# Analysis of weather effects on daily road accidents

***Jennifer Davies***
*Office for National Statistics*
*jennifer.davies@ons.gov.uk*

**This article explores the relationship between rain and temperature and daily counts of killed or seriously injured vulnerable road users in Great Britain.**

## Summary

This analysis follows on from previous studies by ONS and DfT (GSS, 2015, DfT, 2015) using monthly road accident and weather data to assess the impact of temperature and rain on road accidents in Great Britain. The GSS (2015) study analysed data on road accidents where a vulnerable road user (pedestrian, cyclist or motorcyclist) was killed or seriously injured (KSI). It found that warmer than average temperatures coincided with more accidents in most months of the year. The DfT (2015) study extended this work to analyse road accidents for different road users and to include rain in the models.

Following on from GSS (2015), this article presents analysis of daily counts of killed or seriously injured vulnerable road users (KSIVU). Graphical analysis, simple averages of counts on different weather-type days and ARIMA modelling have been used to understand the relationship. Graphical analysis shows that the seasonality in temperature and KSIVU are similar but temperature cannot fully explain all of the seasonal fluctuations. The analysis of simple averages shows that wet days have fewer KSIVU than dry days and the coldest days have fewer KSIVU while the warmest days have more. The time series model supports the analysis of simple averages and suggests that for each millimetre more rain than average there are 0.75 fewer KSIVU and for each degree warmer than average there are 0.96 more KSIVU.

## Data

### Road accident data

The STAT19 dataset contains records of all police-recorded road accidents and casualties in Great Britain. These records were aggregated to provide daily counts of KSIVU.

Figure 1 plots the daily KSIVU time series from 2005 to 2014. There is a strong seasonal pattern in the time series with the number of KSIVU at its highest during the summer. The series drops each year on Christmas day and remains low over the Christmas period. There is a slight downward trend over the ten year period. Figure 2 shows boxplots of the number of KSIVU by month. This shows that KSI increase through the year, peaking in the summer; however it also shows that the number of KSIVU dips in August compared with July and September.

There is also a day of the week effect in the KSIVU data. This can be seen in the boxplots in figure 3. The median level is highest on Friday and lowest on

Sunday.

### Weather data

This analysis used mean daily temperature and rainfall in the United Kingdom. These were provided by the Met Office. There is a slight discrepancy in that the road accident data is for Great Britain and the weather data for the United Kingdom, however weather data was not readily available for Great Britain.

Daily meteorological observations are measured from 9am one day to 9am the next day, as opposed to midnight to midnight. One limitation of this analysis could be that the road accidents refer to calendar days while the weather data does not. This could be overcome by aggregating the road accident data so that days start at 9am.

Figure 4 shows the mean daily temperature between 2005 and 2014. Visually the overall trend of the series appears flat over the ten-year period. Temperatures are seasonal with peaks in the summer and troughs in the winter. Figure 7 provides boxplots of daily temperature by month. These boxplots also show that temperatures increase through the year to peak in July then decrease. They also show that the variability in temperature is different for each month. Typically the range of temperatures in summer months is smaller than the range of temperatures in winter months.

Figure 5 shows the daily rainfall between 2005 and 2014. It is difficult to visually identify any patterns in the rainfall data. Figure 6 provides boxplots of daily rainfall by month. This figure shows that the monthly median rainfall does vary as does the range of rainfall amount.

## Initial analysis

### Graphical analysis

Figures 1 and 4 showed that both KSIVU vulnerable road users and temperature show seasonal patterns. Figure 2 provides boxplots of the daily counts by month and figure 7 boxplots of temperature by month. While both increase through the year and peak in the summer, the number of KSI vulnerable road users dips in August, indicating that while temperature could explain some of the seasonality in KSI vulnerable road users, it does not explain all of it.

It is more difficult to see any relationship between rain and KSIVU vulnerable road users from the boxplots in figures 2 and 6. The range of rainfall is generally larger in the winter months and the median rainfall is higher in October through to February than the rest of the year. If there is a rain effect then the boxplots suggest that months with more rain correspond to months with fewer KSIVU. The median rainfall in September is lower than August and October coinciding with the increase in KSIVU in September compared to these months. This could also suggest that less rain corresponds to more KSIVU. However, overall the link between rain and KSIVU is less clear from the graphical analysis.

### Averages on different types of day

The study of BRSI (2014) compared the average number of accidents on days grouped by their weather, for example the average number of accidents on wet days compared to dry days.

A similar analysis was run on the KSIVU data. For rain, wet days were defined as days where the rainfall was 1mm or greater. Table 1 shows the average number of KSIVU on wet and dry days by month. In all months except for January and December, the average numbers of KSIVU are lower on wet days the dry days. The differences are significant at a 5% level from February through to October.

For temperature, mean daily temperature was calculated for each month and warm days were defined as days with a mean temperature above average and cold as days with a mean temperature below average.

Table 2 shows the average number of KSIVU by month on warm and cold days. In all months there are more KSIVU on warm days than cold days and the differences are significant at a 5% level from November through to July.

This analysis indicates that there could be a relationship between weather and KSI vulnerable road users, however the results of the significance tests are invalidated because the observations are not independent, identically, normally distributed. This can be seen, for example, in the model containing just wet and dry days. The significance test was carried out by fitting a linear least squares regression model with monthly indicator variable for whether the day was wet or dry. The autocorrelation function of the residuals is plotted in figure 8. The model residuals are not independent as they display significant autocorrelation, particularly at lag 7 indicating that a day of the week effect is present in the residuals.

## Time series model

Using a simple linear least squares regression approach did not properly account for the structure of the KSI vulnerable road users time series. In GSS

**Figure 1:** *Daily killed or seriously injured vulnerable road users, Great Britain, 2005-2014*

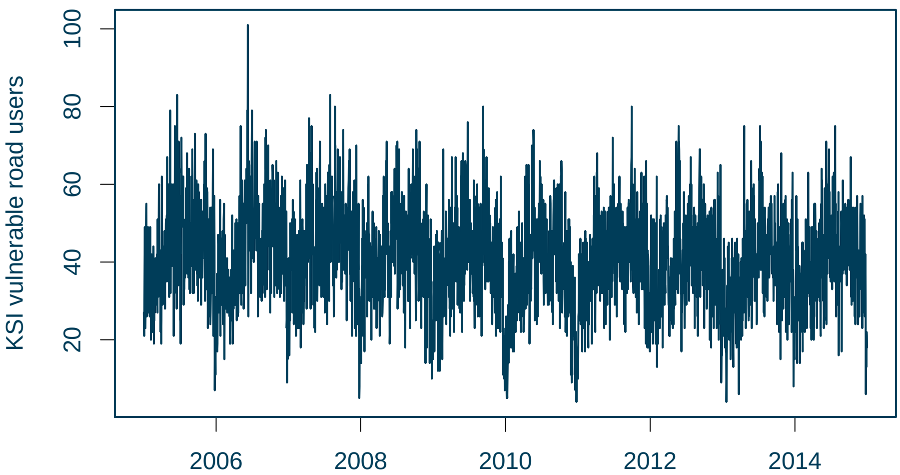



**Figure 2:** *Boxplots of daily killed or seriously injured vulnerable road users by month, Great Britain, 2005-2014*

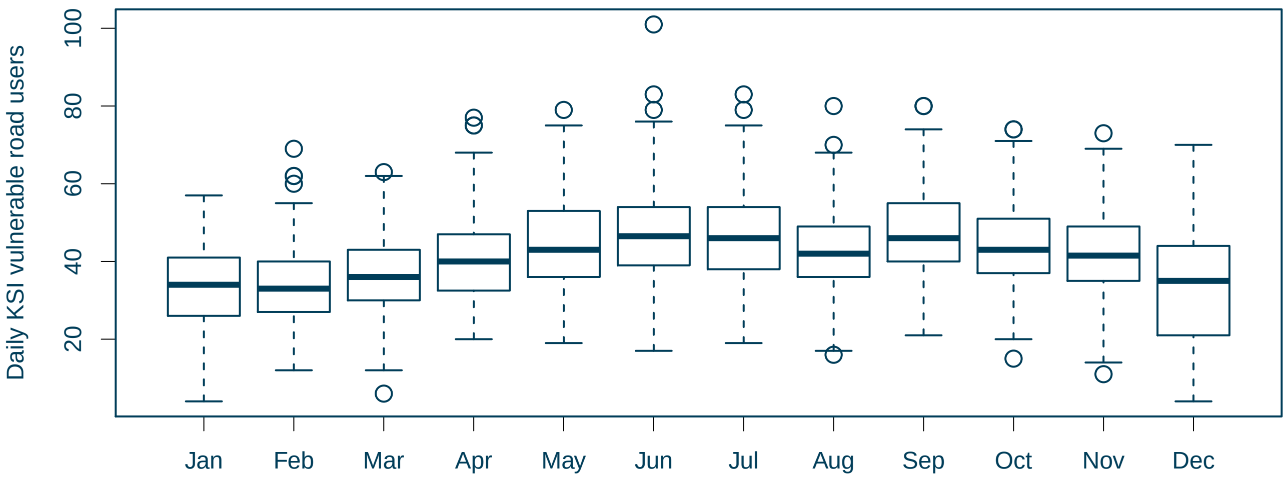



**Figure 3:** *Boxplots of daily killed or seriously injured vulnerable road users by weekday, Great Britain, 2005-2014*

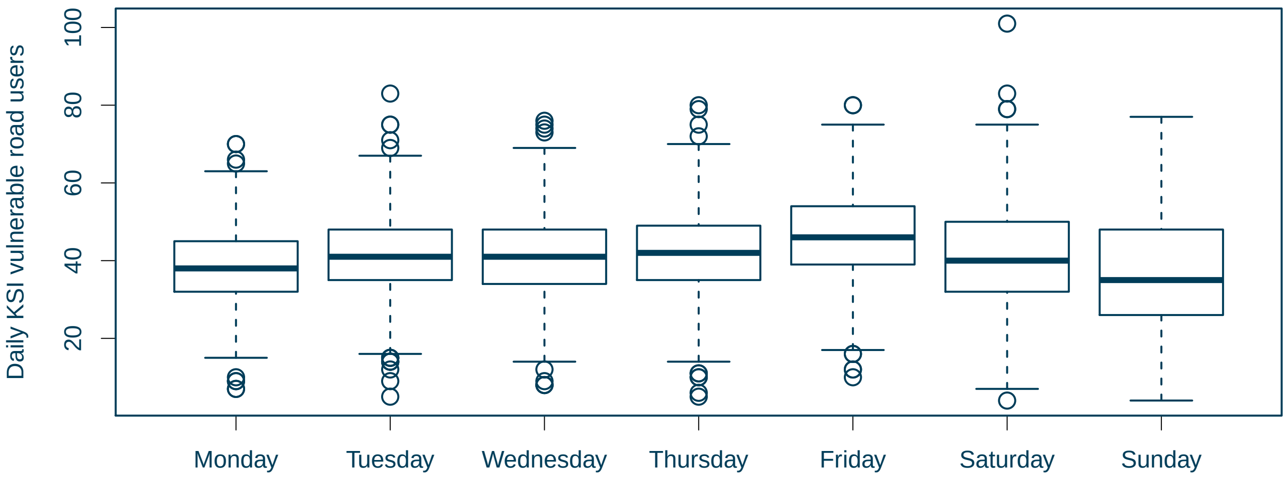


**Figure 4:** *Mean daily temperature in the United Kingdom, 2005-2014*

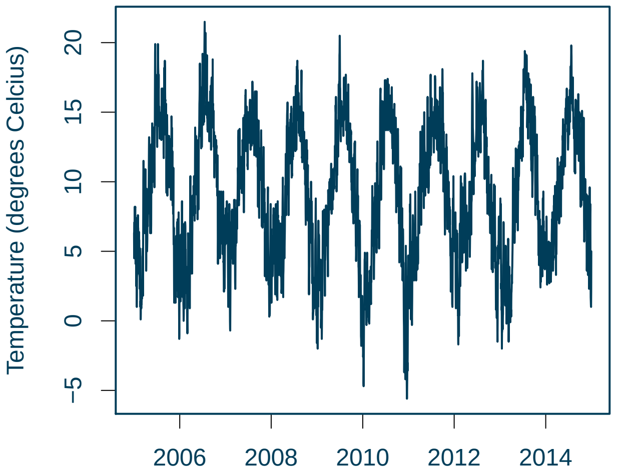



**Figure 5:** *Daily rainfall in the United Kingdom, 2005-2014*

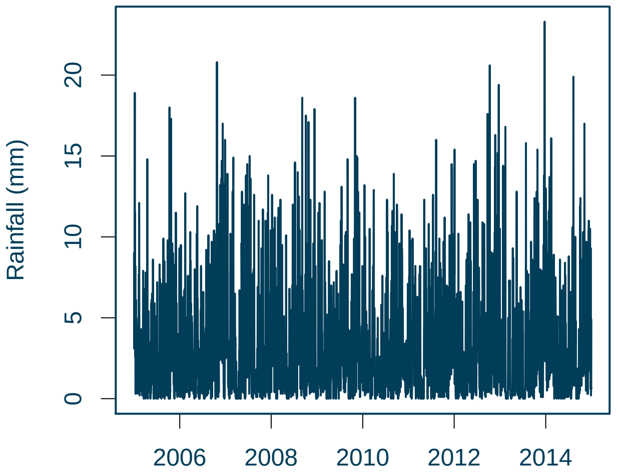



**Table 1:** *Average number of KSI vulnerable road users by month on wet and dry days. Significance levels* *** $< 0.001$, ** $< 0.01$, * $< 0.05$

| Month | Dry | Wet | Significant |
|---|---|---|---|
| Jan | 32 | 34 | |
| Feb | 36 | 32 | ** |
| Mar | 39 | 36 | * |
| Apr | 44 | 38 | *** |
| May | 51 | 41 | *** |
| Jun | 52 | 44 | *** |
| Jul | 52 | 43 | *** |
| Aug | 47 | 41 | *** |
| Sep | 51 | 46 | *** |
| Oct | 47 | 43 | ** |
| Nov | 44 | 41 | |
| Dec | 32 | 34 | |

**Table 2:** *Average number of KSI vulnerable road users by month on cold and warm days. Significance levels* *** $< 0.001$, ** $< 0.01$, * $< 0.05$

| Month | Cold | Warm | Significant |
|---|---|---|---|
| Jan | 30 | 36 | *** |
| Feb | 31 | 36 | *** |
| Mar | 34 | 40 | *** |
| Apr | 37 | 43 | *** |
| May | 41 | 47 | *** |
| Jun | 43 | 52 | *** |
| Jul | 44 | 50 | *** |
| Aug | 41 | 44 | |
| Sep | 47 | 48 | |
| Oct | 43 | 45 | |
| Nov | 40 | 43 | * |
| Dec | 30 | 37 | *** |

(2014), regARIMA models (regression models with ARIMA errors) were used to model the monthly KSI vulnerable road users time series and to test for the effects of temperature and rain. This method has been adapted to apply to daily data.

## ARIMA model

To account for the trading day effect, dummy variables and using a seasonal ARIMA model with period of 7 were both considered. Using dummy variables did not remove autocorrelation at lag 7 so a seasonal ARIMA model was used allowing autoregressive and moving average processes at lag 7.

Seasonal dummy variables were used to capture seasonality. 11 dummy variables were included, one for each month except December which can be derived. An alternative could be to model the seasonality as a trigonometric function.

Outliers have been included to account for the dip on Christmas day and the days between Christmas and New Year where the series gradually recovers to its previous level. Each day from Christmas day to the observed New Year bank holiday inclusive have been set as outliers.

As the seasonality in temperature and KSIVU is similar instead of including temperature, temperature deviations from the monthly average temperature were used. This is to ensure that the seasonality in KSIVU can be captured in the seasonal dummy variables rather than the temperature variable. Similarly rain deviations from the monthly average were included in the model.

**Figure 6:** *Boxplots of daily rainfall by month, United Kingdom, 2005-2014*

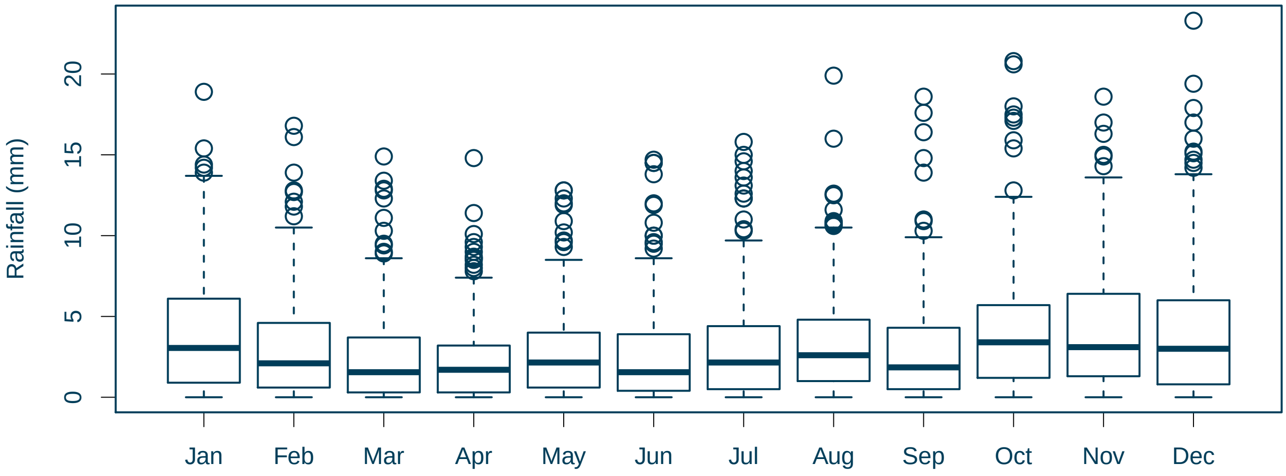



**Figure 7:** *Boxplots of daily mean temperature by month, United Kingdom, 2005-2014*

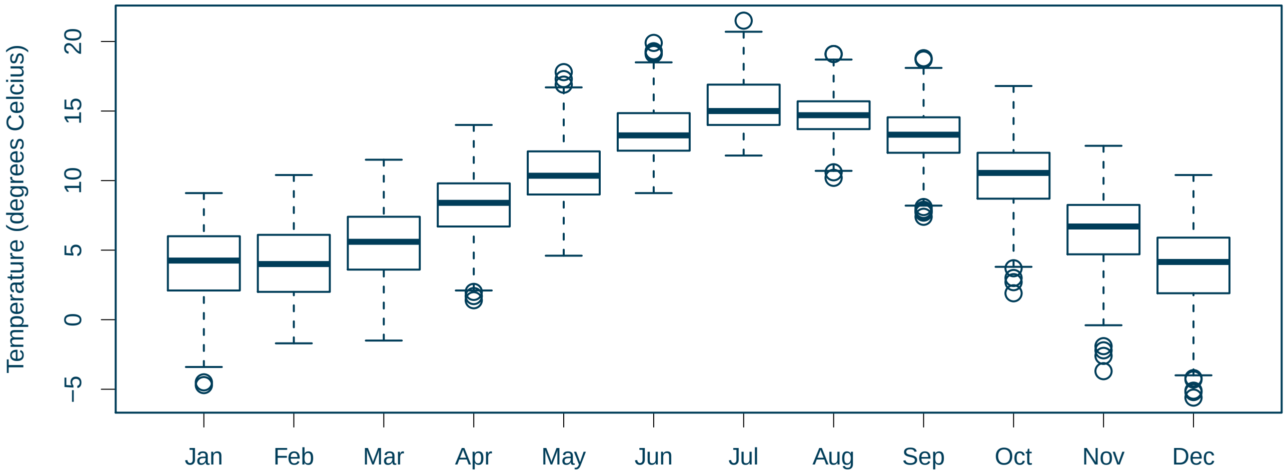
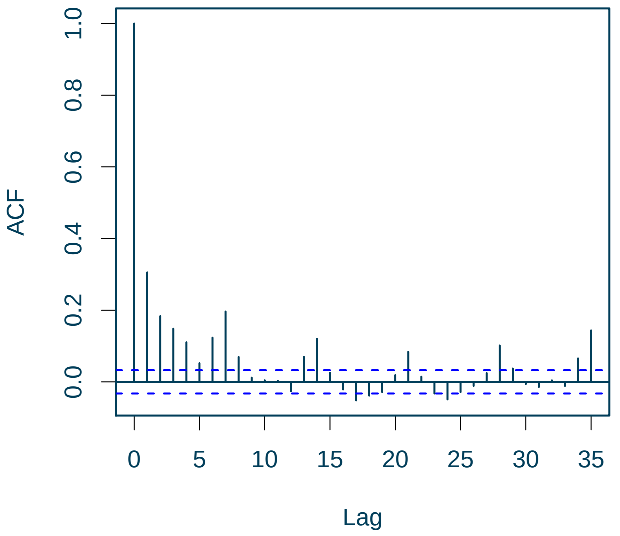


**Figure 8:** *Autocorrelation of residuals from model including monthly wet and dry day indicator variable*




## Results

The auto.arima function was used in R to fit the model. An ARIMA $(1,0,1)(1,0,1)_7$ model was fitted. The results for the weather regressor coefficients are displayed in table 3.

The results in table 3 show that in this model, both rain and temperature have a significant effect on the number of KSIVU. The interpretation of the model is that for, say May, every additional 1mm of rain above the monthly average, all else being equal, the model estimates 1.6 fewer KSIVU. For temperature, for every additional degree above the May average, all else being equal, the model estimates an additional 1.6 KSIVU.

The results suggest that rain has a significant impact in May through to September, and temperature in all months except for February and September through to November. The February coefficient for temperature has a t-value of 3.3 so was close to the critical value of 3.5 used to determine significance. Similarly, the coefficients for rain in March and April were close to the threshold.

The critical value was determined based on a 5% significance level t-distribution with $n$-$p$ parameters where $n$ is the length of the time series and $p$ is the number of estimated parameters including all regressors, ARMA parameters and the variance.

Figure 9 provides diagnostic information about the residuals of the fitted ARIMA model. The ACF plot and Ljung-Box statistics do not indicate any autocorrelation in the residuals. The standardised residuals do not display any obvious patterns. These suggest that the model fit is adequate.

**Table 3:** *Regression coefficients for months, rain and temperature effects from ARIMA model. Rain coefficients can be interpreted as the expected increase (or decrease) in casualties by each additional (or fewer) millimetres of rain compared with average. Similar for temperature the expected increase (or decrease) in casualties by each additional (or fewer) degrees in temperature compared with average.*

| Month | Daily average | Rain effect (1mm) | Temperature effect (1C) |
|---|---|---|---|
| Jan | 33.2 | -0.2 | 1.1* |
| Feb | 33.3 | -0.3 | 0.8 |
| Mar | 36.5 | -0.7 | 1.0* |
| Apr | 40.5 | -0.8 | 1.4* |
| May | 43.6 | -1.6* | 1.6* |
| Jun | 46.7 | -1.6* | 1.8* |
| Jul | 46.4 | -1.3* | 1.5* |
| Aug | 42.6 | -1.0* | 0.9* |
| Sep | 47.7 | -1.1* | -0.1 |
| Oct | 43.8 | -0.5 | 0.4 |
| Nov | 42.0 | -0.3 | 0.6 |
| Dec | 38.4 | 0.2 | 1.3* |

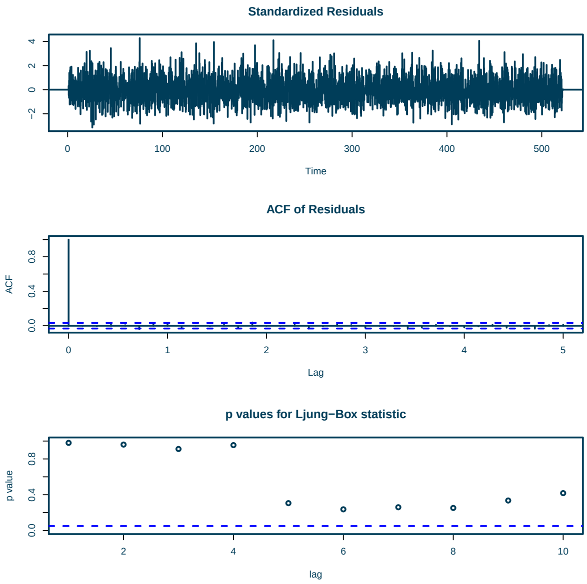
*significant based on t-value of 3.5.

## Future improvements

Some issues identified and future improvements that could be made to the model are as follows.

- Exploring the differences in the results from the auto.arima and arima functions in R as fixing the order of the ARIMA model led to some problems in the diagnostics.
- Using alternative models for the seasonality. Using a trigonometric seasonal could better capture the seasonality as currently it is assumed to be constant within months but is more a gradual change between months. Also some of the seasonal dummy coefficients were not significantly different from the December effect and could be removed. Also some months with similar coefficients could be combined.
- Allowing the weather effects to vary by month. Currently the weather effect is assumed to be constant across the year. In GSS (2014) the analysis found that weather effects varied in size and significance by month. An improvement could be to allow for different effects during different times of the year.

**Figure 9:** *Diagnostic plots for fitted ARIMA model including weather regressors*




- Alternative model for the Christmas effect. All days from Christmas day to the observed New Years day bank holiday were included as outliers; however not all regressors were significant. Alternative models could be considered which better reflect the shape of the time series over the Christmas period.
- Considering alternative methods to ARIMA models, for example state space models.

## Conclusion

This article has presented results of analysis of daily counts of killed or seriously injured vulnerable road users in Great Britain for the effects of weather. It has found some evidence that the counts are related to temperature and rain. Graphical analysis highlighted that the seasonality in KSI vulnerable road users and temperature follow a similar pattern apart from the dip in KSI vulnerable road users in August. This suggests that there could be a relationship, however temperature does not explain all of the seasonality in KSI vulnerable road users.

Comparing the number of KSIVU on different weather type days suggested that there are fewer KSIVU on wet days than dry days, and that colder temperatures leads to fewer KSIVU and warmer temperatures to more. This theory is supported by the results of the ARIMA model which estimates that days with above average rainfall have fewer KSIVU and days with above average temperature have more (and vice versa).

However there are some issues with the ARIMA model and alternative methods could be explored to properly account for all of the features in the daily time series.

## Resources

This section includes the data and R code used to carry out the analysis in this paper as an educational tool for anyone starting to carry out their own weather analysis.

Double click here for R code

Double click here for data

## References

BRSI, 2014, *Are there more accidents in the rain? Exploratory analysis of the influence of weather conditions on the number of road accidents in Belgium*, Macroscopic factors in road safety

DfT, 2015, *Modelling the impact of the weather on road casualty statistics*, Reported road casualties Great Britain: annual report 2014

GSS, 2015, *Exploring the effect of weather and climate on official statistics*