Proceedings of the 21st GSS Methodology Symposium 2016

Methodology: The key to the door of innovation

**Modelling the effect of weather on road casualty statistics**

David Mais, Daryl Lloyd, and Jennifer Davies[[1]](#footnote-1)

**Abstract**

The Department for Transport publish statistics on the number of people killed and injured in road accidents in Great Britain that are reported to the police. It has long been known that the weather impacts on year-on-year changes in these figures. It is important for road safety policy making to be able to assess how much the weather has contributed to these changes.

A cross government group looking at how the weather impacts on different types of statistical series can be assessed was established. This led to the development of a statistical model to produce a weather-adjusted road casualty series. This paper summarises the regARIMA modelling approach that has been used, as well as discussing the adjustments made to the figures in recent years. There was an increase in road deaths between 2010 and 2011, but when the weather is accounted for this increase no longer exists.

Key Words: weather; road; safety.

**1. Introduction**

**1.1 Motivation for producing a weather adjusted series**

The Department for Transport first became aware that the weather affects road casualties when the 2011 figures were published. There was an increase in road deaths in 2011, the first for a number of years (figure 1.1-1). It became clear that this increase was likely to have arisen due to the significant and extended snowfall throughout Britain at the start and end of 2010. This heavy snowfall is likely to have suppressed the number of users on the road leading to a reduction in traffic, accidents and casualties in these periods. There is therefore likely to have been fewer road deaths in 2010 than would have been expected had the conditions been average. As a result an increase in fatalities was seen in 2011. The Department wanted to establish a method to quantify the effect that weather has had on road casualties and estimate the likely number of road casualties that would have occurred each year had weather conditions been more like the long term average.

**Figure 1.1-1**

**Fatalities in reported road accidents: GB, 2005-2014**



**1.2 Cross government working**

The Department for Transport formed part of a cross-government group looking at how the weather impacts on different types of statistical series could be assessed. As part of this work a case study on the Department’s road accident data was undertaken (Davies and Elliot, 2015). RegARIMA modelling was used to test and estimate the effect of deviations from average monthly temperatures on the number of killed or seriously injured vulnerable road users (pedestrians, pedal cyclists and motorcyclists). The final model found significant positive effects (i.e. above average temperatures leading to increases, below average temperatures leading to decreases) for all months of the year except for August, September and November. The Department extended this approach to produce the weather-adjusted series outlined in this paper.

**2. Producing the weather-adjusted series**

**2.1 RegARIMA modelling**

The term “regARIMA” is used to describe the combination of a regression model and an autoregressive integrated moving average (ARIMA) model (Davies and Elliot, 2015). ARIMA models account for correlation in the observed variables and errors, and are typically used to clean time series data prior to seasonal adjustment. The regARIMA model can be thought of as a linear regression model with ARIMA errors. In the equation:

$$y\_{t}= \sum\_{i=1}^{m}β\_{i}x\_{it}+ z\_{t}$$

$y\_{t}$ is the time series of observations to be modelled with observations at time points $t=1, …, T$. $x\_{it}$ are the $i=1, …, m$ regression variables. These might include calendar effects and weather or climate variables. $β\_{i }$are the regression coefficients to be estimated in the model. $z\_{t}$ is an ARIMA process, which deals with certain time series characteristics that could lead to poor inference in our regression model if not appropriately accounted for. In this case, assume we have a monthly time series of road casualties ($y\_{t}$) and a variable ($x\_{it}$) that measures the difference between the observed average temperature in month t and the long term average temperature for that month (for example the monthly averages averaged over 30 years). The estimated coefficient ($\hat{β}\_{i}$) can be interpreted as the amount you would expect $y\_{t}$ to change if the average temperature in any month is one degree above the long term average, holding everything else constant.

ARIMA models assume that time series are stationary. The integrated part of ARIMA modelling can help to make a time series stationary by removing trends and seasonality. However, sometimes a transformation of the time series is also required. A common transformation for many time series models is the logarithmic scale. This is used to stabilise the variance of the time series where, for example, the variance of the observations increases as the level of the series increases. This alters the interpretation of the estimated coefficients ($\hat{β}\_{i}$). Using the previous example and assuming a log transformation of the variable of interest, the estimated coefficient is an estimate of the per cent change of the series of interest ($y\_{t}$) for an average temperature in a month that is one degree above the long term average.

**2.2 Data**

The dataset used in this analysis contains records of all police recorded road accidents in Great Britain. The dataset has information on the time and location of the accident as well as information on casualties, the types of road users affected and the severity of their injuries and a number of other variables, including some information on road and weather conditions. Data are available from January 1979 to December 2015.

For this analysis the records were aggregated to produce monthly time series over 1979 to 2014. The monthly time series were the number of killed, seriously injured and slightly injured casualties for the following road users: pedestrians, pedal cyclists, motorcyclists and car occupants in Great Britain. This resulted in twelve regARIMA models, i.e. one for each road user type in each of the three casualty severities (killed, seriously injured and slightly injured).

The weather data selected for this analysis was mean monthly UK temperature and precipitation from the Met Office. Note the weather series is for the UK and the transport data are for Great Britain. Weather data were not readily available for Great Britain so it was decided to use the UK series as the difference in monthly averages was not expected to be large. An alternative would have been to construct weather data for Great Britain using available weather data for sub-regions of the UK.

**2.3 The models**

Twelve regARIMA models (one for each road user type and severity class) were used to estimate the effect of temperature and precipitation on the number of road casualties in each month:

$$y\_{t}= \sum\_{i=1}^{12}β\_{i}^{temp}x\_{it}^{temp}+\sum\_{i=1}^{12}β\_{i}^{prec}x\_{it}^{prec}+ z\_{t}$$

$y\_{t}$ is monthly casualty data for Great Britain over 1991 to 2014

$x\_{it}^{temp} $and $x\_{it}^{temp}$ are monthly deviations in each month from the long term average UK temperature and precipitation

$β\_{i}^{temp}$ and $β\_{i}^{prec}$ are the temperature and precipitation effects on road casualties in each month to be estimated

Each model was run initially with all regressors included with the order of each ARIMA model selected using X-13ARIMA-SEATS. Backwards selection was then used to select a model with a set of regressors which were all statistically significant. Backwards selection is an iterative process where of all of the regressors which are not found to be statistically significant, the least significant is removed. The model is then refitted and the process is run again until all remaining regressors are statistically significant. The resulting regressors in the final models and their estimated coefficients are discussed in section 2.4 below.

The temperature and precipitation effects have been modelled as linear. This means that the temperature and precipitation effect increases linearly with the temperature and precipitation deviation. For example, it is estimated that a March in which temperature is 1 °C higher than average leads to 38 more seriously injured motorcyclists than if temperature is average (see below), so if March is 2 °C higher than the average temperature for March, then all else being equal, there will be an additional 76 seriously injured motorcyclists. However, for extreme deviations the assumption of linearity may not hold e.g. it is unlikely that if a March was 6 °C above average that this would lead to 228 more seriously injured motorcyclists (6 x 38). However, extreme deviations are rare so the assumption of linearity is likely to be reasonable e.g. the majority of temperature deviations are between -2 °C to 2 °C from the average (see figure 2.3-1). Unlike temperature, precipitation cannot take negative values so the maximum deviation of precipitation below average in any month is capped at the long run average precipitation for that month (which would be achieved if precipitation in the month was 0 mm).

**Figure 2.3-1**

**Time series of temperature deviations from long-run monthly average between January 1979 and July 2015**



**2.4 Interpretation of the model coefficients**

The estimated temperature and precipitation effects for each road user type (pedestrians, pedal cyclists, motorcyclists and car occupants) are summarised in DfT, 2015. As an example, an (0,1,1)(0,1,1)12 regARIMA model with log transformation was selected for killed pedal cyclists with the following estimated temperature and precipitation effects in the final model:

**Figure 2.4-1**

**Estimated temperature and precipitation effects from the pedal cyclist fatalities regARIMA model**



The model suggests that as temperature increases in January, March and December, the number of killed pedal cyclists increases. The largest effect is found for the January regressor with an estimated coefficient of 0.117 which implies that if the temperature in January is 1 °C higher than average this will increase the number of killed pedal cyclists in January by about 11.7%. The model suggests that as precipitation increases in April, the number of killed pedal cyclists decreases. According to the model, if precipitation in April is 1 mm above average then there will be a decrease in the number of killed pedal cyclists by about 0.7%. The direction of the estimated temperature and precipitation effects from the final regARIMA models for each road user type (pedestrians, pedal cyclists, motorcyclists and car occupants) are summarised in figure 2.4-2 below for each severity.

**Figure 2.4-2**

**Direction of the estimated temperature and precipitation effects on road casualties by severity and road user type**

The estimated temperature and precipitation effects were used to produce weather-adjusted road casualty figures in each month. For instance, it is estimated that a March in which temperature is 1 °C higher than average leads to 38 more seriously injured motorcyclists. Therefore, if a March has temperature that is 1 °C higher than average then the adjusted seriously injured figure is found by subtracting 38 from the actual casualty figure. The adjusted figure can then be interpreted as the number of seriously injured motorcyclists we would have expected had the temperature been at the long term average (no statistically significant precipitation effect was found in March – figure 2.4-2).

**3. Analysing the weather-adjusted series**

Figure 3.1-1 shows the actual and weather-adjusted number of road deaths in Great Britain. The weather-adjusted series represents the number of road deaths that would have been expected each year had the temperature and precipitation in each month been at the long term average.

The largest weather adjustments for road deaths in recent years have been in 2010, 2011 and 2014. In 2010 the temperature was well below average in the months of January, February, November and December. In particular, December 2010 was the coldest on record. Owing mostly to the colder temperatures in 2010, we estimate that there were around 20 fewer road deaths than would have been expected if temperature and precipitation had been at average levels. As a result, the weather-adjusted fatality series predicts that, had the temperature and precipitation been average for the year, there would have been roughly 1,870 road deaths in 2010 rather than the 1,850 actually observed.

In contrast, 2011 was a much warmer year, with above average temperatures in most months. In particular, April 2011 was the warmest April on record. Overall, we estimate that the warmer temperature in 2011 may have led to about 20 more road deaths than would have been expected if the temperature had been average. Across the whole year, precipitation was close to average. However, owing mostly to a dry March, April and November, we estimate that there were between 10 and 15 more road deaths in 2011 than would be expected if precipitation had been at average levels. Taken as a whole, therefore, the weather-adjusted fatality figure for 2011 is about 1,867, 34 fewer deaths than the 1,901 that were recorded.

After seven consecutive years of decline, the rise of 51 recorded road deaths between 2010 and 2011 (just under 3%) was a surprise. Once adjusted for the unusual weather patterns in both years, though, this rise is not only lost, it becomes a fall (albeit, statistically insignificant) of about four deaths year-on-year (Figure 3.1-1).

Similar to 2011, 2014 resulted in a rise in recorded road deaths. There were a total of 1,775 fatalities in 2014, an increase of 4% compared with 2013. However, 2014 was the warmest year on record as well as the fourth wettest year on record. Overall, we estimate that there were around 45 more road deaths in 2014 than would have been expected had the weather been closer to the long-term average. But after adjusting for the weather, the year-on-year increase in fatalities is closer to 1% than 4%.

**Figure 3.1-1**

**Actual and weather-adjusted fatalities in reported road accidents: GB, 2010-2014**



**References**

Davies and Elliot (2015), Analysing the impact of weather and climate on official statistics time series. Available online: <http://www.ons.gov.uk/ons/guide-method/method-quality/survey-methodology-bulletin/smb-content-4.pdf>

DfT (2015), Modelling the impact of weather on road casualty statistics:

<https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/463100/weather-on-road-casualties.pdf>

1. David Mais, Department for Transport, David.Mais@dft.gsi.gov.uk; Daryl Lloyd, Department for Transport, Daryl.Lloyd@dft.gsi.gov.uk; Jennifer Davies, ONS, jennifer.davies@ons.gov.uk [↑](#footnote-ref-1)