

# Community Tension

Data Analytics Lab

July 2020

WMP Officers have noted that when incidents of violence occur, there have occasionally previously been reports of tensions within communities. This project sought to identify whether it is possible to use such reports to understand potential conflict and minimise the risk of disorder, violence or crime, i.e. to see whether it was possible to build a model that used information regarding community tensions to predict potential future occurrences of violence.

**It has not proved possible to satisfactorily link reports of community tension to incidents of violence (or other forms of crime) and therefore no predictive model can be built.** This project has therefore been discontinued.

# 1 Table of Contents

2	Introduction.....	3
3	Exploratory Data Analysis.....	4
3.1	Data Extract .....	4
3.2	Topic Modelling.....	5
3.3	Spatial.....	7
4	Modelling.....	8
4.1	Direct Matching.....	8
4.2	Spatio-Temporal .....	8
5	Findings.....	11
6	References.....	12

## 2 Introduction

*Community tension is a state of community dynamics which may potentially lead to disorder, threaten peace and stability, or raise the levels of fear and anxiety in the whole, or a part of the local community. (Cantle 2010)*

Strained relationships may build up within or between communities, or against particular institutions, based on real or perceived events or information or on fears, prejudices, circumstances or specific actions. Tensions may develop over a long period and be inflamed by a 'spark' which precipitates disorder and criminal activity.

WMP Officers have noted previously that when incidents of violence occur, there have occasionally previously been reports of tensions within communities. This project therefore sought to identify whether it was possible to use such reports of community tensions to estimate how likely it is that violence will be seen resulting from these tensions. Hence to understand potential conflict and minimise the risk of disorder, violence or crime.

Whilst this report provides information as to the analyses that have been employed to investigate any links, **it has not proved possible to satisfactorily link reports of community tension to incidents of violence (or other forms of crime) and therefore no predictive model can be built.**

This project has therefore been discontinued.

### 3 Exploratory Data Analysis

#### 3.1 Data Extract

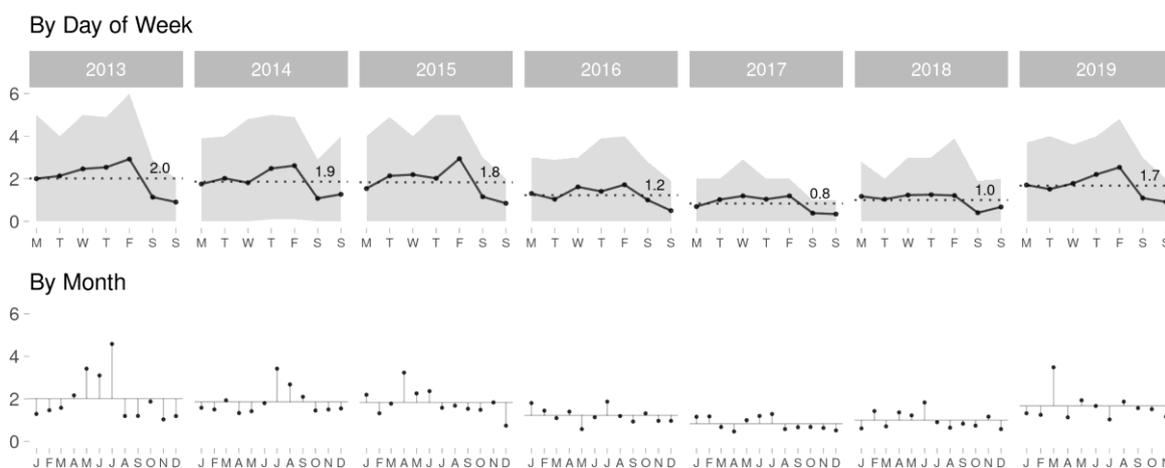
Data were extracted for the period January 2013 - October 2019 from the Crimes, and IMS databases for all IMS logs marked as community tension.

In this period there are 3,720 community tension reports. Reports dropped from 735 in 2013 reaching a low of 305 in 2017 before increasing again to 519 in 2019.

Variable	Type	Comments
log ref	numeric	unique identifier of the IMS log
date log created	date	Date IMS log created
lpu	character	Local Policing Unit
log group	numeric	a grouping between related logs
dets ref	numeric	unique identifier for the IMS log Details
trusted source	logical	Originator Source in ('A', 'B', 'C', '1')
trusted intelligence	logical	Originator Intelligence in ('1', '2', '3', 'A', 'B', 'C')
details	character	free text describing intelligence
grid reference northing	numeric	Location based on ordnance survey 12 digit grid system
grid reference easting	numeric	

From the chart below, there is a weekly pattern with the highest number of reports made towards the end of a week on a Thursday / Friday and a far lower number of reports made over the weekend. There is a less clear pattern month by month.

IMS Logs marked as Community Tension  
Showing Daily Averages with 10% to 90% quantiles

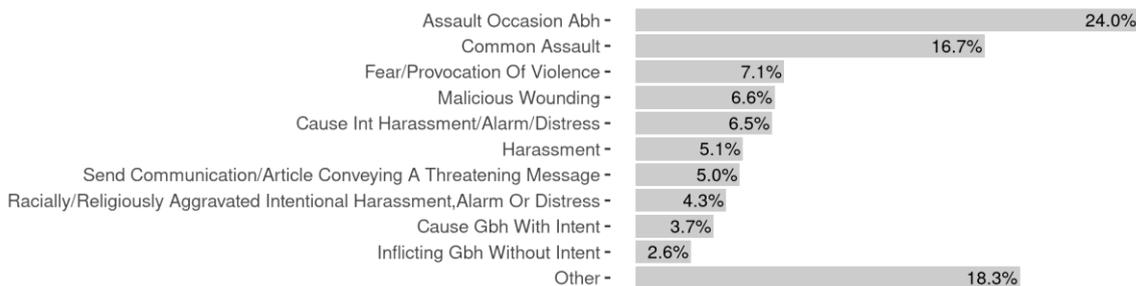


Source: WMP DAL 2020

In the same period there is a year on year increase in violent crime. From an average of 70 incidents per day in 2013 rising to 176 incidents per day in 2019.

Variable	Type	Comments
crime ref	numeric	Unique identifier of the Crime
date record created	date	Date Crime record created
offence title	character	Type of offence
grid reference northing	numeric	Location based on ordnance survey 12 digit grid system
grid reference easting	numeric	

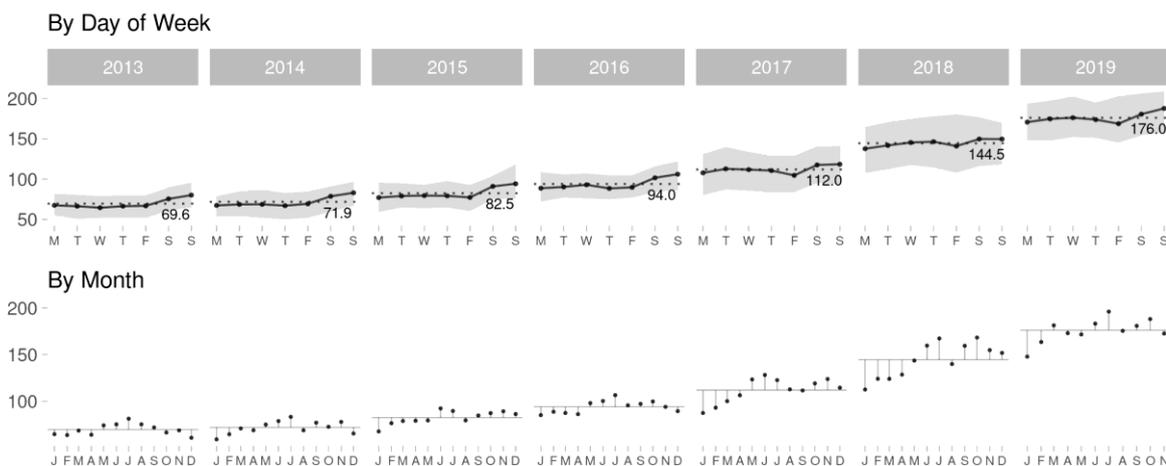
### Types of Violent Crime



Source: WMP DAL 2020

### Violent Crime

Showing Daily Averages with 10% to 90% quantiles



Source: WMP DAL 2020

## 3.2 Topic Modelling

Circumstances that give rise to tension are enumerated by Cante (2018):

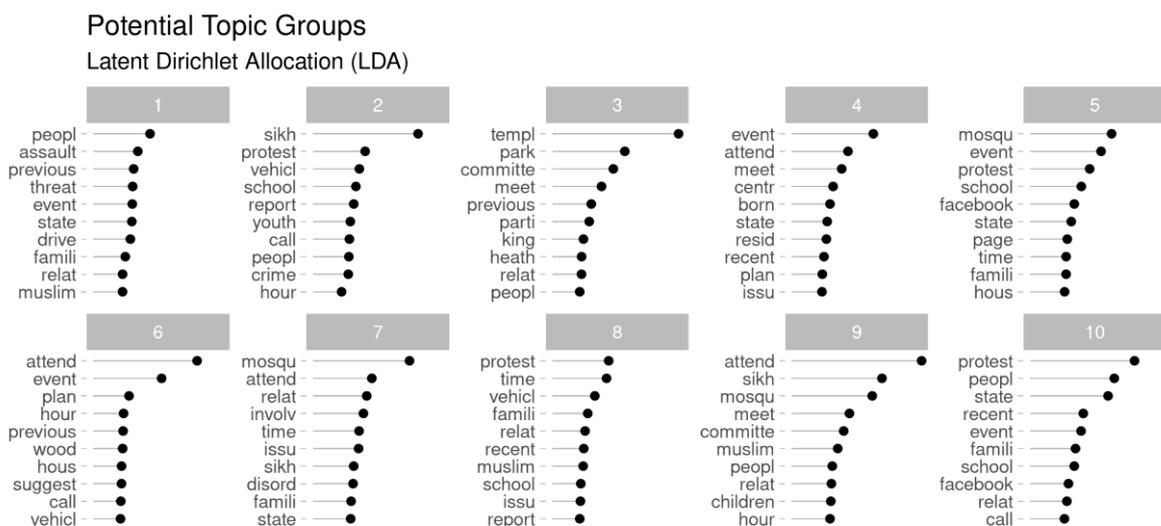
- *Community* – such as tensions between specific communities and incidents between specific communities
- *Criminal* – such as gang activity and antisocial behaviour
- *Future* – such as significant anniversaries and planned demonstrations

- *Immigration, asylum and refugees* - such as local concerns about the effects of migration on the local area; impact of national policy on migration (such as asylum dispersal) on the local area; racist attacks motivated by anti-migrant sentiment
- *National and international* - such as incidents which have impacted upon or have the potential to impact upon the local community
- *Political* - such as extremist political activity (leafleting, graffiti, meetings); elections or by-elections at which extremist candidates are standing; issues potentially detrimental to community cohesion, which are prominent on the public agenda; local demonstrations; local political situations which are exacerbating tensions
- *Racial and religious* - such as actions of racist organisations; racially and/or religiously motivated incidents and offences; concerns within communities - including faith communities about hate crime

The average Community Tension report comprises of 843 characters (90% interval: 285 - 1602) equating to 134 words (90% interval: 44 - 256). Reports do not tend to include considerable detail.

Analysing the IMS text by applying latent dirichlet allocation (LDA) to classify reports into semantically similar topics reveals ten distinct groupings. These only loosely adhere to Cantle’s taxonomy. The number of topics was chosen following a Cao-Juan test. In the diagram below, the words under each topic are those that appear the most in texts that relate to the topic at hand.

- *Criminal* - assaults / threats, 1
- *Racial and religious* - 2, 3, 5, 7, 9
- *Other Event* - 4, 6,
- *Protest* - 8, 10



Source: WMP DAL 2020

### 3.3 Spatial

Crime incidents are not distributed randomly throughout space. Their distribution is dense at some locations while sparse at others. Both the reports of tension and the crime records are highest in high population areas and city centres of the West Midlands and this pattern is moderately stable week on week.

To understand whether it is possible to use such reports to understand future risk of disorder, violence or crime we cast the question in the form of a block design. The question of interest is then whether there is any evidence of a change in the number of violent crimes in the nearby area in the periods following a report of community tension.

Below we plot reports of community tension (blue) against violent crimes (green).

#### Weekly Reports of Community Tension



Source: WMP DAL 2020

## 4 Modelling

### 4.1 Direct Matching

A direct approach is to note the addresses and individuals associated with an IMS log, and then to search for future crimes involving these locations or individuals. Unfortunately, location is not a sufficient basis alone to reliably link and match crimes. Attempts made to match on combinations of location, elapsed time, and topic features proved unsuccessful. Of the ~5% of cases where a match was made, a large proportion related to the original intelligence rather than to events occurring in response to community tension.

Further progress might be possible by a shift in police priorities, to manually match newly raised crimes to intelligence logs.

### 4.2 Spatio-Temporal

Casting the question in the form of a block design. The question of interest is whether there is any evidence of a change in the number of violent crimes in a block in the periods following a report of community tension.

Using a block design we partition the West Midlands into small quadrants, for example, 1Km × 1Km × 1Week, and aggregate data at that level. Varying the block dimensions allows us to check sensitivity and the degree of localisation of any effect.

It is reasonable to suppose that there is an underlying *crime rate* which varies with spatial location, year, and a seasonal component represented by the week of the year. It follows that a Poisson-like model is a plausible starting point to start investigating the count of violent crimes in any particular block.

Based on the subject matter we expect more variation than a simple Poisson, and a negative binomial model is more appropriate.

Here, we begin by fitting a generalised additive model (Wood 2017) as a base model describing the variation in the crime rate to allow for smooth non-linear additive effects.

$$\begin{aligned} y_i &\sim \text{negative\_binomial}(\mu_i, \theta) \\ \log(\mu_i) &= \alpha + f_1(x_{block}, y_{block}) \\ &\quad + f_2(\text{week}_i) + f_3(\text{year}_i) \\ &\quad + f_4(\text{crimes recorded in the previous period}_i) \\ \theta &\sim \text{gamma}(0.01, 0.01) \end{aligned}$$

Where

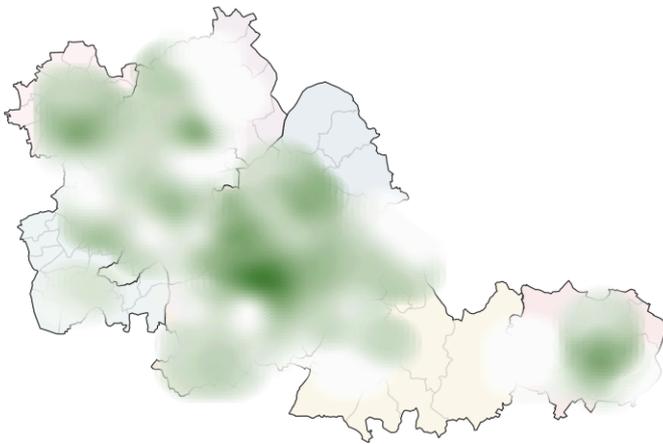
- $f_1$  is the isotropic smoothed spatial effect based on the northing and easting of the block.
- $f_2, f_3$  are the nonlinear temporal effects describing the trend of violent crime by time of year.
- $f_4$  is the ar(1) effect based on violent crime recorded in the previous period.

For a block size of 1Km × 1Km × 1Week, partial effects plots and diagnostics show that the model fit is sensible. The deviance explained is around 80%.

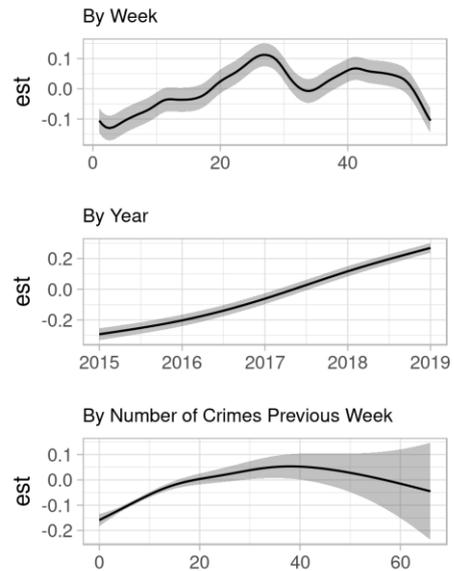
### Relationships to Weekly Crimes

Based on mgcv model. Showing 95% credible intervals

By Geographic Area



Density is proportional to the Local Crime rate.



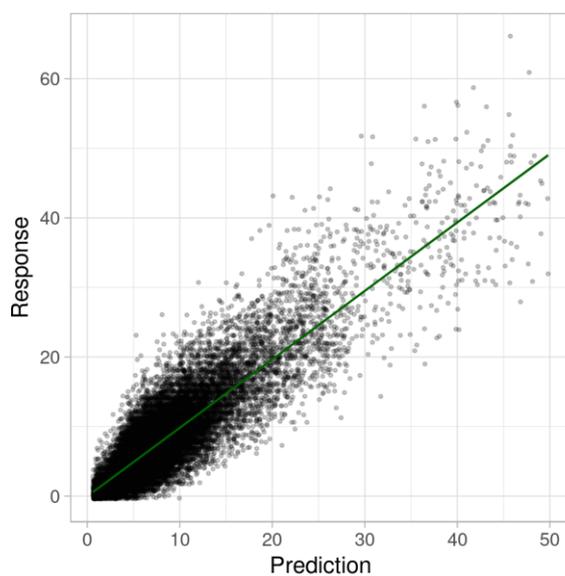
Source: WMP DAL 2020

Approximate significance of smooth terms:

	edf	Ref.df	F	p-value	
s(cx, cy)	145.917	147.938	94.690	<2e-16	***
s(week)	14.002	50.000	7.025	<2e-16	***
s(year)	3.468	3.840	718.472	<2e-16	***
:					
:					

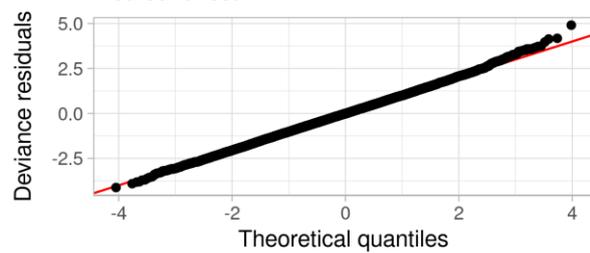
### Model Diagnostics

Response vs Fitted



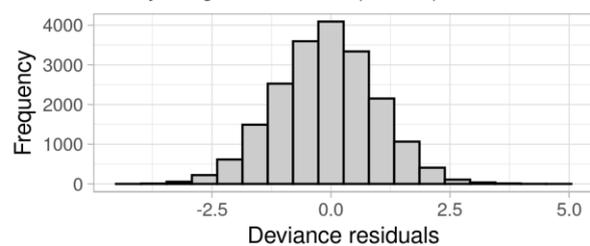
QQ plot of residuals

Method: direct



Histogram of residuals

Family: Negative Binomial(27.542)



Source: WMP DAL 2020

Using this base model we now add the number of previous reports of community tension in each block to the model. Examination of the coefficient table shows that a previous report of community tension does not have a quantifiable effect on an increase in violent (or other) crime.

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.046408	0.016555	123.611	<2e-16 ***
ims_count_lag_1:as.factor(year)2015	0.009868	0.011308	0.873	0.3829
ims_count_lag_1:as.factor(year)2016	-0.007960	0.013555	-0.587	0.5570
ims_count_lag_1:as.factor(year)2017	-0.016070	0.015406	-1.043	0.2969
ims_count_lag_1:as.factor(year)2018	-0.010733	0.013954	-0.769	0.4418
ims_count_lag_1:as.factor(year)2019	0.016842	0.009468	1.779	0.0753 .

Comparing the two models using hypothesis testing (anova) also confirms that including the lagged community tension term does not result in a better model.

Varying the block sizes (0.5Km – 5Km) and lagged periods (1Week – 5Week) to check sensitivity to various geographic and temporal distances fails to find a quantifiable effect of a report of community tension on the increase in violent (or other types of) crime. That is not to say that there is no effect, only that given the large week by week variation in crimes, an effect cannot be detected statistically.

## 5 Findings

There is no ascertainable statistical relationship between intelligence reports of tensions and eventual episodes of violence.

Though reports of tension are arguably causally related to future crime, the available observational data are subject to limitations, which prevent reliable automated direct matching between crimes and intelligence logs. Matching at a block level, using various geographic areas and temporal distances also fails to find a quantifiable effect of a report of community tension leading to an increase in violent (or other types of) crime.

Further progress might be possible by a shift in Police priorities, to manually match newly raised crimes to intelligence logs.

At present however, given that reports of community tensions cannot be linked to later episodes of violence or other crime a predictive model using tension reports as an input feature cannot be developed. This project has therefore been discontinued.

## 6 References

Cantle, Ted. 2010. "Community Tension Monitoring Guidance."

———. 2018. *Community Cohesion: A New Framework for Race and Diversity*. Springer.

Wood, Simon N. 2017. *Generalized Additive Models: An Introduction with R*. Chapman; Hall/CRC.