# IMPACT OF METEOROLOGICAL FACTORS ON THE NUMBER OF INJURY ACCIDENTS

*C. Antoniou, National Technical University of Athens, antoniou@central.ntua.gr*

*G. Yannis, National Technical University of Athens, geyannis@central.ntua.gr*

*D. Katsochis, PLANET S.A., dkats@planet.gr*

## ABSTRACT

Weather conditions are generally considered as one of the most critical sets of factors affecting the number of road accidents and associated casualties in a given network. There have been a few most interesting papers during the last decade dealing with the impact of such parameters on the number of accidents and casualties, through the development of state-space models at the national level, as well as in other forms.
This work is aimed at the exploration of two essential meteorological indicators' (temperature and precipitation) impact on the number of total accidents and fatalities recorded in the wider Athens area for a period of 9 years. The temporal correlation between accidents and fatalities, as well as meteorological variables, is examined. Generalized linear models (GLM) –a family of models including the negative binomial, Poisson and quasi-Poisson distributional assumptions– as well as dynamic GLM (or state-space) models are employed to present model diagnostics and goodness-of-fit measures.
The estimation and prediction accuracy of the various types of models is compared. The examined GLM models show similar estimation and prediction performance. Other aspects of the models are also considered, including models that might give false positive indication of the significance of some values. Furthermore, dynamic GLM/state-space models show a considerably improved performance over the GLM models. The presented models demonstrate a reasonable differentiation across months within a year. Conclusions and recommendations for the practical use of these results in shaping public policy and strengthening road safety campaigns are also discussed.

*Keywords: weather conditions, number of accidents, generalized linear models (GLM), state-space models, meteorological variables, precipitation, temperature, monthly variation*

# INTRODUCTION AND BACKGROUND

## Introduction

Weather conditions are generally considered as one of the most critical sets of factors affecting the number of road accidents and associated casualties in a given network. It is generally accepted that, in the road safety context, they exhibit varying effects on the type of accidents and severity of casualties, primarily depending on the combination of road type (e.g. motorways vs. rural or urban/suburban roads with geometrical characteristics of lower category) and season of the year (inherent property, directly determining the range of the average low/high temperature, rainfall and so on). Beyond the impact of each specific combination of those two factors on the overall environment formed, it has been systematically observed that weather typically affects mobility. It is therefore expected that the effects of weather on the number of injury accidents and casualties are partly differing due to the changes in mobility occurring at the same time.

This work is based on time series accident analyses conducted in Work Package 7 on "Data analysis and synthesis" of the EU-FP6 project "SafetyNet – Building the European Road Safety Observatory". It is aimed at the exploration of two essential meteorological indicators' (temperature and precipitation) impact on the number of total accidents and fatalities recorded in the wider Athens area for a period of 9 years, in the general context of other analyses already carried out and presented on the use of weather variables for analysing changes in the number of road injury accidents (Stipdonk (Ed.), 2008). The suitability of several distributions for modelling road safety data that are temporally correlated is investigated. More precisely, the correlation between accidents and fatalities, and meteorological variables is examined. Exposure data (in the form of vehicles recorded in a regional toll station) is also included in the analysis.

Two families of statistical models are considered in this analysis, namely generalized linear models (GLM) (McCullagh and Nelder, 1989; Dobson, 1990; Gill, 2000) –a family of models including the negative binomial, Poisson and quasi-Poisson distributional assumptions– and dynamic GLM or state-space models. Model diagnostics and goodness-of-fit measures are presented and the explanatory and predictive power of the more involved dynamic GLM models is demonstrated. Different approaches provide insight into the modelled processes and demonstrate their use.

In the concluding section, it is discussed that better understanding of the subtle difference among different model functional forms can make a difference in providing more reliable forecasts. Models that can accurately assess the impact of meteorological parameters on traffic safety can be useful in establishing base-line conditions, in order to assess the performance of safety measures and campaigns.

## Background

### *The research problem*

Time series analysis techniques are widely used to analyse changes in road safety trends as observed at national level, with the use of meteorological variables employed to capture short-term changes in road safety indicators. Fixing the time scale for the analysis constitutes a most important aspect, as the effects of weather conditions on safety indicators may be expected to differ according to the time scale, for example depending on whether the day or the month is selected as the reference base. Daily monitoring generally requires a large number of accident data counts that often reveal notable fluctuation. It is therefore used relatively less to perform analyses at national level. Monthly monitoring tends to be preferred instead, as it facilitates international benchmarking. In France, for instance, short-term monitoring is performed on monthly sets of accident data (ONISR, 2012). This issue is investigated in this paper.

According to existing literature, weather may explain approx. 5% of monthly accident/fatality variability (see Fridstrom et al., 1995, Hermans et al., 2006). Therefore, including the influence of weather conditions in the analysis of road accident trends at aggregate level proves to be especially appropriate for short-term analysis –whether for the past or the near future– as it provides a short-term trend which facilitates identification of the effects of applied safety policy.

There have been a few most interesting recent papers dealing with the impact of the essential meteorological parameters on the number and type of accidents, as well as on the number and severity of casualties. In this context, Hermans et al. (2006) developed state-space models at country level for Belgium for the period 1974-1999. Distinct dependent variables were formed for killed and seriously injured (KSI) and slight injuries. Predictors included factors from 3 main groups: juristic, weather and economic. It was concluded that the quantity of precipitation (in mm, measured as an average for the whole country) was only associated to an increase in the number of accidents and casualties of moderate severity. Eisenberg (2004) also noticed that under adverse weather conditions drivers possibly adopt lower speed resulting, on average, to less severe accidents.

Interestingly, this is similar to respective findings on the impact of traffic congestion in urban and/or suburban areas, where increased density leads in general to lower speed, again resulting to less severe accidents.

### *Current state of knowledge*

The influence of temperature and rainfall on the aggregate numbers of injury accidents and casualties were first studied by Scott (1986) who modelled the changes in their monthly number in the UK from 1970 to 1978. Since then, numerous attempts have been made to build an explanatory model to account for the influence of climate using these two variables independently or simultaneously. The results of these attempts vary, depending on the adopted time scale of the analysis (the day or the month) and the construction of the selected weather variables (the mean values or the extreme values for the considered time scale). In addition, in models where the meteorological parameters prove to be increasingly significant among all selected variables, the magnitude and sometimes even the sign of weather variables' impact may differ.

With respect to the methodology adopted to construct weather variables in the frame of a given analysis, meteorological data including both average and atypical values during the month was first used by Fridstrom and Ingebrigtsen (1991) and Fridstrom et al., (1995) who analysed the changes in the monthly number of accidents of four countries with similar characteristics (Denmark, Finland, Norway and Sweden,). Monthly aggregate models exploiting weather data including both average and atypical values during the month were also fitted to accident data for France, with disaggregation according to the type of road (Bergel and Depire, 2004a and 2004b).

The relationship between weather conditions and road accidents has been investigated intensively in recent years. Respective research results that focus on very short-term links (daily level) and on short-term links (monthly level) have been published (Karlaftis and Yannis, 2010).

# DATA AND METHODOLOGY

## Data used

### *Daily Data*

Daily average temperature and total precipitation are available at daily level for the wider Athens area from the National Observatory of Athens (NOA) and the National Statistical Service of Greece (NSSG). The dataset that is used in this analysis covers the period from the beginning of 1997 until the end of year 2005. Data have undergone some processing from the Department of Transportation Planning and Engineering (School of Civil Engineering) of the National Technical University of Athens (NTUA).

### *Monthly Data*

Furthermore, in an attempt to also consider exposure data, monthly traffic data have been used for the same period from January 1997 to December 2005 (108 months). The values that have been used refer to the Shimatari toll station to the north of Athens, and are assumed as a reasonable proxy to the entire traffic in the Athens area. Information on the traffic composition (passenger cars etc.) is also available and is exploited in this analysis.

Figure 1 presents some key variables of the aggregated monthly data. For the aggregation of the precipitation, a total precipitation for the month was computed. For the aggregation of the temperature, several variables were computed: month average mean temperature, minimum mean temperature for the month, and maximum mean temperature for the month (where mean refers to daily mean).

Several observations can be made. Unusually high precipitation was mainly recorded during winter 2002-03, following two relatively dry winter periods. Data reveal significant improvement (reduction) of fatalities and total accidents. It is stressed, though, that this reduction practically appears as a step in the transition from 2001 to 2002. On the other hand, a steady increase of passenger car traffic is recorded at the annual level.

The proportion of passenger cars on total vehicles travelling through the aforementioned toll station appears to increase over time. In particular, cars amounted up to approximately 75% of all vehicles for the first three years (1997-99). This proportion increased to 80% in 2000, 85% for the period 2001-2004 and about 90% in the second half of year 2005.

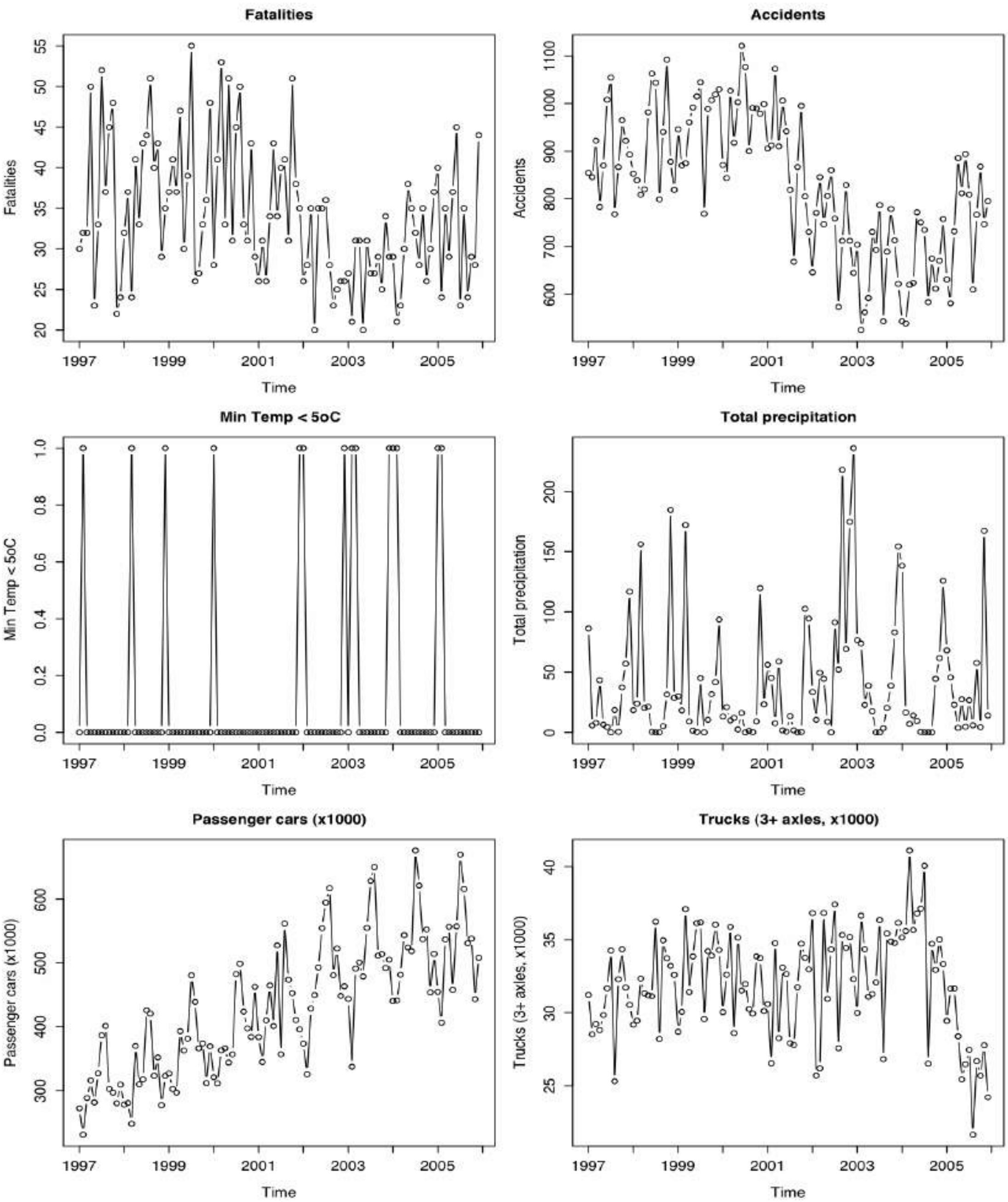



Figure 1 – Descriptive graphs of the monthly dataset used for the analysis

## Methodology

Two families of statistical models have been considered in this analysis to utilise available data, namely generalized linear models (GLM) –a family of models including the negative binomial, Poisson and quasi-Poisson distributional assumptions– and dynamic GLM (or state-space) models. In each case, model diagnostics and goodness-of-fit measures are presented, whereas the explanatory and predictive power of the more involved dynamic GLM models is demonstrated to validate initial assumptions and allow for reliable forecasts. To this end, the selected data-set is split into two parts; the first part is used to fit the models and test their estimation performance, while the second part is used to validate the predictive performance of the models.

## *Generalized Linear Models*

Generalized linear models facilitate the analysis of the effects of explanatory variables in a way that closely resembles the analysis of covariates in a standard linear model, but with less confining assumptions. This is achieved by specifying a *link function*, which links the systematic component of the linear model with a wider class of outcome variables and residual forms.

A key point in the development of GLM was the generalization of the normal distribution (on which the linear regression model relies) to the exponential family of distributions. This idea was developed by Fisher (1934). Consider a single random variable y whose probability (mass) function (if it is discrete) or probability density function (if it is continuous) depends on a single parameter θ. The distribution belongs to the exponential family if it can be written in the form (Eq. (1)):

$$f(y;\theta) = s(y)t(\theta)e^{a(y)b(\theta)} \tag{1}$$

where a, b, s, and t are known functions. The symmetry between y and θ becomes more evident if Eq. (1) is rewritten as Eq. (2):

$$f(y;\theta) = \exp\left[a(y)b(\theta) + c(\theta) + d(y)\right] \tag{2}$$

where s(y)=exp[d(y)] and t(θ)=exp[c(θ)]. If a(y)=y then the distribution is said to be in the canonical form. Furthermore, any additional parameters (besides the parameter of interest θ) are regarded as nuisance parameters forming parts of the functions a, b, c, and d, and they are treated as though they were known. Many well-known distributions belong to the exponential family, including –for example– the Poisson, normal, and binomial distributions. On the other hand, examples of well-known and widely used distributions that cannot be expressed in this form are the student's t-distribution and the uniform distribution.

The generalized linear model can be defined in terms of a set of independent random variables $y_1, \ldots, y_n$, each with a distribution from the exponential family with the following properties:

1. The distribution of each $y_i$ is of the canonical form and depends on a single parameter $\theta_i$ (not necessarily the same parameter for all variables) (Eq. (3)):

$$f(y_i;\theta_i) = \exp\left[y_i b_i(\theta_i) + c_i(\theta_i) + d_i(y_i)\right] \tag{3}$$

2. The distributions of all the $y_i$ s are of the same form (e.g. all normal or all binomial) so that the subscripts on b, c, and d are not needed.

The joint probability density function of $y_1, \ldots, y_n$ is then (Eq. (4)):

$$f(y_1,\ldots,y_N;\theta_1,\ldots,\theta_N) = \exp\left[\sum_{i=1}^{N}\left(y_i b(\theta_i) + c(\theta_i) + d(y_i)\right)\right] \tag{4}$$

When specifying a model, the *N* parameters $\theta_I$ are usually not of direct interest. Instead, for a GLM, a smaller set of *p* parameters $\beta_1, \ldots, \beta_p$ is considered (where $p < N$), such that a linear combination of the βs is equal to some function of the expected value $\mu_i$ of $y_i$, i.e. (Eq. (5)):

$$g(\mu_i) = \mathbf{x_i}^{\mathbf{T}}\boldsymbol{\beta} \tag{5}$$

where
*g* is a monotonic, differentiable function called the link function;
$\boldsymbol{x}_i$ is a $(p \times 1)$ vector of explanatory variables (covariates and dummy variables for levels of factors); and
$\boldsymbol{\beta} = [\beta_1, \ldots, \beta_p]^T$ is the $(p \times 1)$ vector of parameters.

To recapitulate, in the univariate case, a generalized linear model has three components:

1. A response variable y assumed to follow a distribution from the exponential family (which is a generalization of and includes the normal distribution);

2. A set of parameters $\boldsymbol{\beta}$ and explanatory variables $\mathbf{X} = [\mathbf{x_1}^{\mathbf{T}}, \ldots, \mathbf{x_N}^{\mathbf{T}}]^{\mathbf{T}}$

3. A monotonic link function g such that $g(\mu_i) = \mathbf{x_i}^{\mathbf{T}}\boldsymbol{\beta}$, where $\mu_I = E(y_i)$

## *Dynamic GLM/State Space models*

The DGLM models are a certain form of SS models and were run at this context using the Poisson distribution and log link for the dependent variable (using the sspir package in R) and as a GLM (again using the Poisson distribution with a log link function). This is considered a standard, acceptable approach according to international literature due to the suitability of the Poisson distribution to describe counts data (Lord et al., 2005). In this section Gaussian state space models are first presented and following the presentation in Dethlesen and Lundbye-Christensen (2006) it is shown how generalized linear models can naturally be extended to allow the parameters to evolve over time. Components (e.g. trend and seasonal components) that separate the time series into parts that may be inspected individually after analysis are defined.

The Gaussian state space model for univariate observations involves two processes, the state (or latent) process, $\theta_k$, and the observation (or measurement) process, $y_k$.

Let $\{y_k\}$ be measured at regular intervals $t_k$ for k = 1, . . . , n. The state space model can then be defined as

$$y_k = \boldsymbol{F}_k^T\boldsymbol{\theta}_k + v_k, \;\; v_k \sim N(V_k) \tag{6}$$
$$\boldsymbol{\theta}_k = \boldsymbol{G}_k\,\boldsymbol{\theta}_{k-1} + \boldsymbol{\omega}_k, \;\; \boldsymbol{\omega}_k \sim N_p(0, \boldsymbol{W}_k) \tag{7}$$
$$\boldsymbol{\theta}_0 \sim N_p(\boldsymbol{m}_0, \boldsymbol{C}_0) \tag{8}$$

where $v_k$ and $\omega_k$ are white error noise vectors of appropriate dimension.

Assuming that the state process is Gaussian and the sampling distribution belongs to the exponential family:

$$p(y_k|\eta_k) = \exp\{y_k|\eta_k - b_k(\eta_k) + c_k(y_k)\} \tag{9}$$

which contains several distributions (such as the Gaussian, Poisson, gamma and the binomial) as special cases. The natural parameter $\eta_k$ is related to the linear predictor $\lambda_k$ by the equation $\eta_k = v(\lambda_k)$ or equivalently $\lambda_k = u(\eta_k)$. The linear predictor in a generalized linear model is of the form $\lambda_k = Z_k \beta$, where $Z_k$ is a row vector of explanatory variables and $\beta$ is the vector of regression parameters. The link function, g, relates the mean, $E(y_k) = \mu_k$ , and the linear predictor, $\lambda_k$: $g(\mu_k) = \lambda_k$.

The static generalized linear model can be extended by adding a dynamic term, $X_k \beta_k$, to the linear predictor, where $\beta_k$ is varying randomly over time according to a first order Markov process, resulting in

$$\lambda_k = \boldsymbol{Z}_k\boldsymbol{\beta} + \boldsymbol{X}_k\boldsymbol{\beta}_k \tag{10}$$

where $\beta$ is the coefficient of the static component and $\beta_k$ are the time-varying coefficients of the dynamic component. The model is fully specified by the initializing parameters $m_0$ and $C_0$, the matrices $F_k$, $G_k$, and the variance parameters $V_k$ and $Var(\omega_k)$.

The variation in the linear predictor, random or not, may be decomposed into four components: (i) a time trend, (ii) harmonic seasonal patterns, (iii) unstructured seasonal patterns, and (iv) a regression with possibly time-varying covariates. Each component may contain static and/or dynamic components, which is specified by zero and non-zero diagonal elements in $Var(\omega_k)$, respectively. The long term trend is usually modelled by a sufficiently smooth function. In static regression models, this can be done by e.g. a high degree polynomial, a spline, or a generalized additive model. In the dynamic setting, however, a low degree polynomial with time-varying coefficient may suffice.

# MODEL ESTIMATION RESULTS AND FINDINGS

Several models have been developed using the data described above: a dynamic GLM (DGLM)/state space (SS) model, using the Poisson distribution and log link for the dependent variable (using the SSPIR package in R), and as a GLM (again Poisson with log link, but also quasi-Poisson due to the known issues with the standard errors due to not modelling overdispersion by Poisson) (Yannis et al., 2007).

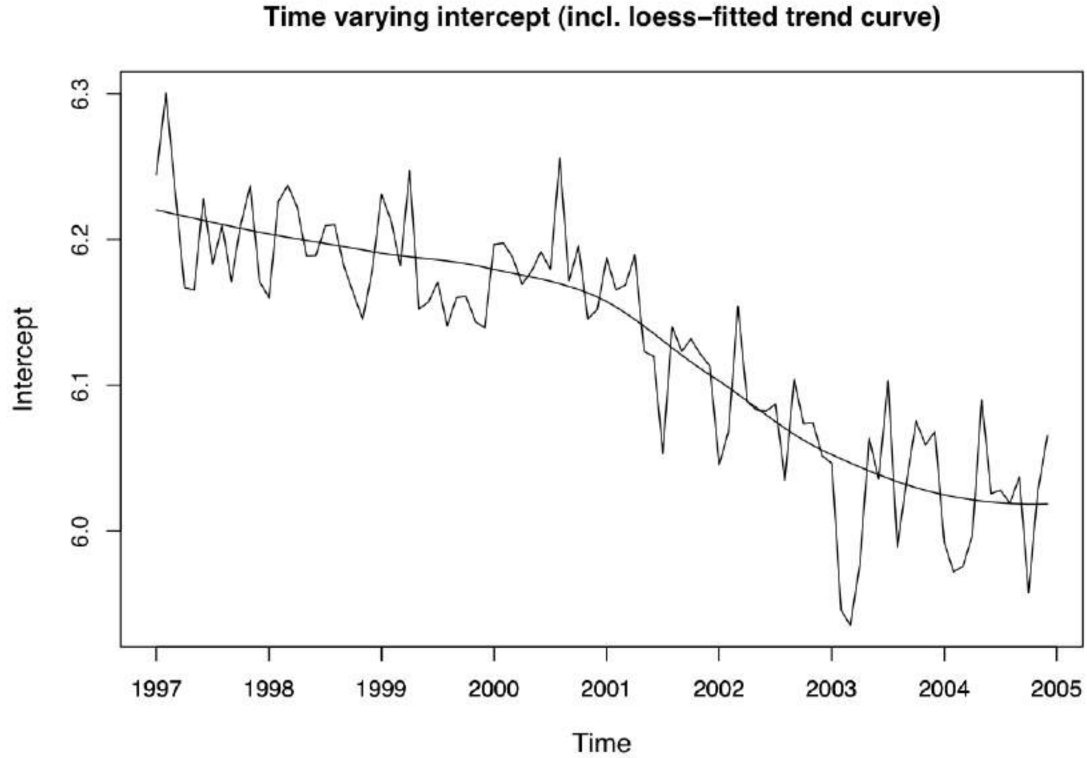
In this section, the approach of dynamic generalized linear models (DGLM), which allows for –among other things- explicitly modelling serial correlation, is used instead of the static GLM framework. While state-space models (SS) and the traditional Kalman Filter techniques have normality requirements, as mentioned in the previous section, this family of models allows distributions of the measurement equation that fall within the exponential family (Dethlefsen and & Lundbye-Christensen, 2006).

A large number of explanatory variables, in different forms, have been considered. In the final models, the following explanatory variables have been used:

- MinMeanTempLT5: binary (0/1) variable taking the value 1 (or TRUE) if the minimum mean temperature of at least one day in the month was less than 5 degrees C.

- SumTotalPrecipitation: sum of total precipitation in month (mm)
- HeavyTrucks: number of heavy trucks (three axle or more) passing from the toll station in the entire month
- Motor2Wheelers: number of motorized two-wheelers (mopeds and motorcycles)

While the two types of models DGLM and SS have many similarities, there are also some distinct differences. In particular, the intercept is allowed to vary in the SS model, while there is also an unstructured seasonal pattern (as described in Claus Dethlefsen and Soren Lundbye-Christensen, 2006). The differences between the two models are therefore the time-varying intercept in the state-space model and the seasonal pattern. Figure 2 illustrates the time varying intercept (top subfigure), in which a loess-fitted trend line is also fitted, for illustration purposes of the decreasing trend. The bottom subfigure of Figure 2 illustrates the month-specific seasonal components, verifying that there is a lower number of accidents during the summer months (possibly due to low exposure and improved weather conditions), while there is a higher number of accidents during the winter months (possibly due to inclement weather conditions). The lowest seasonal component is observed in August, while the highest one in December.



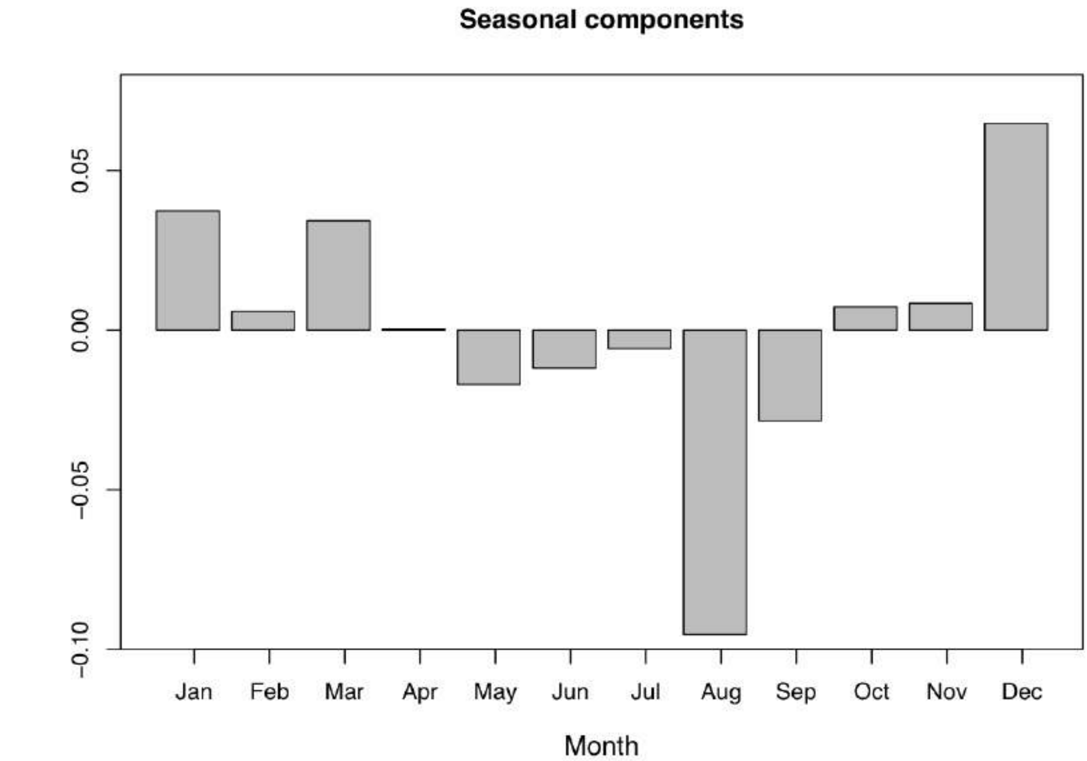




Figure 2 – Selected estimated parameters (state-space, accidents)

Table 1 summarizes the model estimation results for the GLM models. In the top row, the models with the Poisson assumption are presented, followed by the models with the quasi-

Poisson assumption in the bottom row. In each case, a model without dummies for the months is presented first, followed by a model with month dummies (to somewhat approximate the seasonal components of the state-space model). While the coefficients for most of the retained parameters are significant at the 95% level, the significance of some falls well below this threshold.

Table 1 – Model estimation results for GLM models (top: Poisson models, bottom: quasi-Poisson models)

| | **GLM Poisson (glm1.p)** | | **GLM Poisson with month dummies (glm2.p)** | |
|---|---|---|---|---|
| | **Coef.** | **t-test** | **Coef.** | **t-test** |
| Intercept | 6.119 | 155.165 | 6.057 | 147.94 |
| Min Temp < 5oC | 4.851x10-2 | 3.549 | 1.934x10-2 | 1.322 |
| Total Precip. | 1.602x10-4 | 2.14 | 1.118x10-4 | 1.324 |
| Heavy Trucks | -9,891x10-3 | -8.387 | -8.263x10-3 | -6.810 |
| Motor.2wheel. | 1.863x10-3 | 42.163 | 1.882x10-3 | 41.787 |
| | | | | |
| Dummy Jan | N/A | | 7.207x10-2 | 5.079 |
| Dummy Aug | N/A | | -2.111x10-3 | -2.068 |
| Dummy Dec | N/A | | 6.198x10-2 | 3.921 |
| | | | | |
| Null deviance | 2669.6 | (95 d.o.f.) | 2669.6 | (95 d.o.f.) |
| Residual deviance | 391.54 | (91 d.o.f.) | 352.0 | (88 d.o.f.) |
| AIC | 1222.3 | | 1188.8 | |
| | **GLM quasipoisson (glm1a.p.disp)** | | **GLM quasipoisson with month dummies (glm2.p.disp)** | |
| | **Coef.** | **t-test** | **Coef.** | **t-test** |
| Intercept | 6.107 | 72.877 | 6.043 | 72.072 |
| Min Temp < 5oC | 5.316x10-2 | 1.932 | 2.079x10-2 | 0.718 |
| Total Precip. | --- | --- | 1.251x10-4 | 0.730 |
| Heavy Trucks | -9.647x10-3 | -3.982 | -8.507x10-3 | -3.468 |
| Motor.2wheel. | 1.83x10-3 | 20.341 | 1.925x10-3 | 20.611 |
| | | | | |
| Dummy Jan | N/A | | 7.013x10-2 | 2.430 |
| Dummy Aug | N/A | | -2.360x10-3 | -1.098 |
| Dummy Dec | N/A | | 6.460x10-2 | 1.959 |
| | | | | |
| Null deviance | 605.372 | (95 d.o.f.) | 649.594 | (95 d.o.f.) |
| Residual deviance | 91.335 | (92 d.o.f.) | 86.958 | (88 d.o.f.) |
| AIC | 280.26 | | 296.69 | |

The estimation and prediction accuracy of the various types of models is compared. While the various GLM models show similar estimation and prediction performance, other aspects of the models are also considered, as –for example– models that do not model over-dispersion correctly underestimate standard errors and might give false positive indication of the significance of some values. Furthermore, dynamic GLM/state-space models show a considerably improved performance over the GLM models.

The presented models demonstrate a reasonable differentiation across months within a year, with June yielding more accidents than each month of the autumn period, probably because more vehicle-km are driven on most road networks during early summer. Intuitive expectations seem to be justified to some extent from the use of models that also utilise exposure data. In particular, it appears that low temperature during wintertime, mostly, corresponds to some reduction of recorded accidents. The same is the case as total precipitation in a month increases, probably due to reduced mobility under rainy weather, but this effect is much less pronounced.

Table 2 summarizes the estimation and prediction performance of the various estimated models, as they have been quantified using the RMSPE metric:

$$\mathrm{RMSPE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(\frac{u_n^e - u_n^o}{u_n^o}\right)^2} \tag{11}$$

where *N* is the number of observations and superscripts *e* and *o* denote estimated (respectively predicted) and observed fatalities. All models perform satisfactorily, with an estimation error (as expressed by the RMSPE metric) of about 7% for the GLM models and less than 4% for the state-space model. The predictive performance is naturally a bit deteriorated, but still the error is below 10% for the GLMs and well below 5% for the state-space model.

It is noted that the GLM models that use the Poisson distribution do not model the overdispersion in the data correctly (c.f. Yannis et al., 2007) and therefore may underestimate standard errors. However, from an estimation and prediction point of view there is no clear reduction in predictive capabilities, when compared to the more appropriate quasi-Poisson assumption.

Table 2 – Estimation and prediction results for the various models

| **RMSPE** | **Estimation** | **Prediction** |
|---|---|---|
| State Space model | 0.0386 | 0.0461 |
| GLM Poisson (glm1.p) | 0.0727 | 0.0984 |
| GLM Poisson – month dummies (glm2.p) | 0.0684 | 0.0914 |
| GLM quasi-Poisson (glm1a.p.disp) | 0.0730 | 0.0981 |
| GLM quasi-Poisson – month dummies (glm2.p.disp) | 0.0685 | 0.0948 |

Figure 3 provides a visual representation of the estimation and prediction accuracy of the presented models. Naturally, visual inspection confirms the quantitative results presented in Table 2.

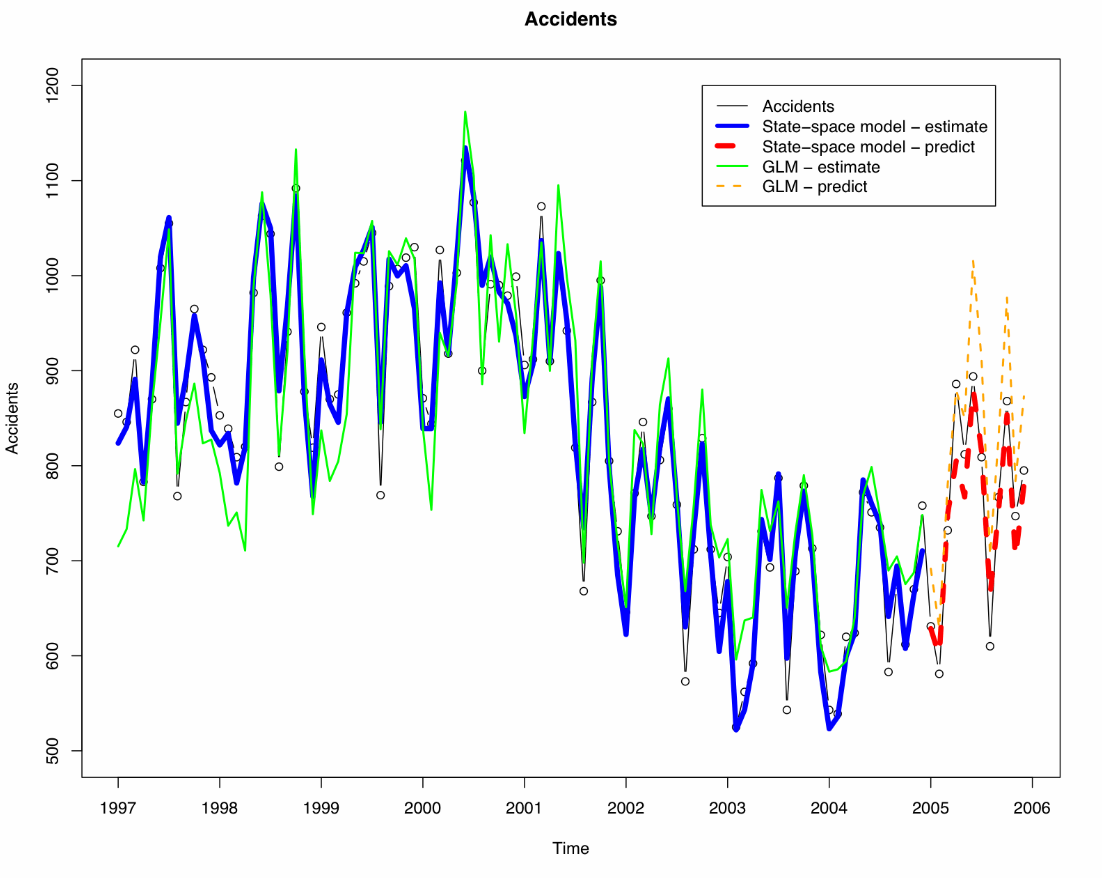



Figure 3 – Estimation and prediction results of GLM and DGLM/state-space models

# IMPLICATIONS FOR RESEARCH/POLICY

This work concludes with certain conclusions and recommendations for the practical use of the results presented above in shaping public policy and strengthening road safety campaigns. Better understanding of the subtle difference among different model functional forms can make a difference in providing more reliable forecasts.

Models that can accurately assess the impact of meteorological parameters on traffic safety can be useful in establishing base-line conditions, in order to assess the performance of safety measures and campaigns. However, one observation that can be made from the research presented in this paper is that (as perhaps expected) one cannot really separate exposure from weather impacts.

In the previous sections of this work, certain attempts to investigate the impact of meteorological variables on trends of road traffic accidents and fatalities have been presented. Results may be considered encouraging, as there are consistent trends regarding the impact of temperature and precipitation on the examined dependent variables.

As expected, models solely built around meteorological variables only demonstrate limited potential in interpreting trends and may only be used as indicative descriptive tools. These rather simple models demonstrate a reasonable differentiation across months within a year, with June yielding more accidents than each month of the autumn period, probably because more vehicle-km are driven on most road networks during early summer. Intuitive expectations seem to be justified to some extent from the use of models that also utilize exposure data. In particular, it appears that low temperature during wintertime, mostly, corresponds to some reduction of recorded accidents. The same is the case as total precipitation in a month increases, probably due to reduced mobility under rainy weather, but this effect is much less pronounced.

In future steps, some more elaborate models can be developed on the same topic, further elaborating in terms of texture and level of detail compared to the preceding analyses. This shall largely depend on data availability and context overall. To this end, one could, perhaps, investigate the potential impact of precipitation quantity on slight accidents frequency (perhaps only involving property damage); which also introduces the issue of reporting, as there is a tendency in several countries for many slight accidents to not be reported (Yannis et al., 2010).

Different road network users may call for varying data treatment as well. One may claim that, in territories where detailed exposure data are not available –and are not expected to be so soon-, there is a necessity to develop more parsimonious models, placing focus on precision in variables selection and modelling technique approach. In such cases there is often some “omitted variables” problem, simply because it is not possible to utilize what is reasonably considered as an ideal set of predictors. Still, one may want to obtain some idea of the impact of the remaining parameters to be used. So, if one clearly sets a research objective, it is most probable that even simple tools will be sufficient to reach reliable and applicable results. Models built with predictive application in minds would also in general be simpler, to minimize the needs to predict a large number of variables, before they can be applied.

# REFERENCES

Bergel R., Depire A. (2004a) Recherche d’une forme fonctionnelle d’un modèle agrégé de risque routier. Actes Inrets n° 90 du groupe de travail 2001 du Séminaire Modélisation du tafic, Arcueil (in French).

Bergel-Hayat R., Depire A. (2004b). Climate, road traffic and road risk – an aggregate approach. Proceedings of 10th WCTR (CD-Rom), World Conference on Transport Research Society, Istanbul, Turkey.

Dethlefsen C, Lundbye-Christensen S (2006). “Formulating State Space Models in R with Focus on Longitudinal Regression Models.” Journal of Statistical Software, 16(1), pp. 1-15. URL http://www.jstatsoft.org/v16/i01

Dobson AJ. (1990) An Introduction to Generalized Linear Models, 2nd ed., Chapman and Hall, London.

Eisenberg, D. (2004). The Mixed Effects of Precipitation on Traffic Crashes, Accident Analysis and Prevention, Vol. 36, pp. 637-647.

Fisher RA. (1934) Two New Properties of Mathematical Likelihood, Proceedings of the Royal Society A, Vol. 144, pp. 285–307.

Fridstrøm L., Ifver J., Ingebrigtsen S., Kulmala R., Thomsen L.K. (1995). Measuring The Contribution Of Randomness, Exposure, Weather, And Daylight To The Variation In Road Accident Counts. Accident Analysis & Prevention 27 (1), pp. 1-20.

Fridstrom, L., Ingebrigtsen, S. (1991). An aggregate accident model based on pooled, regional time-series data. Accident Analysis and Prevention 23(5), pp. 363-378.

Gill J. (2000) Generalized Linear Models: A Unified Approach, Sage University Papers Series on Quantitative Applications in the Social Sciences, 07–134, Sage, Thousand Oaks, CA.

Hermans, E., Wets, G., Van Den Bossche, F. (2006). Frequency and Severity of Belgian Road Traffic Accidents Studied by State-Space Methods. Journal of Transportation and Statistics, Vol. 9 no1, pp.63-76.

Karlaftis M, Yannis G. (2010) Weather Effects on Daily Traffic Accidents and Fatalities: A Time Series Count Data Approach, Proceedings of the 89th Annual meeting of the Transportation Research Board, Washington.

Lord D, Washington SP, Ivan JN. (2005) Poisson, Poisson-Gamma and Zero-Inflated Regression Models of Motor Vehicle Crashes: Balanc- ing Statistical Fit and Theory, Accident Analysis and Prevention, Vol. 37, pp. 35–46.

McCullagh P, Nelder JA. (1989) Generalized Linear Models, 2nd ed., Chapman Hall, New York.

ONISR (2012) La sécurité routière en France: bilan de l'année 2011. Paris, France: La Documentation Française.

Scott P. P. (1986). Modelling Time-Series of British Road Accident Data. Accident Analysis & Prevention, Vol. 18 n°2, pp.109-117.

Stipdonk H. (Ed.) (2008). Time series applications on road safety developments in Europe. Deliverable D7.10 of the EU FP6 project SafetyNet http://www.erso.eu

Yannis, G., C. Antoniou and E. Papadimitriou (2007). Road casualties and enforcement: Distributional assumptions of serially correlated count data. Traffic Injury Prevention, 8:3, pp. 300 - 308.

Yannis G., Petridou E., Evgenikos P., Chaziris A., Terzidis A., Desypris N. (2010), "Calculation of the underreporting level of road accident casualties in Greece", Proceedings of the 4th International Congress on Transport Research in Greece, Volos, September