Knife Crime Prediction

Data Analytics Lab

(December 2020)

This version changed May 2021

This project aims to make predictions of knife crime in terms of the likely number and location of knife crimes over a 4 weekly period.

1 Contents

2 Introduction

The overall aim of this project is to develop a statistical model to predict the likely levels of knife crime (used causing injury) over time and space within the WMP area.

Nationally and locally there is concern about the increase in knife crime, particularly involving young people. The Parliamentary Youth Select Committee's report 'Our Generation's Epidemic: Knife Crime' (2019) stated that the number of fatal stabbings in the year ending March 2018 in England and Wales was the highest on record since data collection began in 1946. In 2019, over 100,000 people signed an online petition demanding a Parliamentary debate on knife crime.

Increases in knife crime in the West Midlands reflect the national trend; the Force has seen an increase in reporting from a low in 2012 to levels that were last seen in the early 2000s. Indeed, the mean number of monthly knife crimes (where they were used causing injury) in 2019 represents an increase of 160% over the mean number of monthly crimes in 2012.

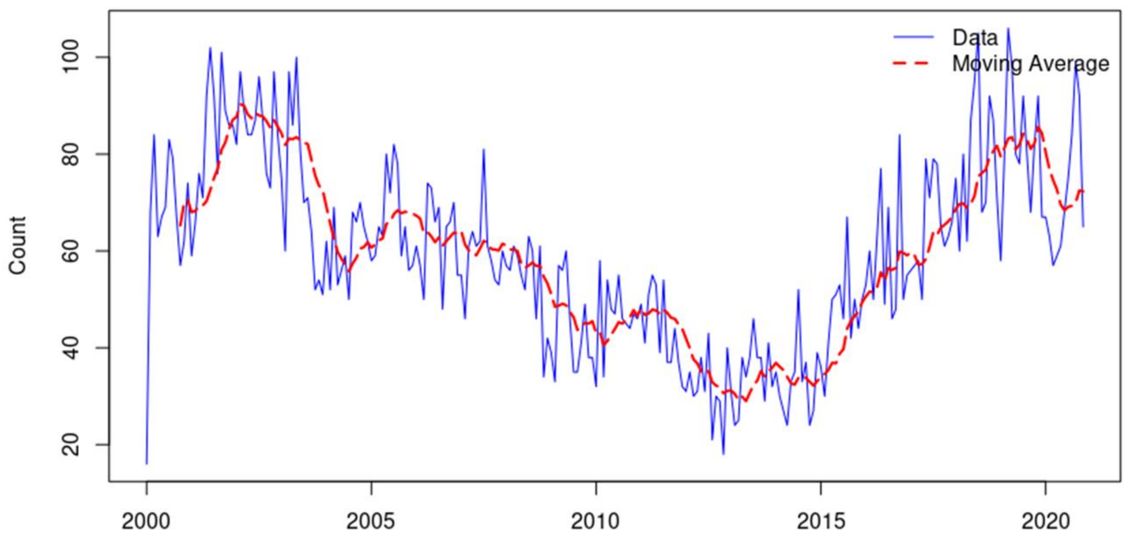


Figure 1: weekly knife crimes 2000 - 2020

In addition, the rate per 1000 residents of 'wounding (serious and other)' offending in the West Midlands is above the average when compared to our most similar forces of Merseyside and West Yorkshire (Office for National Statistics Crime in England and Wales: Police Force Area data tables year ending Sept 2019 – data for Greater Manchester currently unavailable).

In April 2019 the Home Office gave West Midlands Police (WMP) £7.62million in police surge funding with the mandate that it is to be used to reduce serious violence in public spaces, with a focus on reducing knife crimes among young people. The force's response has been to create a two year project, known as Project Guardian. The team works closely with the Violence Reduction Unit (VRU) launched the same year by the Police and Crime Commissioner (PCC) and other local agencies.

3 Exploratory (Spatial) Data Analysis

The data used in this project relate to knife crime (which is any bladed weapon) where "used causing injury". For the exploratory (spatial) data analysis data going back to the year 2000 have been used.

In a spatial sense, the incidents of knife crime tend to exhibit clustering, particularly around Coventry, Birmingham and Wolverhampton (and to a lesser extent Walsall) city centres.

Figure 2: incidents of knife crime (left to right, top then bottom) 2000, 2012, 2019 and 2020

It can also be seen from the chart below that over recent years (2018 – 2020) there have been between 60 – 100 crimes per month:

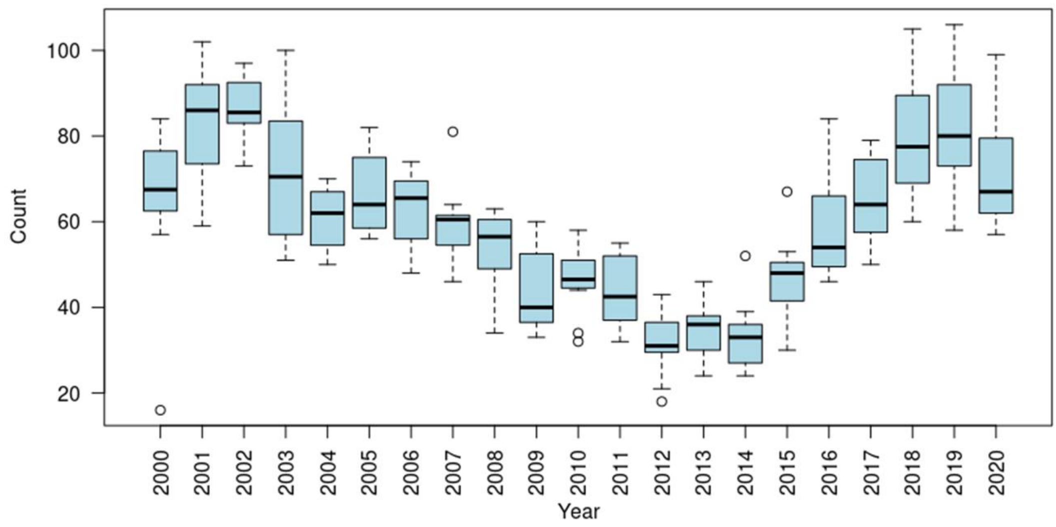


Figure 3: distribution of monthly count of knife crimes

Over the course of 2019, there have tended to be more knife crimes at weekends:

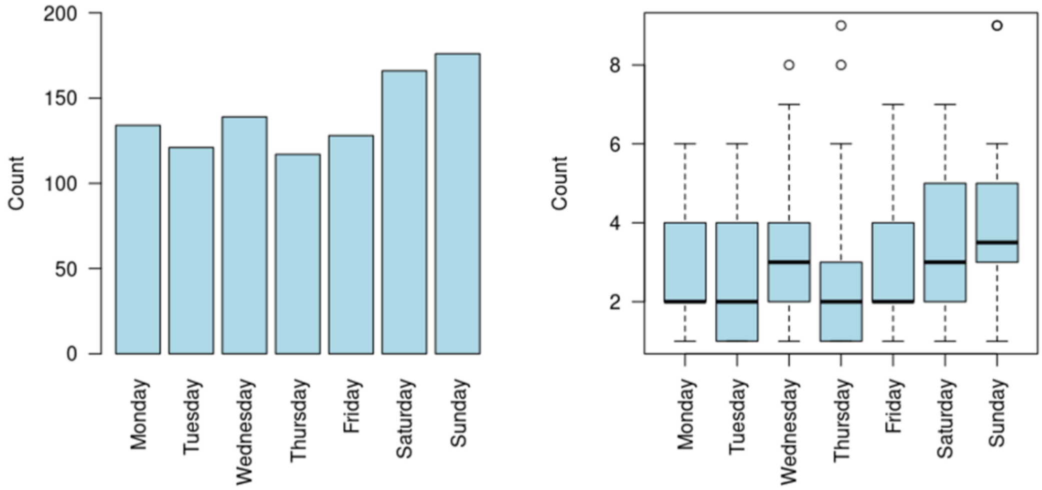


Figure 4: knife crimes by day of week, 2019 (total and distribution)

As can be seen from the charts below, depending on the day of the week, the majority of crimes occur either in the late afternoon / early evening or between circa 22:00 – 01:00.

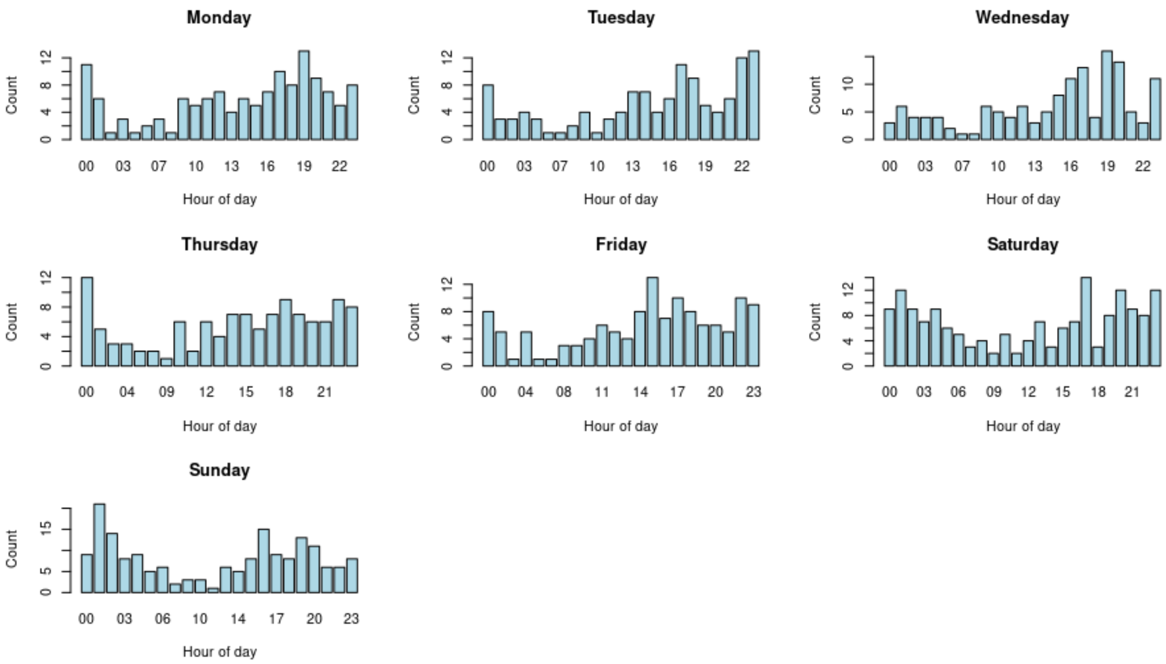


Figure 5: count of crimes by hour of day, 2019

On average, there is circa 10 hours between sequentially occurring knife crimes:

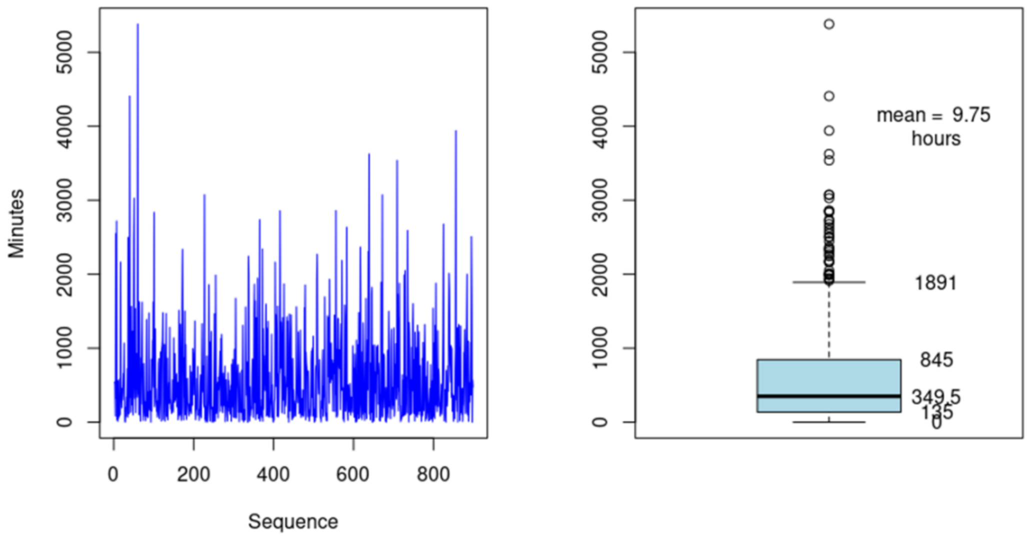


Figure 6: time between sequential knife crimes 2019

For those crimes neighbouring in time, the average spatial distance between them is circa 15 km:

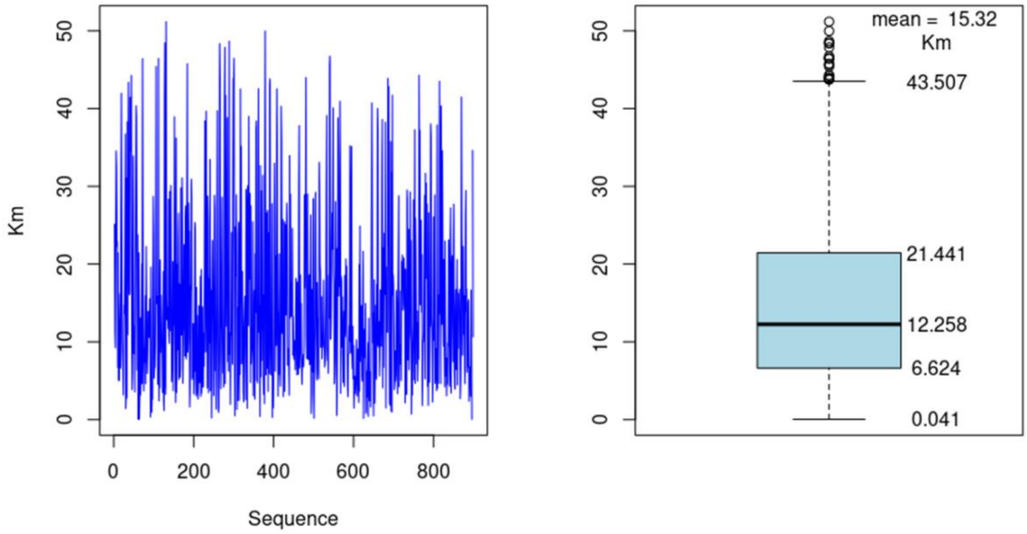


Figure 7: spatial distance between crimes neighbouring in time 2019

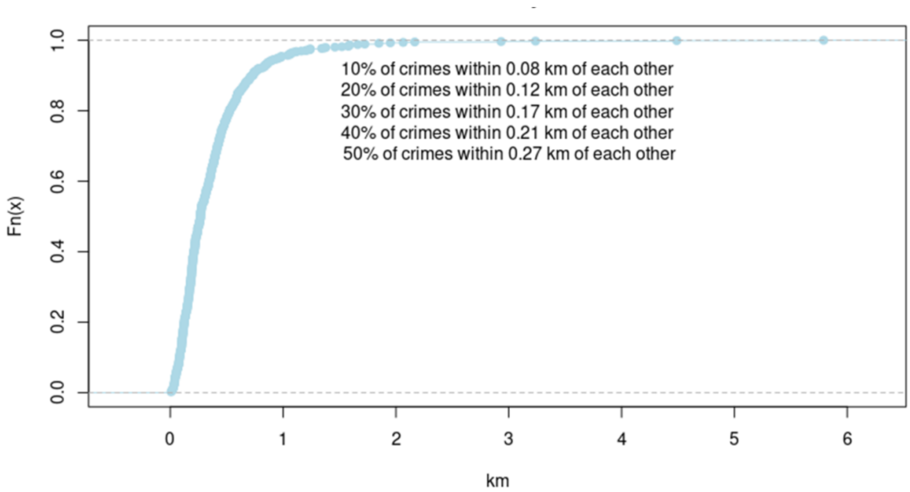


Figure 8: CDF of distance to nearest crime for crimes within 1 day of each other

For those crimes neighbouring spatially, the average time difference is circa 100 days:

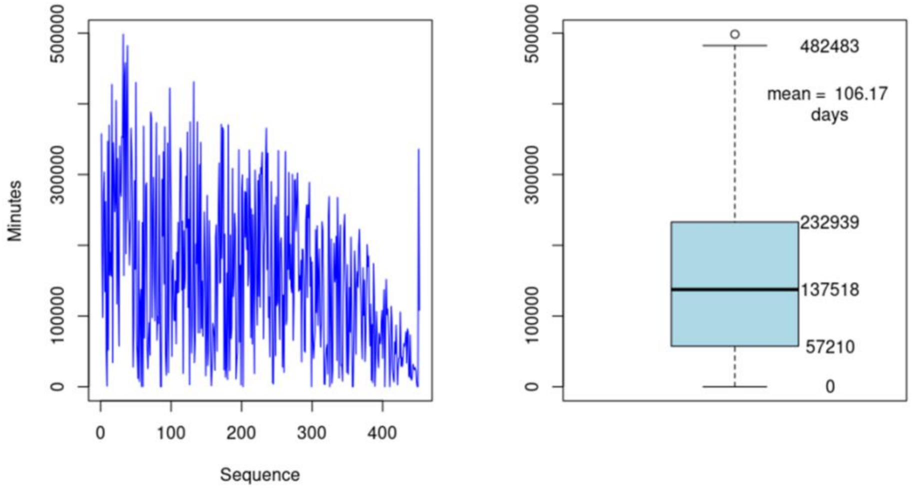


Figure 9: time difference between spatially neighbouring (a crime to its nearest neighbour) knife crimes 2019

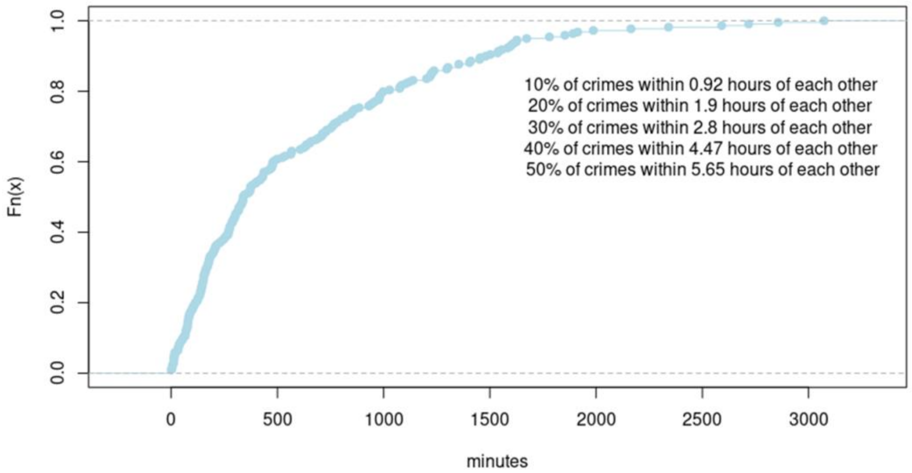


Figure 10: CDF of time between crimes for crimes within 0.15 km of each other

4 Modelling Approach

The preceding analyses and those contained in the appendix show that a spatio-temporal model using past levels of crimes as an additional feature is appropriate, augmented with a time series model. Details of both these analyses and the final models are in the appendix (where it is also shown that using the model(s) performs better than using a general hotspot approach).

Due to identifiable areas being most useful operationally, for the purposes of this project the WMP area has been split into a grid with each grid being circa 1 square km.

An example of the resultant output is shown in the figure below (for actual use, it is intended that this will be shown on a dashboard):

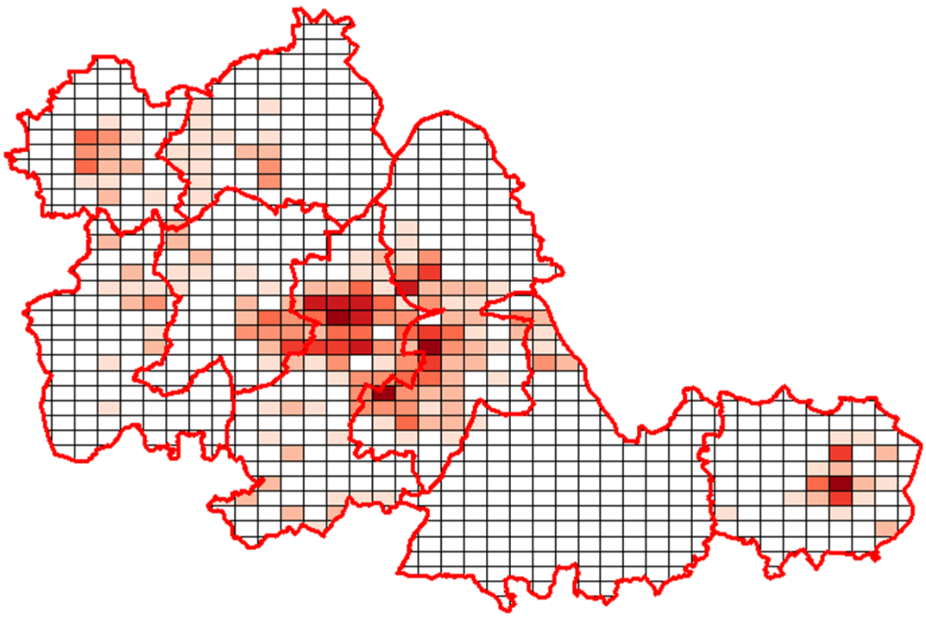


Figure 11: example of predictions for a 4 week period

Appendix – Modelling Details

From the previous sections it can be seen that there is correlation between knife crimes both over space and time. This appendix analyses this correlation and outlines the approach taken to the predictive modelling.

Whilst the majority of incidents occur within and around the main 4 city centres, there are incidents occurring outside the central areas and their surrounds. Despite this, when examined as a spatial point pattern (on a monthly basis), the knife crime incidents do not follow complete spatial randomness (CSR)[1]:

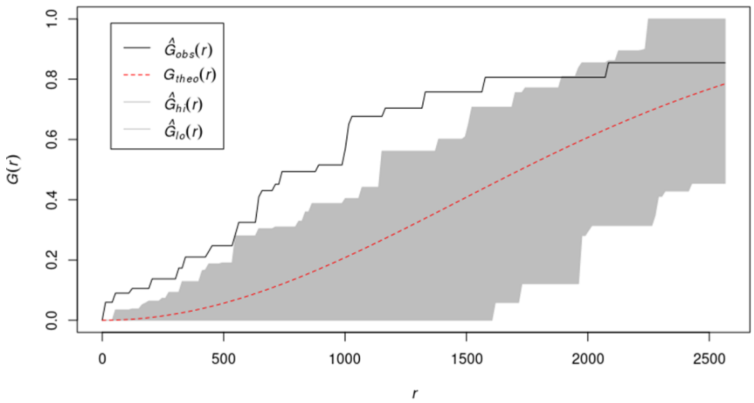


Figure 12: testing for CSR.

NOTE: this chart shows the result of testing of the locational patterns of knife crimes against a hypothesis of CSR; the black line being outside of the grey boundaries shows that there is clustering and the patterns of knife crime incidents do not follow CSR.

Due to identifiable areas being most useful operationally, for the purposes of this project the WMP area has been split into a grid with each grid being circa 1 square km. It is on this gridded pattern (and now therefore areal, as opposed to point pattern, data) that the modelling has been undertaken.

The figure below shows that when the knife crimes are examined for a 4 week period (October – November 2020), there is a small degree of global spatial autocorrelation:

[1] This is the case whether incidents inside the city centres are excluded or not and whether or not it is assumed that there is spatial non-stationarity.

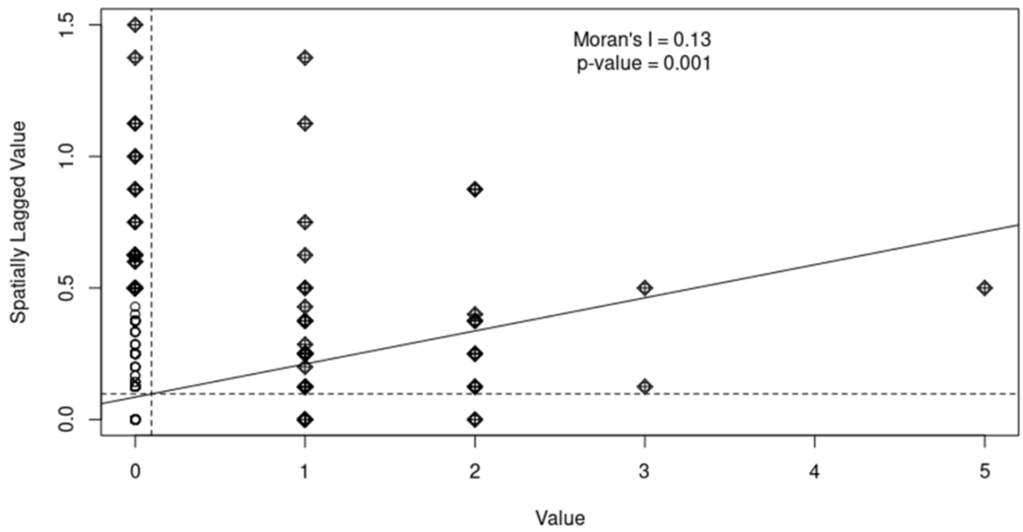


Figure 13: global spatial autocorrelation (4 week period Oct - Nov 2020)

A more localised analysis shows the presence of spatial clusters[2]:

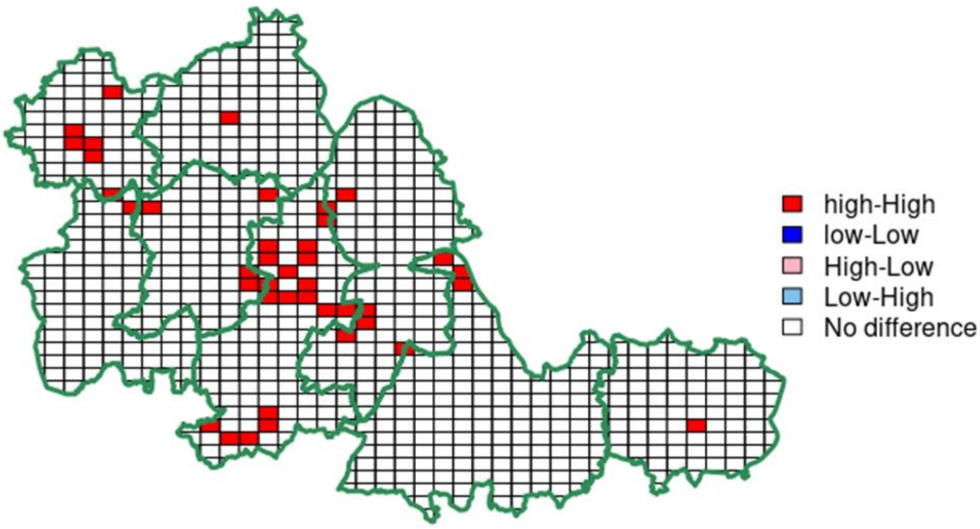


Figure 14: cluster map of local indicator of spatial association (4 week period Oct - Nov 2020)

On the time dimension, it can be seen from the chart below that there would appear to be a degree of seasonality amongst knife crimes, but of greater importance is an apparent long-term autocorrelation in the trend:

---

[2] It should be noted that these are locations of likely clusters, not knife crimes *per se*. Derived following calculation of a local indicator of spatial association statistic (see Anselin 1995).

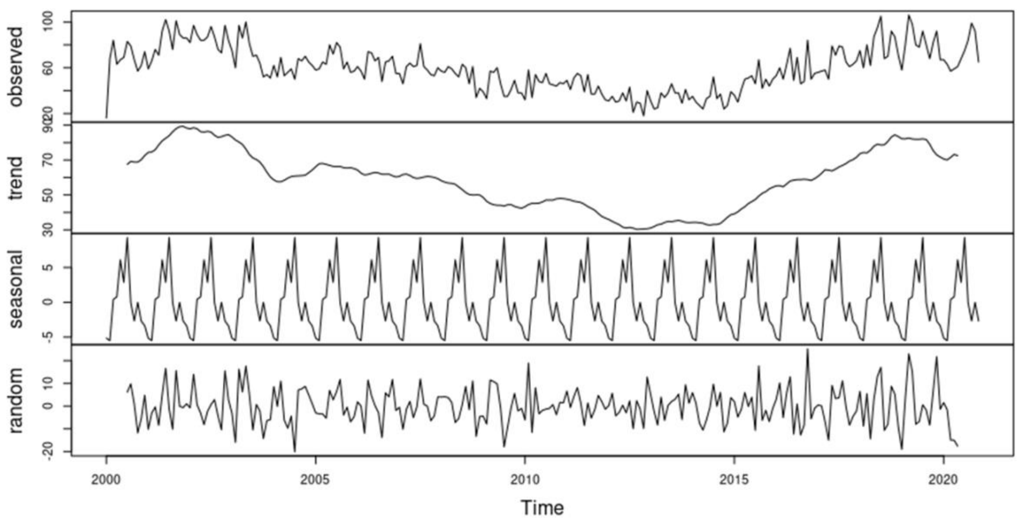


Figure 15:decomposition of knife crime time series

This long-term autocorrelation is confirmed in the ACF and PACF charts below:

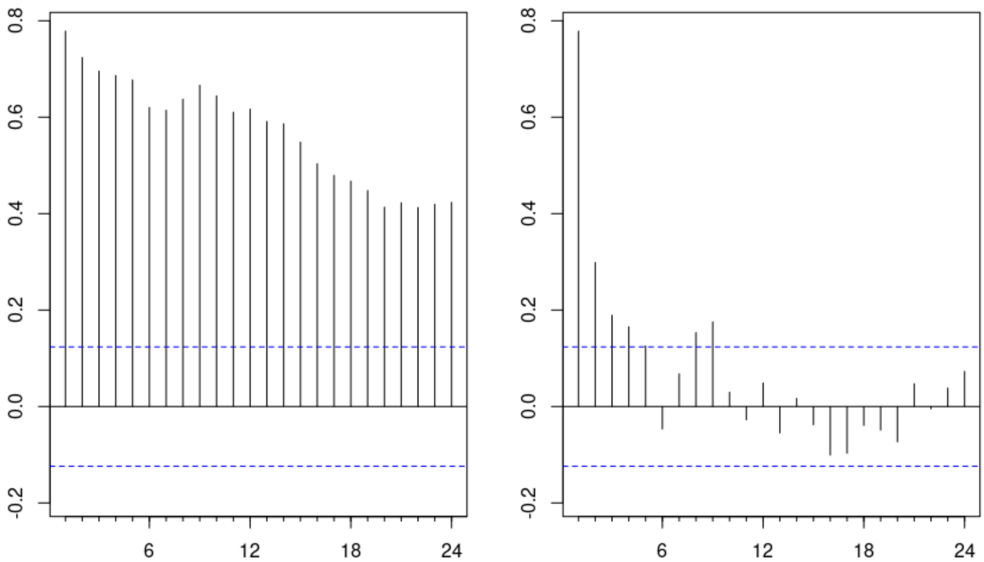


Figure 16: autocorrelation and partial autocorrelation functions of knife crime (lags on the x axis, correlation on the y axis)

Figure 18 below shows the relative score of a space time scan statistic. This shows that there are locations that exhibit higher than expected counts of knife crime incidents and that there are clusters of such areas.

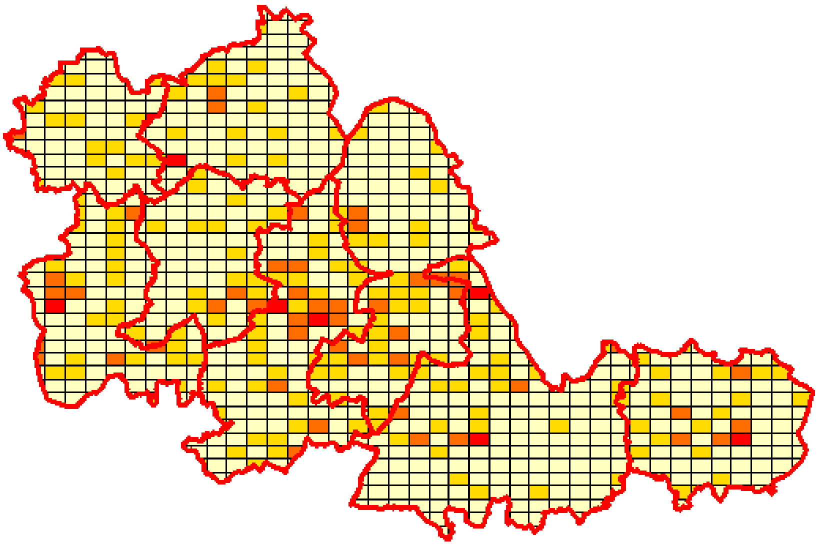


Figure 17: map of relative score of an empirical Bayes scan statistic

This points towards any relevant model being space-time inseparable[3] (i.e. some form of interaction between space and time will likely be useful within any such model).

From the above, there is information in both the temporal and spatial dimensions that can be useful for making predictions of knife crime incidents. Therefore, the approach taken here is essentially to use previous occurrences of knife crime incidents to predict (over a coming 4 week period) future knife crime incidents. This is akin to using a time series to predict future levels of that time series (an often used methodology). The data used are the count of 4 weekly knife crime incidents over the previous 2 years.

For additional information the levels of (previous time period) general crime are also used. The number of pubs, clubs, etc. has also been examined (which does not change over time), however these data do not provide useful information for predicting knife crime incidents.

The exploratory spatial data analysis (ESDA) undertaken shows that there is essentially correlation over both time and space of knife crime incidents and this therefore naturally leads to the use of a spatio-temporal model. The results of the scan statistic noted in figure 18 point towards the use of a space-time inseparable model[4]. Separable model(s) have been tested and it is indeed found that inseparable models are preferable[5].

---

[3] Haining, R. and Guangquan, L., 2020.

[4] A space-time separable model essentially models the data generation process as overall spatial + overall temporal whilst an inseparable model is applied to data generation processes that cannot be fully described by this structure; there is an interaction between space and time.

[5] The different types of models being tested by way of the deviance information criterion (DIC).

The final model chosen for the spatio-temporal elements used is a pCAR model with a type I space-time interaction[6], the basic form of which is (conditional on the y_{it}not being a structural zero):

$$y_{it} = (zeroinf)NB(\mu_{it}, \pi_t)$$

$$log(\mu_{it}) = a + S_i + v_i + \delta_{it} + \boldsymbol{x}_{i,t-1}\boldsymbol{\beta}$$

$$S_{i:N} \sim pCAR(\boldsymbol{W}_{sp}, {}^{1}/_{\sigma_S^2}, \tau)$$

$$v_{1:T} \sim pCAR\left(\boldsymbol{W}_{RW2}, {}^{1}/_{\sigma_v^2}\right)$$

$$\delta_{it} \sim N\left(0, {}^{1}/_{\sigma_\delta^2}\right)$$

$${}^{1}/_{\sigma_S^2} \sim loggamma(1{,}0.0005)$$

$${}^{1}/_{\sigma_v^2} \sim loggamma(1{,}0.0005)$$

$${}^{1}/_{\sigma_\delta^2} \sim loggamma(1{,}0.0005)$$

$$\tau \sim loggamma(1{,}1)$$

$$\pi_t \sim loggamma(1{,}1)$$

$$a \sim N(0{,}0.01)$$

$$\boldsymbol{\beta} \sim N(0{,}10)$$

Where a zero inflated negative binomial distribution is used due to the large number of zeros, $S_{i:N}$ is the spatially structured element of the pCAR model, v_i is the temporal element, $\delta_{it}$is the space-time interaction element, $\boldsymbol{W}_{sp}$ is a spatial weights matrix employing first order rook contiguity and $\boldsymbol{W}_{RW2}$ is the adjacency matrix for the random walk of order 2 temporal component. The pCAR model is the intrinsic conditional autoregressive model (essentially the number of knife crime incidents in a location is a function of the numbers in neighbouring areas). The additional covariate (the $\boldsymbol{x}_{i,t-1}$) is the lag of the number of (general) crimes.

In order to assess the choice of final model a spatio-temporal model with no additional feature was compared with a spatio-temporal model with the additional lag of general crime feature. It was found that the additional lag of general crime feature improved the model (assessed by way of comparing the DIC). Different formulations of the basic spatio-temporal model were assessed (the iCAR, pCAR and Besag, York, Millie (BYM) models). Whilst there was little difference, the pCAR formulation generally had the lower DIC.

Once the basic form of the spatio-temporal model was identified, different formulations of the temporal element were assessed, notably a random walk of order 1, a random

[6] The type of space-time interaction has been tested for by way of the DIC.

walk of order 2 and an auto-regressive (AR) of order 1. A similar model specification to these but using a generalised additive model (GAM) with multiple smooth terms (a version with and without spatial interaction) were also assessed. These were compared using the root mean square error / mean absolute error (RMSE / MAE) where predictions had been made for a total of 5 (4 week) time periods:

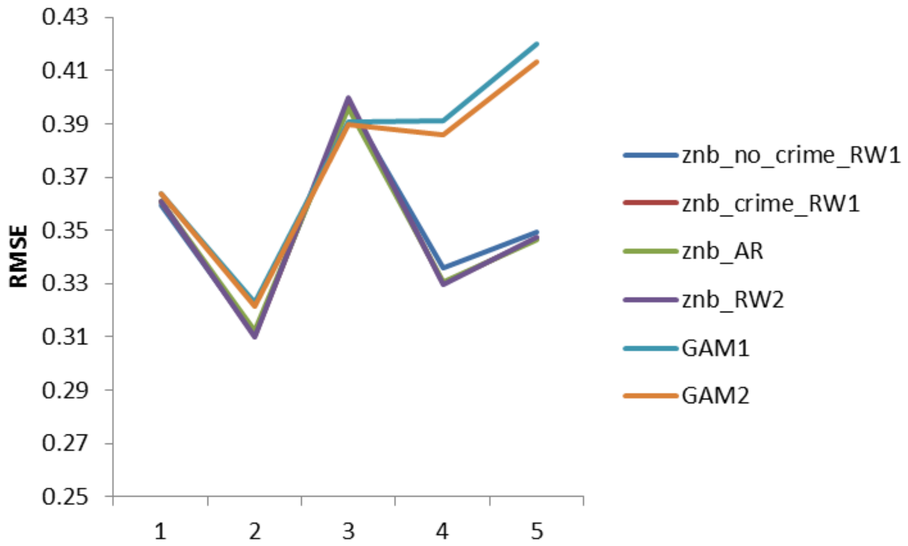


Figure 18: RMSE of different models

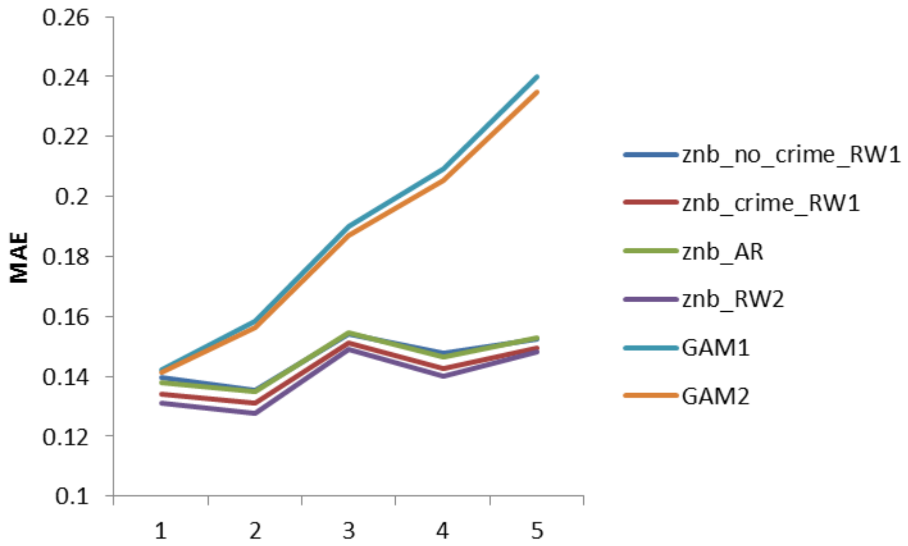


Figure 19: MAE of different models

Although the models follow similar patterns in the errors (and for actual usage only 1 time period would be predicted) it can be seen that the ZNB spatio-temporal model with a random walk of order 2 consistently produces lower errors[7].

[7] Different priors were also tested and made essentially no difference.

Knife crime resulting in injury is a relatively rare crime compared to other types of crime and this means that within the grid pattern used there is a large number of zero entries in any one time period. This results in a consistent under-estimation of the number of potential knife crimes in each of the locations (although the location of potential knife crimes is estimated well, see later). Because of this, the time dimension has been forecasted separately.

For the time series forecasting, an ensemble of neural nets, a seasonal and trend decomposition using Loess forecasting (STLF) and a Bayesian Structural Time Series (BSTS, aka a dynamic linear model) were used (all including a lag of general crime as an added feature) and compared:

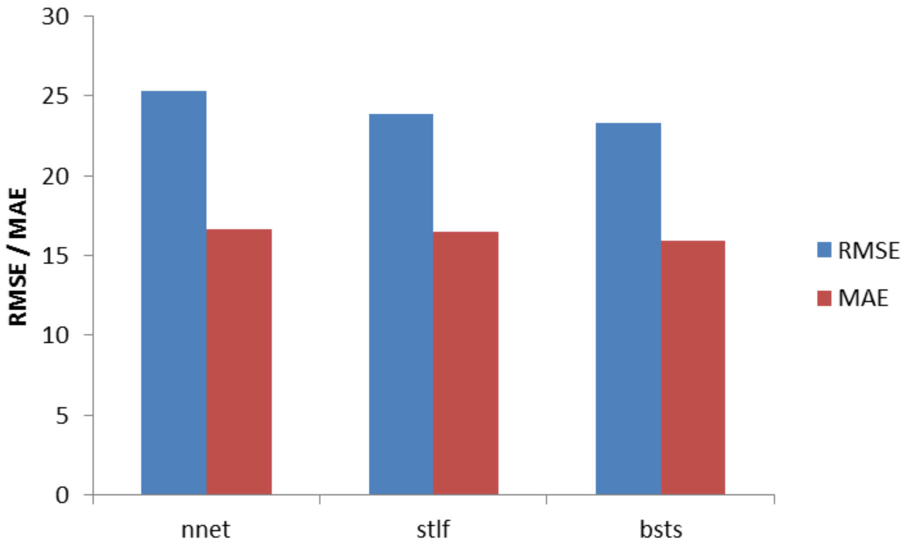


Figure 20: comparison of time series forecasting models

It can be seen that over the 5 period forecasting window the BSTS model produced the lowest errors.

In terms of locations, because of the large number of zero entries the spatio-temporal model has a tendency to allocate small numbers to a large number of locations. To minimise the potential for allocating resources to locations where a knife crime is actually unlikely, only those areas where the predicted number of knife crimes is at or above the 80th percentile of the distribution of predicted crimes are used. This is based on assessing the predictions as a binary yes / no (i.e. did an actual knife crime occur in the location v. did the model predict a knife crime to occur in the location).

Using the 80th percentile on this basis leads to a sensitivity of between 61% - 72% (over the 5 forecasting periods) and a specificity of 83% - 84% (an F1 score of 72% - 78%). Using all the forecasts produces a sensitivity of circa 100%, but a specificity of circa 27%.

Comparing different cutoff levels shows that the sensitivity generally decreases, whilst the specificity and precision generally increase:

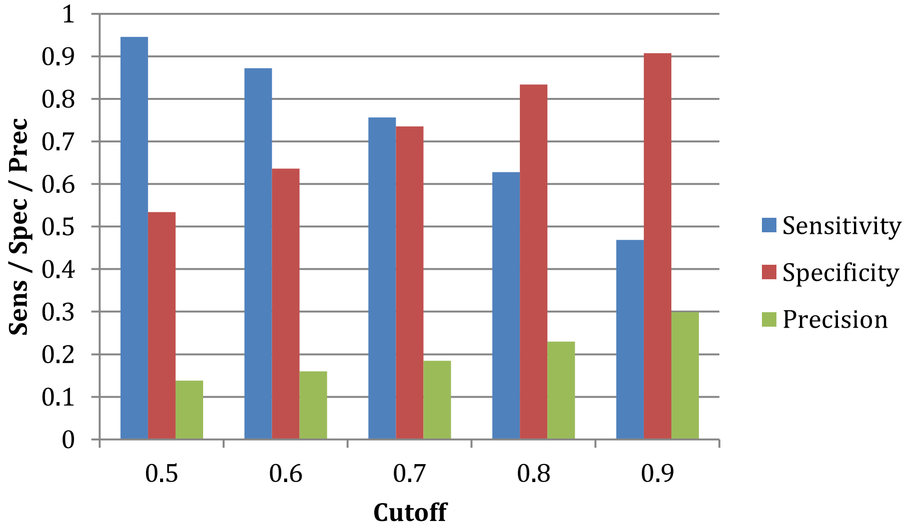


Figure 21: accuracy measures over different cutoff levels

However, if the costs of police time are taken into account (where APC = average PC salary in ((Precision * APC) – (False Positive Rate * APC)) * (Sensitivity – 0.5) where the latter term penalizes cutoff levels where the true positive is below 0.5 shows that 0.8 is the optimal cutoff:

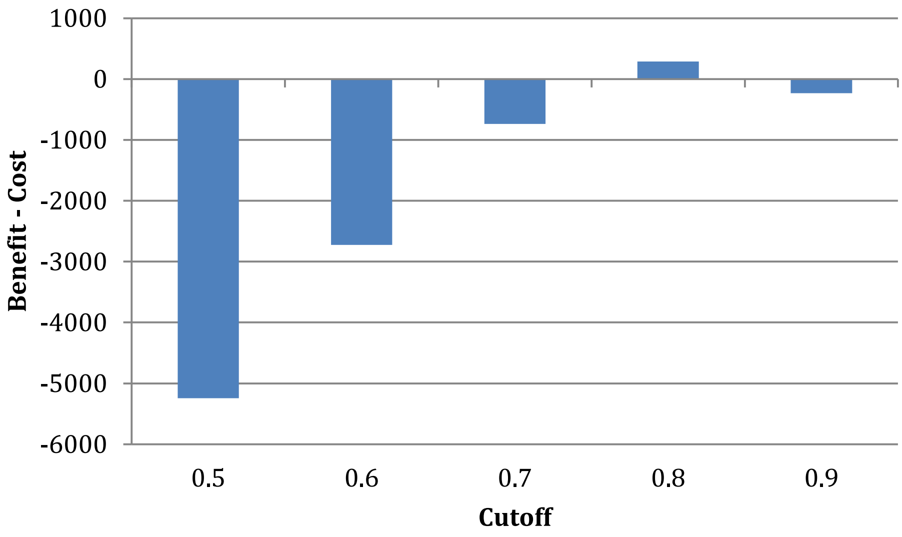


Figure 22: benefit - cost over different cutoff levels

Essentially the number of locations where a prediction is made is reduced.

The predictions for this reduced number of areas are then augmented with the time series model so that the total number of predicted crimes sum to the same value.

As a check on the efficacy of this approach (and bearing in mind the findings of the exploratory spatial data analysis), the predictions arising from the model(s) are compared to those arising from a naïve approach, namely using the mean number of crimes over the previous 2 years to identify those locations with a greater likelihood of a crime occurring.

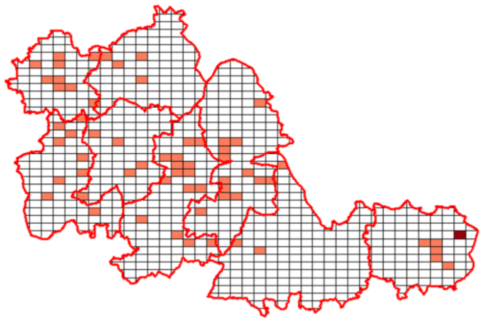


Figure 23: actual knife crimes 1st forecasting period

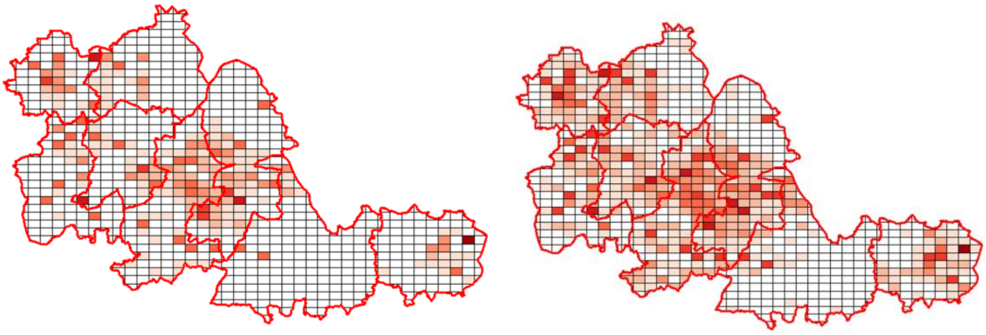


Figure 24: residuals from model (left) and application of mean (right)

It can be seen that using the mean number of knife crimes produces a larger number of locations where a prediction was made and therefore using this produces a lower specificity and could lead to the misallocation of resources. Calculating the weighted average predicted error (WAPE) arising from the model and comparing to that using the mean shows the model producing between 8% and 16% lower errors.

Using the distribution of the mean level of knife crime, these can also be examined by way of the 80th percentile (of the distribution of the mean):

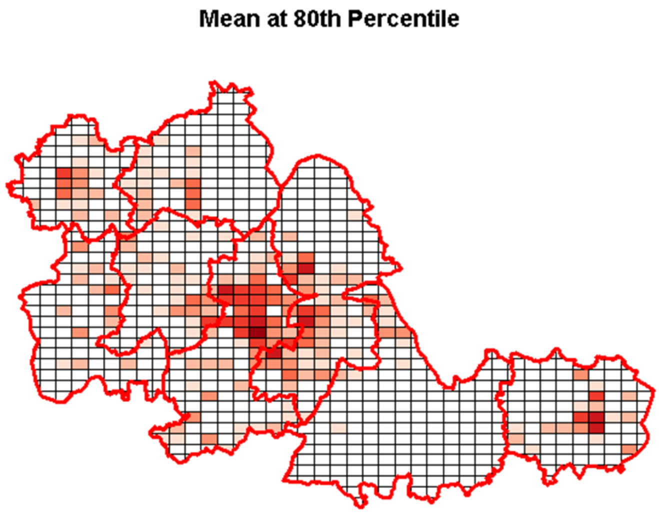


Figure 25: distribution of mean of crimes (80th percentile)

This also reduces the number of observations.

Taking this approach improves the sensitivity but reduces the specificity. On average, the WAPE is 80% better using the model and the benefit – cost measure of effectiveness whilst positive using this approach (at 82) is still less than that generated from the model.

References

Anselin, L., 1995, Local Indicators of Spatial Association – LISA, *Geographical Analysis*, Vol. 27(2), pp. 93 – 115.

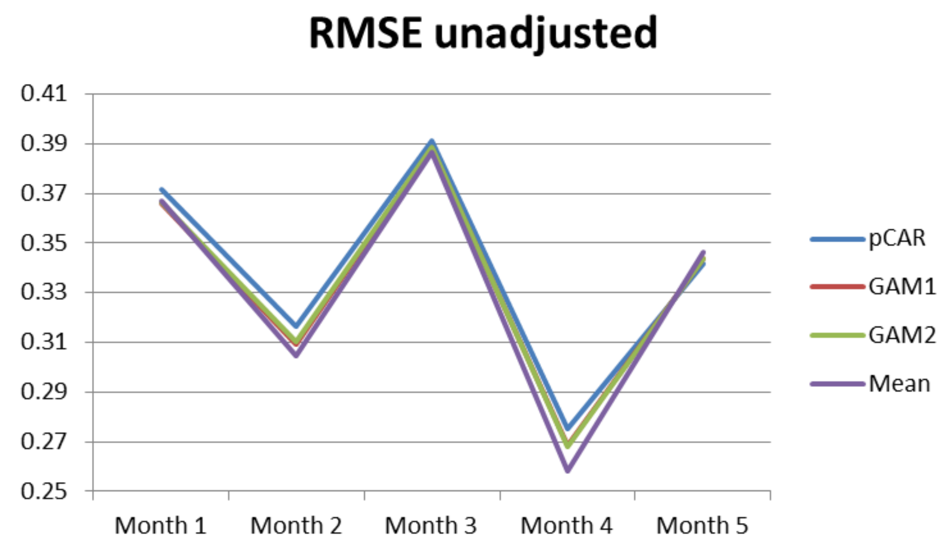
Haining, R. and Guangquan, L., 2020, *Modelling Spatial and Spatial-Temporal Data A Bayesian Approach*, CRC Press, Boca Rotan, Fl.

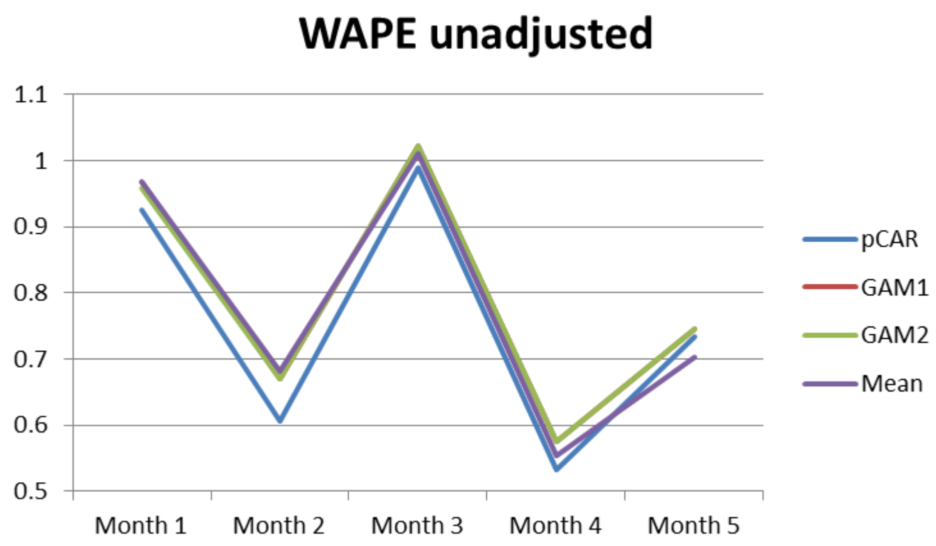
5 Addendum

In view of comments raised at (and post) the last Ethics Committee of early December, the analyses below show the results of applying the model to a moving window of 5 months (actually 4 weekly periods) using data from the 2 years previous to each prediction period (which start from November 2020 through to early March 2021).

For comparison, the model chosen above (the proper conditional autoregressive model with a random walk of order 2 for the temporal dimension) is compared to a generalized additive model (GAM) using splines to smooth through space and time as well as using the mean number of crimes per square area as a means of a more traditional "hotspot" (naïve) methodology.

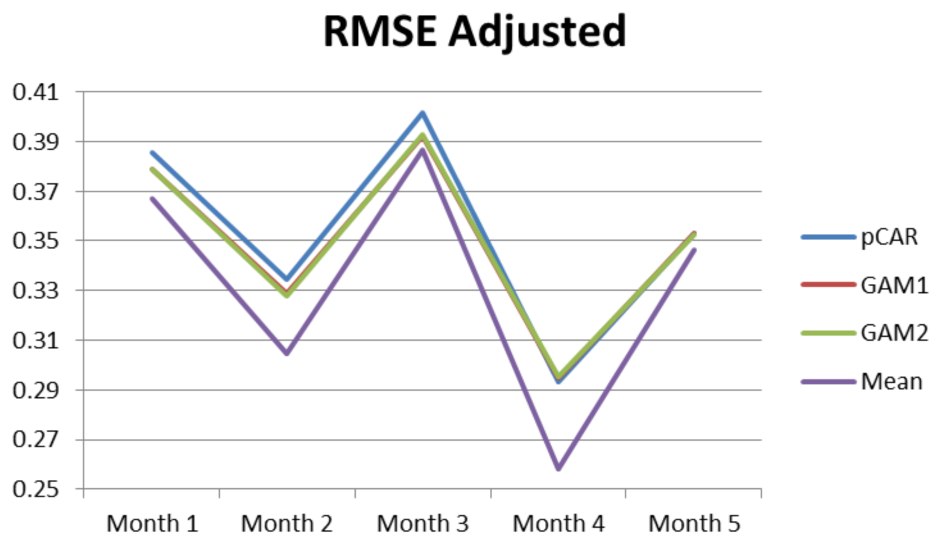
Using the unadjusted output (see above) the results are:

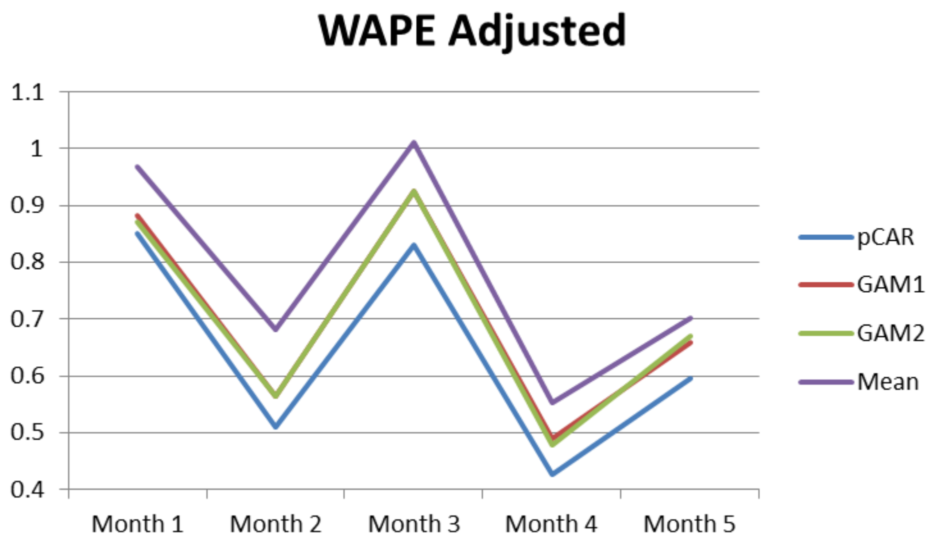




Using the RMSE, the mean number of crimes has the lowest errors, but using the WAPE (comparison of positive valued predictions and observations) shows that the pCAR model produces the lowest errors.

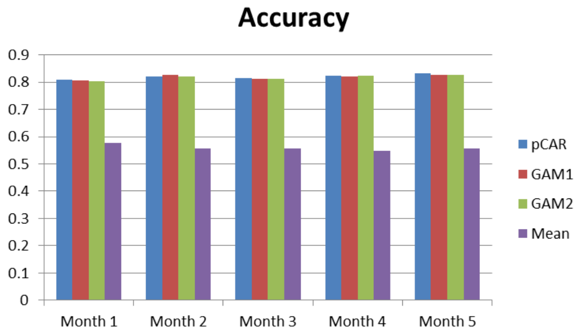
However, as noted in the main report, due to the sparse nature of knife crime and the consequent sparse occurrences, an overall better model is produced whereby predictions at the 80th percentile are used and adjusted in light of the predictions of a time series model (predicting the total number of knife crimes for each 4 week period). These are termed the "adjusted" predictions:

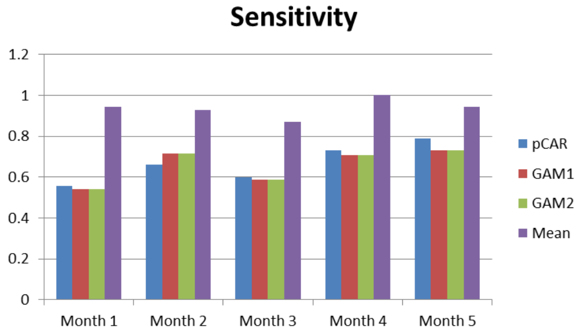


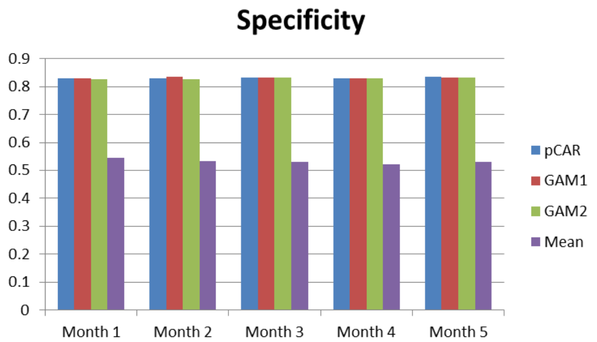


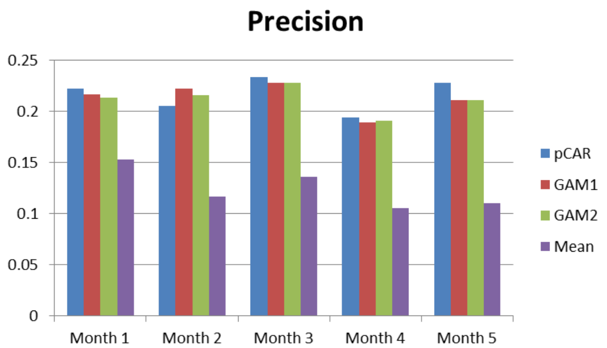
Qualitatively, the result is the same as for the unadjusted predictions, however it will be noted that the overall WAPE levels are now lower for the models, with the (adjusted) pCAR model having the lowest overall WAPE.

Due to the sparse nature of knife crime, it is also instructive to examine the predictions in terms of a dichotomous whether a knife crime occurred / didn't occur viewpoint.









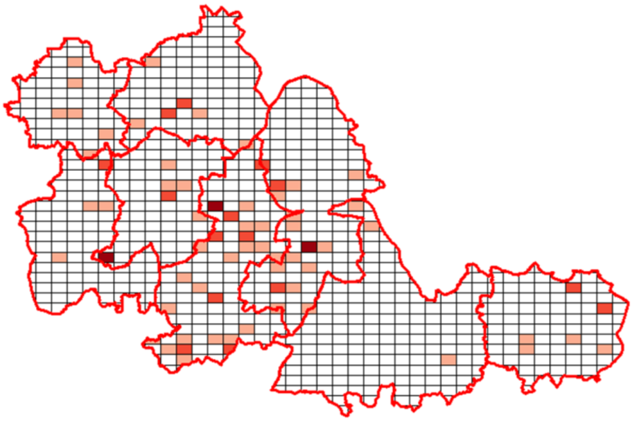
From these charts it can be seen that in terms of overall accuracy, the naïve approach fairs less well than the modelling approach (which are all similar).

The naïve approach does provide better sensitivity, but this comes at the expense of a far worse specificity. Use of the naïve approach would therefore likely lead to a wasteful use of resources in that many areas would be identified as requiring attention when in fact they are unlikely to see a knife crime.

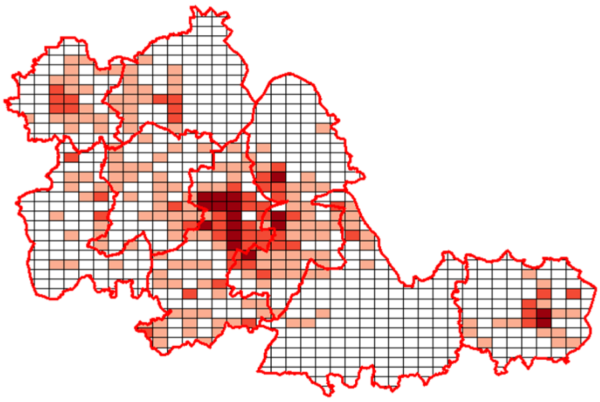
The modelling approach also produces a better precision than the naïve approach with the pCAR model having slightly better precision in 4 out of the 5 months than the GAM models.

Examples:

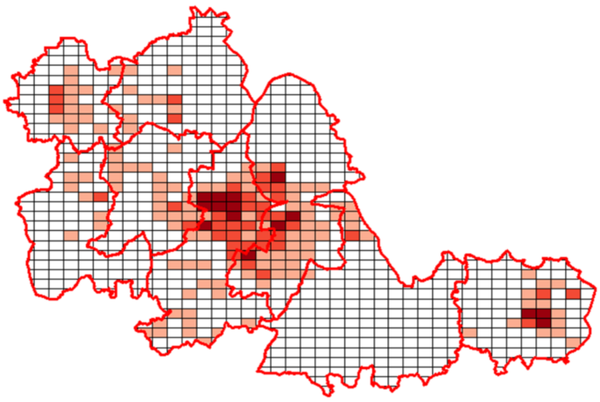
Actual Knife Crime
Month 1



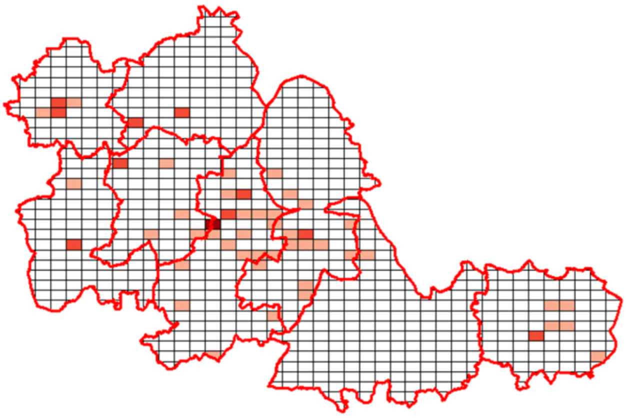
Using Mean Knife Crime
Month 1



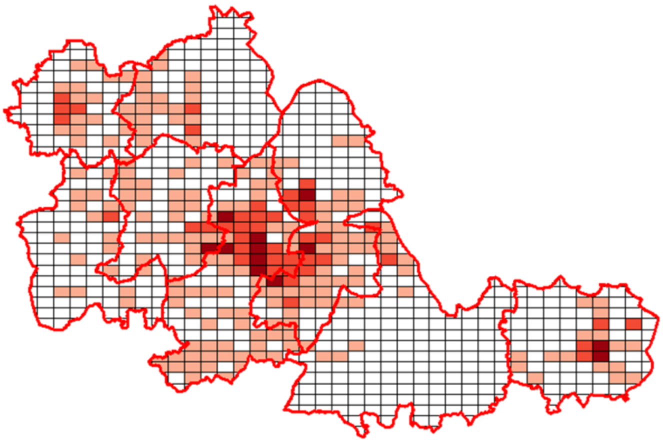
Predictions, Knife Crime
Month 1



Actual Knife Crime
Month 5



Using Mean Knife Crime
Month 5



Predictions, Knife Crime
Month 5