3	3364.94449	2	0	1	7.518	3.345164	3.241399	3.140853	3.043425	2.630085	2.556349	-0.054706	9.54013646	8.75207745	8.02911643	7.36587313
47	157	unknown	3	9241.98556	10	4	3	3.383	3.486973	3.378809	3.274	3.172442	2.630085	1.244807	-0.212399	6.63801201	5.20126975	4.07549812	3.19339028
48	158	Signalized	4	4338.52066	4	3	1	2.298	6.221792	5.557393	4.963943	4.433864	2.51853	0.53291	-0.02462	3.48300436	3.03540964	2.64533486	2.30538736
49	159	unknown	4	38515.34	11	3	8	5.925	4.288059	4.155046	4.026159	3.901269	2.630085	1.53571	-0.046331	7.22457926	6.68353684	6.18301264	5.7199708

VISUM Safety: Predicted clusters



4. Benefits of the approach

Peer-reviewed approach accessible to road safety practitioners to aid decision-making

Information about ‘true’ effect of road safety interventions on collision/casualty reduction

Predictions of collision/casualty frequency in a future time period: site prioritisation

Evidence-led and proactive approach to road safety investment

References

Thorpe N, Fawcett L. (2012) 'Linking road casualty and clinical data to assess the effectiveness of mobile safety enforcement cameras: a before and after study.' BMJ Open, 2(6), e001304. http://bmjopen.bmj.com/content/2/6/e001304?ct

Fawcett, L.; Thorpe, N. (2013) Mobile safety cameras: estimating casualty reductions and the demand for secondary healthcare. Journal of Applied Statistics 40(11), 2385-2406 http://www.tandfonline.com/doi/full/10.1080/02664763.2013.817547

Fawcett, L.; Thorpe, N.; Matthews, J.; Kremer, K. (2017) A novel Bayesian hierarchical model for road safety hotspot prediction. Accident Analysis & Prevention, 99, pp.262-271. http://www.sciencedirect.com/science/article/pii/S0001457516304341

RAPTOR logins and further information available from Neil.Thorpe@ncl.ac.uk