Integrating Road Safety and Environmental Impact via Telematics: Modeling Traffic Accident Risk Using Vehicle Emissions

—

Juan Sebastian Yanez
Department of Econometrics,
Riskcenter-IREA,
Universitat de Barcelona,
Barcelona, Spain
yanez.juan_sebastian@ub.edu

Montserrat Guillen
Department of Econometrics,
Riskcenter-IREA,
Universitat de Barcelona,
Barcelona, Spain
mguillen@ub.edu

Paulina Roszkowska
Bayes Business School,
City St George's, University of London, United Kingdom
SGH Warsaw School of Economics,
Warsaw, Poland
paulina.roszkowska@city.ac.uk

Jens Perch Nielsen
Bayes Business School,
City St George's, University of London, United Kingdom
Jens.Nielsen.1@city.ac.uk

Abstract

High-emission private vehicles contribute disproportionately to urban air pollution, which, together with traffic accidents, constitutes a major public health threat and a leading cause of death among young adults. The tension between individual driving habits and collective environmental responsibility highlights the pressing need for strategies promoting safer driving. This paper introduces a novel approach to evaluating road crash risk by using air pollutants as exposure measures, thereby encouraging drivers to reduce both their environmental footprint and mitigate their crash risk. We analyze a dataset of over 1,500 at-fault crash-related claims recorded over two years by an insurance company, merged with detailed telematics driving data for individual vehicles. We show that emission factor models can effectively incorporate emission-based exposure measures into crash risk models. We also provide empirical evidence that incorporating behavioral telematics data allows pollutant-driven models to perform as efficiently as traditional distance-driven models. The approach reported has the potential to enhance road safety and reduce air pollution by directly linking environmentally conscious driving practices with lower road crash risks.

Keywords air pollution; crash risk; public health; road accident; telematics; transportation

1 Introduction

Private vehicles represent a growing concern to public health, making a significant contribution to accident-related deaths and posing serious socioeconomic challenges at both individual and national levels. According to the World Health Organization (WHO (2023)), road traffic injuries are the leading cause of death among children and young adults aged 5–29 years, serving to highlight their devastating personal impact. Moreover, these injuries disproportionately affect working-age individuals: for example, in 2023, 46% of road fatalities in the EU occurred among those aged 15–49 (European Commission (2024)). On the broader scale, road traffic accidents have substantial economic costs, estimated at between 1 and 3% of gross domestic product in most countries (WHO (2023)). The impact of private vehicles is undeniable; yet despite extensive efforts and campaigns to promote road safety, risky driving behaviors persist, intensifying the associated dangers.

Concurrent with these concerns, there has been a shift in societal priorities because of the effects of human-driven climate change and air pollution. The land transport sector is a major culprit here, with private vehicles making a substantial contribution to global carbon dioxide (CO_2) emissions, the greenhouse gas that drives climate change. In 2022, the domestic transport sector accounted for 23.77% of total CO_2 emissions in the EU (EEA (2024)). In addition, the WHO reports that the sector emits other harmful pollutants, including nitrogen oxides (NOx) and carbon monoxide (CO) (WHO (2021)). Urban areas, where motor vehicle density is highest, bear the brunt of these impacts: in 2022, 97% of the EU's urban population was exposed to fine particulate matter levels exceeding the WHO's 2021 guidelines (EEA (2021)). Electric vehicles (EVs) offer a pathway to reduce emissions; however, in many parts of the world, they remain outnumbered by internal combustion engine (ICE) vehicles. Thus, in the 27 EU Member States in 2023, EVs accounted for just 22.7% of new car registrations (EEA (2024)).

In short, the sector affects public health on two key fronts: road safety, on the one hand, and air pollution, on the other. Both are critical and have dedicated research fields in which academic debates are wide ranging. However, the lack of access to datasets that include both accident and emission statistics severely limits their joint study. Yet, there is great potential in so doing, as particle emissions and road accidents share many risk factors that, if diminished, could simultaneously reduce the risk of accidents and a driver's environmental footprint. In this way, safer driving practices, in terms of both air quality and accident risk reduction, can be promoted by advocating for the dual benefit of such practices. Moreover, by linking driving not only to crash risk but also to pollutant release, insurance schemes can be designed that price driving risk and environmental damage simultaneously.

To the best of our knowledge, this is the first paper to seek to integrate the two pivotal public health issues affecting road transportation: road safety and air pollution. We do so by modeling road accident risk as a function of particle emissions as opposed to relying on more conventional measures such as distance traveled or observation period (typically referred to as 'risk exposure'). Ours is an intuitive approach, as the literature has already proposed a variety of alternative proxies for exposure, including travel time and, even, fuel consumption.

As with other complex variables (e.g. time spent behind the wheel), data availability is the key limitation when seeking to employ emissions as a risk exposure measure. Yet, recent advances in telematics mean we can now circumvent this problem. Telematics data, gathered using on-board devices (OBD) and smartphone applications, provide information about driving behavior and road conditions on specific routes, including the number of kilometers driven (an exposure measure in itself), road type (urban, rural, etc.), speed, harsh braking events, and accelerations. In studies of pollutant emissions, telematics data are helpful insofar as they provide information about two of the three factors affecting vehicle emissions: driving behavior and road conditions (the third being vehicle characteristics). In accident prevention research, telematics data allow behavioral risk factors, such as the number of harsh accelerations and nighttime driving, to be identified.

The case study reported herein benefits from these recent technological advances. Specifically, we have access to a dataset containing two years' worth of telematics data from 17,405 drivers along with records of at-fault accidents. We complement this information with a macroscopic emissions model, adapted to our data, that converts telematics data into particle emission estimates. We then propose various models (ranging from simple to more complex) to predict traffic accident risk. We provide empirical evidence that modeling the number of accidents as a function of air pollutant emissions is as effective as modeling them as a function of distance, with the performance of the models being similar in terms of the precision of their estimations.

This paper contributes to the academic literature by proposing an alternative and innovative risk exposure measure hitherto unexplored in the road accident and prevention literature. In this way, we complement ongoing research on exposure measures in road accident modeling, earlier examples of which include traffic volume (Qin *et al.* (2006); Van den Bossche *et al.* (2005)), travel time (Chipman *et al.* (1992); Chipman *et al.* (1993), Pei *et al.* (2013); Shen *et al.* (2020)) and fuel (Fridstrom *et al.* (1995)).

From a modeling perspective, emissions represent a conceptually distinct exposure dimension that links risk assessment with environmental externalities. This dual relevance is particularly important in light of ongoing technological developments: advances in telematics, onboard diagnostics, and standardized emissions modeling are rapidly reducing measurement costs and increasing availability. As a result, emissions are becoming a feasible and scalable variable in applied settings.

The approach we report can be used, moreover, as a template for many practical applications. In insurance, an ecological spin might be given to policies based on telematics data, with premium discounts being linked to a client's environmental footprint. While this would have only a minimal impact on the premium cost for the underlying risk, it could provide the insured with ecological incentives to drive more safely. From a governmental perspective, key metrics for the joint risk factors, capturing the amount of emissions before an accident and the practices that reduce both effects, could be usefully included in the reports of road safety or environmental agencies.

2 Background

Understanding the environmental and safety impacts of land transportation is critical for addressing contemporary public health challenges. Research focused on the prediction of pollutant emissions, fuel consumption, and road accident risks provides both motivation and a foundation for the present study. Indeed, several authors have hinted at the importance of considering these two major problems simultaneously. For instance, Segui-Gomez *et al.* (2011) advocated reducing the distance driven to avoid accidents, while highlighting the potential benefits for air pollution reduction. Likewise, in the ambient air pollution survey conducted by Anenberg *et al.* (2016), the authors called for the development of assessment tools that address both air pollution and related health risks, including vehicle accidents. Here, our access to extensive telematics data allows us to bridge this gap by modeling road accident risk as a function of particle emissions.

2.1 Pollutant Emissions and Fuel Consumption Predictions

Over the last 30 years, air pollution attributable to private vehicles has been studied using different statistical methods and datasets. Researchers have focused on two distinct objectives: predicting fuel consumption per kilometer (petrol or diesel) and estimating pollutant emissions per kilometer (CO_2, NOx, CO, or hydrocarbons). Whichever variable is studied, be it particles or fuel, the complexity of models depends on the level of detail of the input data and the scale at which the measurements are taken.

The literature on predicting pollutant emissions and fuel consumption can be divided into two broad categories: macroscopic and microscopic (Nejadkoorki *et al.* (2008); Nocera *et al.* (2017)). Macroscopic models estimate emissions (i.e., the mass of pollutants) or fuel consumption per unit distance using aggregate values over large areas and long periods. For instance, government agencies such as the United Kingdom's National Atmospheric Emissions Inventory (NAEI) and the European Environment Agency (EEA) provide big datasets for nationwide emissions. Other large-scale models estimate emissions based on mean travel speed, the case of the U.S. Environmental Protection Agency's Mobile Source Emissions Factor (MOBILE), the California Air Resources Board's EMission FACtor (EMFAC) and the Calculations of Emissions from Road Transport (COPERT) (Ntziachristos *et al.* (2009)). Smit *et al.* (2008a) and Smit *et al.* (2010) list other models that incorporate traffic conditions as input variables, including the Handbook Emission Factors for Road Transport (HBEFA; De Haan & Keller (2004)), the Assessment and Reliability of Transport Emission Models and Inventory Systems (ARTEMIS; Andre *et al.* (2004)) and Traffic Emissions and Energetics (TEE; Negreti (1996)).

In microscopic models, the goal is to measure emissions on a much smaller scale, that is, for specific vehicles, trips, or driving cycles (acceleration, deceleration, etc.). The models vary in complexity and typically need detailed input data, requiring engineering expertise to develop. Examples include the Comprehensive Modal Emission Model (CMEM; Scora *et al.* (2006)), the Virginia Tech Microscopic energy and emission model (VT Micro; Ahn *et al.* (2002); Rakha *et al.* (2004)), and the Passenger Car and the Heavy-Duty Emission Model (PHEM; Hausberger *et al.* (2013)). The main limitation of these models is their reliance on detailed vehicle parameters that can generally only be acquired directly from the manufacturer. In addition to these models that use engineering data, scholars have developed multivariate regressions that use vehicle class and driving pattern factors (e.g. speed, driving cycles, etc.), such as the Mobile Emissions Assessment System for Urban and Regional Evaluation (MEASURE; Fomunung *et al.* (2000)) and the VERSIT+ model (Smit *et al.* (2007)).

Recently, the increasing availability of telematics data from global positioning system (GPS) devices, OBDs, and smartphone applications has facilitated the collection of trip- and driver-specific information. Trip-specific data include road type and driving conditions, while driver-specific capture behavior such as aggressiveness and speeding. These datasets have allowed researchers to identify behaviors that affect fuel consumption (i.e. eco-driving). Relevant studies using smartphone data include Pirayre *et al.* (2022), Yao *et al.* (2020) and Meseguer *et al.* (2017) (for a general overview see Fafoutellis *et al.* (2020)). Other studies have focused on predicting emissions directly. For instance, Rodriguez *et al.* (2023) and Thibault *et al.* (2016) studied behavioral profiles based on geographical data from smartphones, while Sentoff *et al.* (2015) and Huboyo *et al.* (2017) developed models using data directly from OBDs.

2.2 Road Accident Risk

Studying the factors that influence road accident risk presents several challenges, primarily due to the relative infrequency of accidents compared to more common driving events, such as acceleration and braking. This risk can be influenced by a range of factors that include the weather, driver-specific behavior, and vehicle dynamics. In a recent study, Wang *et al.* (2025) found evidence in China of a significant association between road accidents and extreme pollution levels.

Several studies have investigated the impact of weather on accident risk. For instance, El-Basyouny *et al.* (2014) employed a multivariate Poisson model to examine how time and weather conditions influence accidents. Similarly, Naik *et al.* (2016) used high-resolution data from weather stations across the United States to assess the effects of weather on road safety. In Finland, Malin *et al.* (2019) proposed a Palm probability framework to study the relationship between the weather and accident risk. More recently, Becker *et al.* (2000) used a generalized additive model (GAM) to analyze the combined effects of weather and congestion on accident risk.

Road type and congestion levels are critical determinants of accident risk. Ahn *et al.* (2004) applied a logistic distribution to examine the impact of traffic conditions on accidents. Yu *et al.* (2013) used data on freeway congestion levels in a multivariate model to predict accident risk. More recently, Guo *et al.* (2021) employed a random forest approach to identify risk factors and understand the interplay between traffic flow and risky driving behavior. Comprehensive reviews by Wang *et al.* (2013) and Retallack *et al.* (2019) summarize this stream of research.

Advances in telematics have enabled detailed studies of driver-specific behaviors and there has been a plethora of such research in insurance economics (Guillen *et al.* (2021); Duval *et al.* (2023); Bauer *et al.* (2024); Holzapfel *et al.* (2024); Soleymanian *et al.* (2025)). Both Ma *et al.* (2018) and Quddus (2002) found that harsh braking and rapid acceleration increase accident rates. Location-based analyses also show that areas with frequent acceleration and braking events are associated with higher accident risk (Stipancic *et al.* (2018)). In actuarial science research, telematics have been used to develop detailed driver risk profiles. For example, Guillen *et al.* (2019) showed that specific behavioral data (e.g. speed limit violations and urban driving) captured by OBDs increase expected accident frequency. Wüthrich (2017) designed speed-acceleration heatmaps to identify driver risk profiles, while Gao & Wüthrich (2019) employed a convolutional neural network to classify telematics data, providing detailed insights into varying risk levels. Similarly, Huang & Meng (2019) applied logistic regression and machine learning techniques to telematics-derived data to classify drivers into distinct risk categories. Across these and other studies, acceleration/deceleration and speed have been consistently identified as critical factors in road accident analysis.

An essential component of accident risk modeling is the exposure measure, which reflects how long a driver is at risk of an accident. In the literature, the most commonly used exposure measures have been either distance driven or time (i.e. the observation period). From a macroscopic perspective, time is often used as a simplified proxy due to limited access to detailed driver data, a common limitation in the insurance industry (see Ohlsson *et al.* (2010), for a comprehensive overview). From a microscopic perspective, both distance and time have been used as exposure measures. For example, Ayuso *et al.* (2014) and Ayuso *et al.* (2016a) examined the distance driven before an accident using Weibull distributions, while Boucher *et al.* (2017) employed GAMs incorporating both distance traveled and time. Recognizing the critical importance of employing these two exposure measures, several studies have explored the relationship between them (Pei *et al.* (2013); Shen *et al.* (2020); and Chipman *et al.* (1992)).

3 Methodology

Our approach to modeling road accident risk is motivated by the availability of telematics data, which have proven to be a valuable resource in both accident risk assessment and emissions research. Figure 1 provides a graphical overview of the data used in this study, highlighting the variables selected and the modeling process.

Before specifying our model, it is important to describe the makeup of a typical telematics dataset and its relevance for modeling emissions and accident risk. The variables derived from telematics data can be broadly categorized into two groups: factors that significantly influence both accidents and emissions, and factors that primarily impact on only one of these outcomes. A key link between emissions and road accidents is that the primary drivers of emissions – that is, vehicle characteristics, road conditions, and driving behavior – also play a critical role in modeling accident risk.

In conventional modeling approaches, an exposure measure, such as distance traveled, is typically included alongside other predictors specific to estimating either emissions or accident risk. Indeed, the choice of factors in the existing literature is strongly influenced by the availability of input data.

This study builds upon the positive correlation observed between emissions and accident risk by leveraging the risk factors shared by the two outcomes. In so doing, our primary contribution lies in the substitution of traditional exposure measures, such as distance traveled, with emission estimates. In this way, we model accident risk as a function of vehicle emissions, a novel approach that is illustrated in Figure 1.

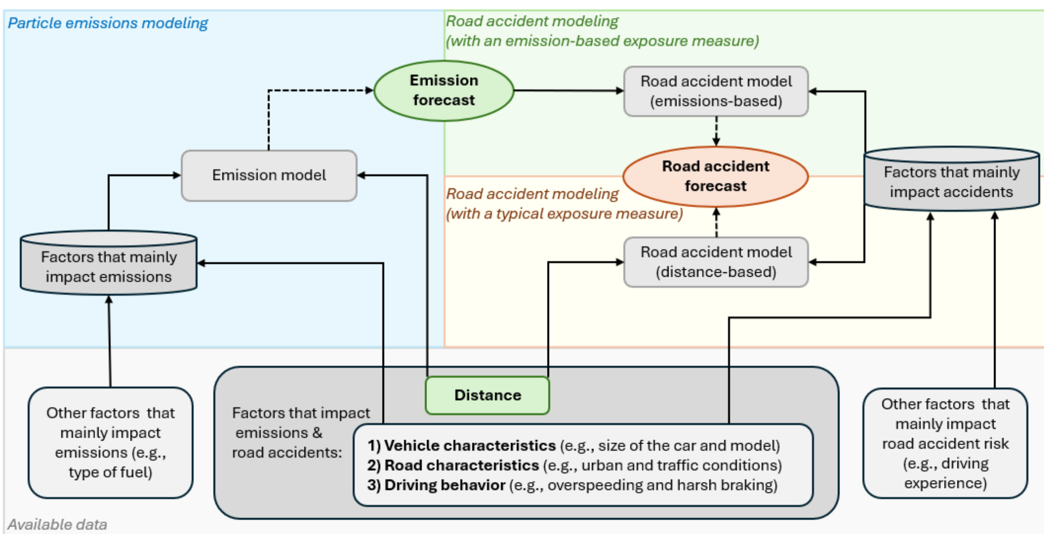

Figure 1: Graphical overview of the emissions and road accident risk modeling process.

After broadly outlining the impact of our telematics data on the modeling process, we next describe the statistical methods employed in our empirical analysis. We begin by introducing the notation used for each variable in this paper and follow this with an in-depth discussion of the probability distributions considered in the analysis.

3.1 Statistical Framework

Let $\mathcal{D}$ denote a dataset containing observations from I drivers over J years, where, for each observation (i, j) (corresponding to the $i-th$ driver in the $j-th$ year), the accident counts are recorded. In addition, suppose that for each driver, static covariates describing individual and vehicle characteristics are also available. Let us also assume that yearly telematics data capturing driving behavior and distance driven are recorded. Specifically, we consider three distinct telematics-based exposure measures k: distance driven (k=1), estimated NOx emissions (k=2), and estimated CO emissions (in grams) (k=3).

Based on our assumptions, for $k = 1, 2, 3$ and $i = 1, \ldots, I$ and $j = 1, \ldots, J$, we define the following variables:

- $N_{\mathrm{i,j}}$; the number of at-fault accidents for driver i in year j,
- $T^{(0)}_{\mathrm{i,j}} = 1$ $\forall$i,j; the exposure in years for driver i in year j,
- $T^{(k)}_{\mathrm{i,j}}$; the telematics-based exposure of type k for driver i in year j,
- $Z_{\mathrm{i,j}}$; a vector containing additional telematics data for driver i in year j,
- X_i; a vector containing static, non-telematics covariates for driver i.

Next, we identify the statistical framework of our accident frequency models, conditional to the telematics and static covariate vector. Specifically, we consider a classical generalized linear model (GLM) structure, where the distribution of the response variable is defined through its mean, and we consider three different model structures for each exposure measure (see Table 1).

Table 1: Mean parameter $\left(\mu^{(k)}_{i,j}\right)$ structure for each exposure measure k

Tag Number	$\mu^{(k)}_{i,j}$ formula	Description
(0)	$T^{(0)}_{i,j} \exp\left(\mathbf{X_i'}\boldsymbol{\beta}\right)$	Benchmark model, uses time as exposure measure
(I)	$T^{(k)}_{i,j} \exp\left(\mathbf{X_i'}\boldsymbol{\beta}\right)$	Telematics-based exposure as an offset
(II)	$\exp\left(\mathbf{X_i'}\boldsymbol{\beta} + \ln\left(T^{(k)}_{i,j}\right)\alpha^{(k)}\right)$	Log of the telematics-based exposure as covariate
(III)	$\exp\left(\mathbf{X_i'}\boldsymbol{\beta} + \ln\left(T^{(k)}_{i,j}\right)\alpha^{(k)} + \mathbf{Z_{i,j}'}\boldsymbol{\gamma}\right)$	Model (**II**), with the other telematics covariates

β, $\alpha^{(k)}$, and γ are the parameter vectors that combine linearly to their respective covariates.

3.2 Distributions

The Poisson regression model is the classical choice for modeling count variables due to its simplicity and ability to account for exposure either as an offset term or as a covariate influencing the mean parameter (μ). Consequently, we adopt the Poisson model as our baseline specification and use it as a benchmark against which to evaluate more complex, flexible alternatives.

The negative binomial regression is another widely used model for count data. Unlike the Poisson regression, it introduces an additional dispersion parameter, which is typically set as a constant. This added complexity allows the model to relax the equidispersion assumption inherent to the Poisson (i.e. that the expected value of the response variable is equal to its variance). The flexibility provided by the negative binomial model makes it well suited for overdispersed data, a common feature of low-frequency events such as road accidents.

The zero-inflated Poisson (ZIP) regression provides a method for addressing an excessive number of observations with no events (see Tang *et al.* (2014))). Specifically, the response variable is assumed to be a random draw from one of two events: (i) with probability $\sigma \in [0, 1]$, we assume that the observation always equals 0; (ii) with probability $1 - \sigma$, we assume that the observation follows a Poisson distribution. The excessive-zero component is modeled by a Bernoulli variable, which is equal to 0 when we draw from the Poisson model and 1 otherwise. Given that the Bernouilli regression is based on the probability of the variable being equal to 1 as a parameter (i.e., $\sigma \in [0, 1]$), incorporating an exposure measure as an offset to σ may violate the boundaries in which the parameter is defined (e.g. obtaining values where ($\sigma > 1$). To address this, exposure needs to be incorporated in the model as a covariate, ensuring that the probability of drawing from the Poisson distribution appropriately reflects the driver's level of exposure.

We next summarize the main components of each regression using the variables defined in Section 3.1. Note that the parameters of each model are obtained by maximizing the log-likelihood function.

3.2.1 *Poisson*

The Poisson model is defined as in Equation (1):

$$\left(N_{i,j} | \mathbf{X}_i, \mathbf{Z}_{i,j}, T^{(k)}_{i,j}\right) \sim \text{PO}\left(\mu^{(k)}_{i,j}\right), \text{ for } T^{(k)}_{i,j} > 0, \tag{1}$$

where, $\mu_{\mathrm{i,j}}^{(k)}$ is the mean parameter of the Poisson distribution (PO) and can be defined with the structures outlines in Table 1).

The mean and the variance of the Poisson model are given by:

$$\begin{aligned} \mathrm{E}\left[N_{i,j}|\mathbf{X}_i,\mathbf{Z}_{i,j},T_{i,j}^{(k)}\right] &= \mu_{i,j}^{(k)}, \\ \mathrm{Var}\left[N_{i,j}|\mathbf{X}_i,\mathbf{Z}_{i,j},T_{i,j}^{(k)}\right] &= \mu_{i,j}^{(k)}. \end{aligned}$$

3.2.2 *Negative Binomial*

The negative binomial model is defined as in Equation (2):

$$\left(N_{i,j}|\mathbf{X}_i,\mathbf{Z}_{i,j},T_{i,j}^{(k)}\right) \sim \mathrm{NB}\left(\mu_{i,j}^{(k)},\sigma^{(k)}\right), \text{ for } T_{i,j}^{(k)} > 0, \tag{2}$$

where $\mu_{\mathrm{i,j}}^{(k)}$ and $\sigma^{(k)}$ are, respectively, the mean parameter and the dispersion parameter of the negative binomial distribution (NB). The former can be defined with the structures outlined in Table 1; the latter is a constant.

The mean and the variance of the negative binomial model are given by:

$$\begin{aligned} \mathrm{E}\left[N_{i,j}|\mathbf{X}_i,\mathbf{Z}_{i,j},T_{i,j}^{(k)}\right] &= \mu_{i,j}^{(k)}, \\ \mathrm{Var}\left[N_{i,j}|\mathbf{X}_i,\mathbf{Z}_{i,j},T_{i,j}^{(k)}\right] &= \mu_{i,j}^{(k)} + \sigma^{(k)}\left(\mu_{i,j}^{(k)}\right)^2. \end{aligned}$$

3.2.3 *Zero-Inflated Poisson*

Two components define the ZIP model. We set $N_{\mathrm{i,j}} = 0$ with probability $\sigma_{\mathrm{i,j}}^{(k)}$. Then, we set a Poisson distribution with probability $1-\sigma_{\mathrm{i,j}}^{(k)}$. The ZIP distribution is defined as in Equation (3):

$$\left(N_{i,j}|\mathbf{X}_i,\mathbf{Z}_{i,j},T_{i,j}^{(k)}\right) \sim \mathrm{ZIP}\left(\mu_{i,j}^{(k)},\sigma_{i,j}^{(k)}\right), \text{ for } T_{i,j}^{(k)} > 0, \tag{3}$$

where, $\mu_{\mathrm{i,j}}^{(k)}$ and $\sigma_{\mathrm{i,j}}^{(k)}$ are the mean parameter of the Poisson distribution and the probability of not drawing from the Poisson distribution, respectively. The former can be defined with the structures outlined in Table 1; the latter is a function of the exposure measure, as in Equation (5),

$$\sigma_{i,j}^{(k)} = \frac{\exp\left(\psi_0^{(k)} + \ln\left(T_{i,j}^{(k)}\right)\psi_1^{(k)}\right)}{1 + \exp\left(\psi_0^{(k)} + \ln\left(T_{i,j}^{(k)}\right)\psi_1^{(k)}\right)}, \text{ for } T_{i,j}^{(k)} > 0, \tag{4}$$

where, $\psi_0^{(k)}$ is the intercept and $\psi_1^{(k)}$ is the parameter of the log of the k^{th} exposure measure. Note that when the exposure is measured in years and all policy holders have a complete one-year contract, we have $T_{\mathrm{i,j}}^{(0)} = 1$, ∀i,j. Thus, the log of the exposure becomes 0, which makes $\sigma_{\mathrm{i,j}}^{(0)}$ a constant.

The mean and the variance of the ZIP model are given by:

$$\begin{aligned}
\mathrm{E}\left[N_{i,j}|\mathbf{X}_i, \mathbf{Z}_{i,j}, T_{i,j}^{(k)}\right] &= \left(1 - \sigma_{i,j}^{(k)}\right)\mu_{i,j}^{(k)}, \\
\mathrm{Var}\left[N_{i,j}|\mathbf{X}_i, \mathbf{Z}_{i,j}, T_{i,j}^{(k)}\right] &= \left(1 - \sigma_{i,j}^{(k)}\right)\mu_{i,j}^{(k)}\left(1 + 2\mu_{i,j}^{(k)}\right).
\end{aligned}$$

3.2.4 *Combined Actuarial Neural Network*

To complement the GLM and GAM approaches adopted in our case study, we also implement supervised learning techniques. Among available options, we consider the combined actuarial neural network (CANN) model, as proposed by Wüthrich & Merz (2019) to be especially relevant here for three reasons. First, its primary strength is that it manages to combine the flexibility and pattern recognition capabilities of classical neural network (NNs) with the familiar structure of classical regression models such as GLMs. In this way, we can complement the information provided by a GLM. Second, given that we have already formulated our models in terms of the mean parameter, the CANN model directly extends this structure by incorporating an NN architecture. Third, the CANN model has already shown promising results in accident count modeling using telematics data Duval *et al.* (2024).

To implement the CANN here, we follow the CANN-Poisson algorithm outlined by Duval *et al.* (2024). Specifically, we consider a two-layered multilayer perceptron (MLP) with rectified linear unit (ReLU) activation functions, combined with batch normalization between the fully connected layers. For the output activation function, we opt for the exponential function introduced by Wüthrich & Merz (2019). Although Duval *et al.* (2024) recommend the softplus function to mitigate the numerical instability dependent on the input data, we find the exponential function to be preferable in our context. It allows for the seamless incorporation of an offset component to capture risk exposure – a key aspect of our analysis – and, importantly, we did not observe stability problems when using this function.

The CANN approach employed can also be summarized through the lens of the mean parameter. For example, when complementing model (**III**) in Table 3.1, we obtain

$$\mu_{i,j}^{(k)} = \exp\left(\mathbf{X_i'}\boldsymbol{\beta} + \ln\left(T_{i,j}^{(k)}\right)\alpha^{(k)} + \mathbf{Z_{i,j}'}\boldsymbol{\gamma} + a\left(\mathbf{X_i'}, \mathbf{Z_{i,j}'}, \theta\right)\right) \tag{5}$$

where β, $\alpha^{(k)}$, γ are the parameter vectors that combine linearly with their respective covariates. θ are the weight matrices and bias vectors of the MLP, which is denoted by $a(\cdot)$. A schematic overview is provided in Figure 2.

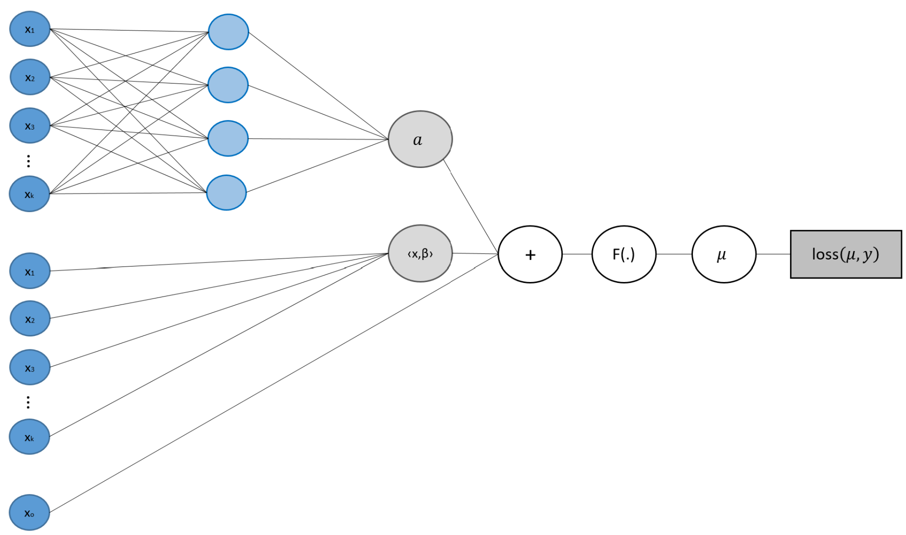

Figure 2: CANN architecture for the Poisson specification. The MLP's preactivation output value is added to the log-linear model's preactivation output value of the GLM component, defined as the linear combination of parameters and covariances, before being transformed with the exponential function. The loss function is specified as the Poisson deviance.

4 Results

4.1 Data and Descriptive Statistics

In our study, we draw on a dataset containing information about 17,405 drivers insured by a Spanish company between 2017 and 2018. All insured were covered for the two entire years. A major advantage of this dataset is its size, which includes over 1,700 claims at-fault reported over the two-year observation window. Claims not at-fault are not considered in our study. Table 2 summarizes the distribution of claim counts per driver for this period.

Table 2: Frequency of claims (17,405 drivers, in 2017 and 2018)

	At-fault claim counts				
	0	1	2	3	4
Absolute Frequency	15835	1450	112	6	2

In addition to claim counts, we have access both to static covariates about each driver and vehicle and to telematics data collected via OBDs. These data include total kilometers driven, disaggregated into three distinct road types: Urban, Rural, and Motorway. Table 3 presents the descriptive statistics for the mean yearly distance traveled on each road type. We observe that, on average, drivers in our sample tend to cover the greatest distance on motorways, followed by rural roads, and to cover the least distance in urban areas.

Table 3: Descriptive statistics for the mean yearly distance driven by road type (in km)

Descriptive	Road type			
Statistics	Urban	Rural	Motorway	All
Mean	2,349	2,741	5,142	10,233
1st Quartile	1,311	977	1,660	5,473
Median	2,050	1,985	3,809	9,103
2nd Quartile	3,038	3,636	7,220	13,692
Standard Deviation	1,557	2,619	4,746	6,463

We further consider telematics indicators related to risk-taking behavior, namely the percentage of kilometers driven at night and the percentage driven above the speed limit. Table 4 presents the descriptive statistics for both telematics and static covariates, distinguishing between drivers with and without claims.

Table 4: . Descriptive statistics of driver and vehicle covariates

Covariate	All sample $n = 17405$ (100%)		Drivers without claims* $n = 15835$ (91%)		Drivers with claims $n = 1570$ (9%)	
	Mean	SD	Mean	SD	Mean	SD
Driver Age	29.07	4.62	29.10	4.60	28.77	4.78
Gender Male %	43.36	-	43.19	-	45.10	-
Vehicle Power	102.71	30.37	102.58	30.19	104.06	32.10
Km at Night %	10.02	7.80	9.89	7.73	11.32	8.38
Km overspeed %	3.60	6.07	3.58	6.04	3.84	6.33
Km urban %	26.39	15.93	26.21	15.89	28.23	16.22

* only claims at-fault are considered. The drivers without at-fault claims that we consider in this segmentation could have been involved in claims not at-fault.

Based on these descriptive results, we note differences between drivers with and without at-fault claims. Among the static covariates, the riskier cohort tends to be younger on average and includes a slightly higher proportion of male drivers. Additionally, the vehicles of the riskier cohort generally have more powerful engines. Among the telematics covariates, drivers with claims present a higher average percentage of night-time driving, overspeeding, and urban driving. However, the differences between drivers with and without claims are quantitatively very small. Specially when we take into account that the relative small size oof the cohort of clients that declared clamis during thsi two year period.

Our data set does not contain direct measures of air pollution attributable to drivers. Therefore, we opt to estimate these particle levels using an emission forecasting model.

4.2 Emission Modeling

We employ a macroscopic model to estimate emissions, based on emission factors provided by the UK's national agency responsible for reporting annual emissions, the NAEI. These factors enable us to broadly estimate drivers' yearly emissions by multiplying the respective factors by the kilometers driven by each road type. As discussed, our data offer essential insights into road accident risk as they combine telematics variables and claim counts for a relatively large population of drivers. While a microscopic model could yield more precise estimates, we lack access to sufficiently detailed telematics information about drivers' trips necessary to effectively implement current state-of-the-art microscopic approaches.

To account for differences in the national-level fleet composition between the UK and Spain, we turn to the Ministry for the Ecological Transition (MITECO), the Spanish agency responsible for pollution reports. Unfortunately, the MITECO does not report emission factors for cars under distinct road conditions. However, it reports the total yearly mileage by vehicle as well as the fleet composition by fuel type (diesel or petrol) (see Ministerio para la Transición Ecológica (2020)). We have summarized this information specifically for cars during the observation period 2017-2018 (see Table 9). Moreover, for comparative purposes, we have also included the parallel table for the UK's fleet for the same years. This latter table was constructed by cross-referencing data from the NAIE and the "Road Traffic Estimates: Great Britain" from the Department for Transport (see Department for Transport (2018) and Department for Transport (2019)).

For this analysis, we focus on two pollutants with substantial public health impacts: nitrogen oxides (NOx) and carbon monoxide (CO). The choice of these pollutants are primarily motivated by air quality and public health considerations, rather than direct greenhouse gas effects (Note, the NAIE did not report CO_2 emissions from private cars in 2017 and 2018.) The emission factors from the NAIE are summarized in Table 5[1]. Additional descriptive statistics for our emission estimates are presented in the Appendix (see Table 10). According to these emission factors, vehicles emitted the highest levels of pollutant particles in urban areas, followed by motorways and, then, by rural roads.

Table 5: Estimated combined hot exhaust and cold start emission factors (g/km), by road type

Country		Spain		UK		Spain		UK	
Year		2017	2018	2017	2018	2017	2018	2017	2018
In g/km		**NOx**	**NOx**	**NOx**	**NOx**	**CO**	**CO**	**CO**	**CO**
	urban	0.5536	0.5361	0.3728	0.3555	0.6705	0.5819	0.7957	0.7027
All cars	rural	0.3984	0.3795	0.2798	0.2662	0.1291	0.1271	0.2607	0.2463
	m-way	0.4522	0.4281	0.3537	0.3367	0.1726	0.1727	0.3078	0.2925

[1]'Cold start' emissions are produced when the engine and catalytic converter are not yet at their optimal operating conditions leading to higher pollutant emissions (Reiter *et al.* , 2016).

We run a bivariate kernel density estimation (KDE) for the variable pairs (km, CO) and (km, NOx), based on the 2017 and 2018 Spanish factors, to visualize the relationship between particle emissions and kilometers driven. The KDE for each pair is constructed using a multivariate normal distribution and a normal scale bandwidth selector, following standard practice in the multivariate KDE literature (see, for example, Chacón & Duong (2018), Chs. 2-3). Figure 3 provides a graphical representation of the results using heatmaps. We observe that the variance depends on the pollutant selected in each pair. Specifically, the pair (km, NOx) presents a strong positive relationship: longer distances are clearly associated with higher emissions. In contrast, the other pair (km, CO) presents a weaker association, particularly at longer distances where the variance in CO emissions is greater.

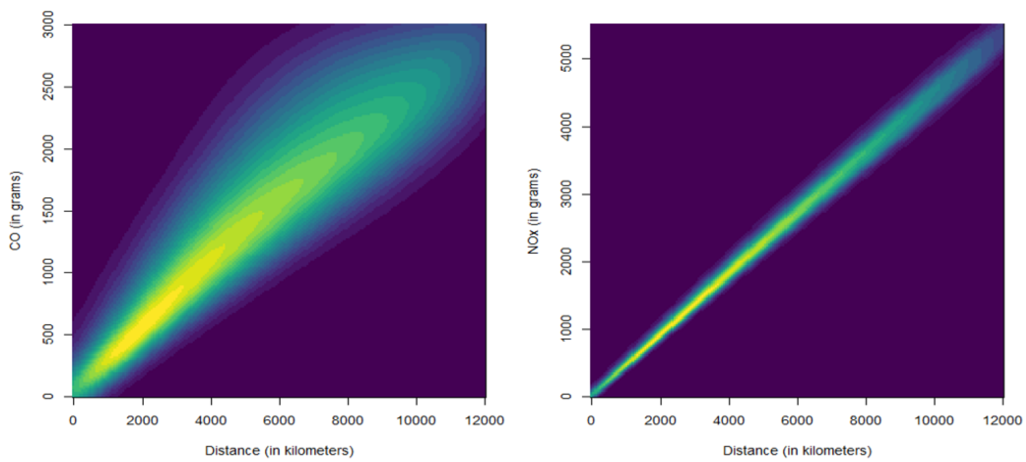

Figure 3: Bivariate kernel density distribution for kilometers driven and pollutant particles. Left: CO. Right: NOx.

In short, the macroscopic model yields a linear relationship between distance driven and particle emissions, but the strength of this association varies by pollutant. These estimates provide the emission inputs for the accident risk models discussed in Subsection 4.3.

4.3 Road Accident Modeling

To model road accident risk using estimates of particle emissions, we first conduct a comparative analysis of the impact of different exposure measures. We then introduce the models used for predictive analysis, compare their predictive power in terms of goodness-of-fit, and, finally, perform an out-of-sample simulation to compare emission-based and distance-based approaches.

4.3.1 *Exposure Measure Comparison*

We compare the impact of alternative exposure measures on road accident risk. Figures 4 and 5 visually illustrate these relationships. For the analysis, we split the dataset into two subsamples: one containing driver-year observations with no at-fault reported claims, and another including observations with at least one reported at-fault claim. Using the bivariate kernel approach outlined in Section 4.2, we estimate the density distributions for each subsample.

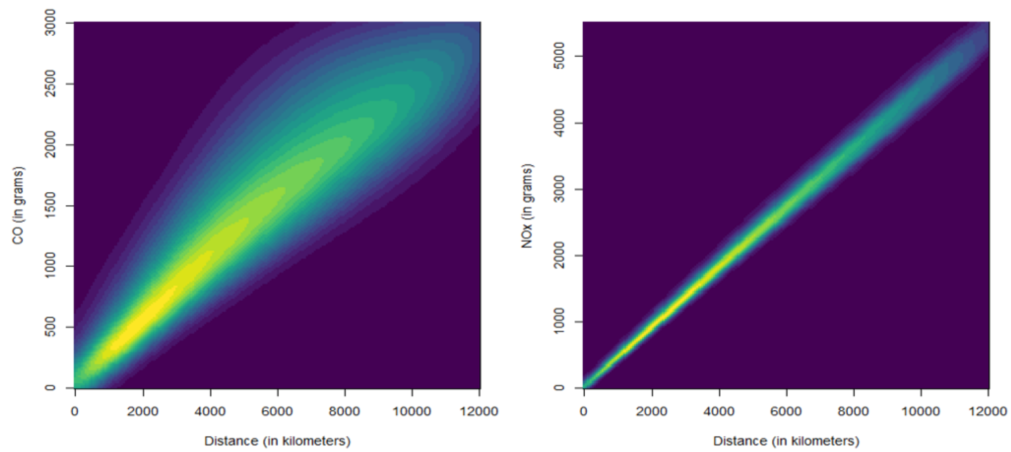

Figure 4: Bivariate kernel density distribution for kilometers driven and pollutant particles in the sample without at-fault claims. Left: CO. Right: NOx.

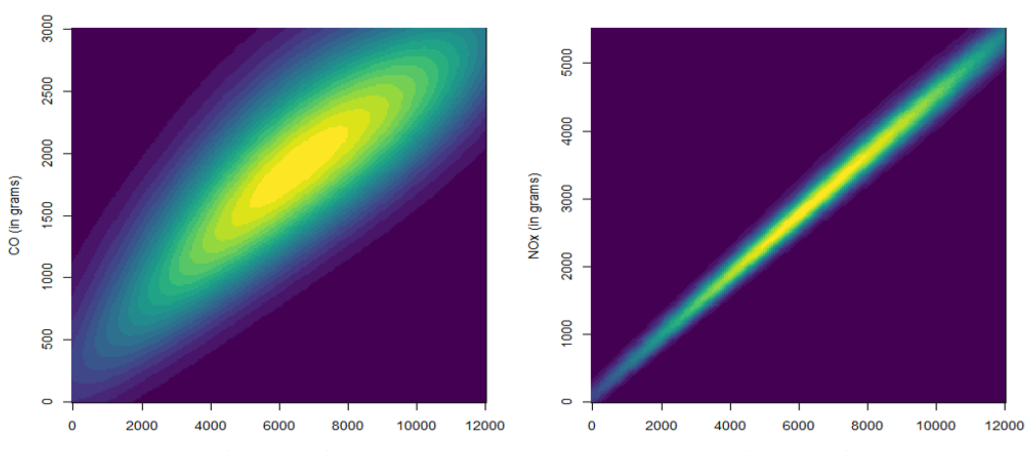

Figure 5: Bivariate kernel density distribution for kilometers driven and pollutant particles in the sample with claims. Left: CO. Right: NOx.

Figures 4 and 5 present heatmaps for the subsamples without and with at-fault claims, respectively. Notably, driver-year observations with at least one at-fault claim tend to exhibit a higher number of kilometers driven and higher emissions as the mode of the distributions shifts proportionally. Additionally, consistent with findings in Section 4.2, the pollutant under observation affects the variance of the distributions: the (km, CO) pair exhibits greater dispersion than that of the (km, NOx) pair. Visual inspection suggests that all three exposure measures – kilometers driven (Km), CO emissions, and NOx emissions – are relevant and provide a slightly different perspective in explaining the variation in road accident frequency.

4.3.2 *Model Description*

Following the comparative analysis of exposure measures using kernel distributions, we train predictive models and evaluate the significance of covariates based on the three statistical distributions introduced in Section 3: the Poisson (Equation (1)), the negative binomial (Equation (2)), and the ZIP (Equation (3)). For model development, 80% of the 2017 data is used as the training set and the remaining 20% as the validation set. Out-of-sample comparisons in this section are therefore based on the validation data set. The data for 2018 are used in the next section. Moreover, our main results will consider the Spanish emission factors from 2017 (see Table 5).

The core driver, vehicle, and behavioral characteristics highlighted are already explicitly incorporated in our model and more information is provided through standard telematics covariates. This is an important point: emissions are not intended to substitute for these variables, nor to proxy for omitted driver heterogeneity. Rather, they are introduced in addition to a rich set of controls that already capture driving style, usage, and individual risk factors. Consequently, any incremental contribution of emissions should be interpreted as over and above what is already explained by driver, vehicle, and telematics characteristics.

Hereon in, we opt to focus specifically on type (**III**) models, and we analyze covariate significance using *z-tests*. Table 6 presents the results for the negative binomial model, while comparable results for the other two distributions, in terms of significance, are provided in the Appendix (Tables 11- 12).

Table 6: Negative binomial model with static covariates, telematics covariates, and exposure measures

	Risk exposure measures							
	Years		Km		NOx		CO	
	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value
(μ intercept)	-3.317	< 0.001	-5.391	< 0.001	-9.950	< 0.001	-9.463	< 0.001
Driver Age	-0.017	0.036	-0.014	0.094	-0.014	0.092	-0.014	0.089
Gender Male	-0.015	0.838	-0.060	0.414	-0.061	0.407	-0.065	0.379
Vehicle Power	0.003	0.002	0.003	0.019	0.003	0.019	0.003	0.022
Km Night	0.018	< 0.001	0.013	0.001	0.013	0.001	0.012	0.003
Km overspeed	-0.002	0.666	-0.008	0.147	-0.009	0.136	-0.008	0.157
Km Urban	0.010	< 0.001	0.026	< 0.001	0.024	< 0.001	0.014	< 0.001
$\log\left(T_{i,j}^{(k)}\right)$			0.752	< 0.001	0.752	< 0.001	0.769	< 0.001
σ	1.883	0.021	3.227	0.080	3.253	0.081	3.431	0.089

Across all models, a consistent pattern emerges. When using a 5% p-value threshold, the coefficients of driver age and vehicle power stand out as significant static covariate effects. Among the telematics covariates, the coefficients of percentages of kilometers driven in urban areas and at night consistently fall below the 5% significance threshold, while gender and the percentage of kilometers driven above the speed limit consistently underperform in terms of their *z-test* results. The non-significance of gender aligns with prior findings indicating that it loses explanatory power when telematics variables are considered (see, for example, Ayuso *et al.* (2014)). However, the non-significance of overspeeding contradicts the consensus presented in the literature (see, for example, Steg & van Brussel (2009); Pawar *et al.* (2020)), suggesting that this effect is captured by correlated components of the model, such as vehicle power and distance driven in urban areas.

Finally, the exposure measures demonstrate substantial predictive power across all models. Having assessed the covariates and exposure measures in our models, we can now compare model performance using a range of statistical tools.

4.3.3 *Goodness-of-fit Comparison*

We compare the performance of the aforementioned models using the Akaike information criterion (AIC) and both the in-sample and out-of-sample mean squared error (MSE), as summarized in Table 7. Across all distributions and evaluation criteria, incorporating any of the three exposure measures – either as offsets or as covariates (models (**I**) and (**II**), respectively) – enhances the baseline models (**0**). Notably, treating the exposure measures as covariates consistently reduces both the AIC and MSE across all models.

Adding telematics covariates leads to additional model improvements, as evidenced by the general performance gains observed when comparing models (**III**) and (**II**). These improvements are consistent across all distributions and exposure measures, highlighting the value of incorporating telematics information in these risk models.

Table 7: Akaike information criterion (AIC) and mean squared error (MSE) (in-sample and out-of-sample) for selected models with various exposure measures. Bold figures indicate the best exposure measure for each model. These results use Spain's emission factors from 2017.

		Poisson			Negative Binomial			Zero-Inflated Poisson			CANN		
Criteria	Model	Km	NOx	CO	Km	NOx	CO	Km	NOx	CO	Km	NOx	CO
AIC	**(0)**†	5552.28	5552.28	5552.28	5547.71	5547.71	5547.71	5548.88	5548.88	5548.88			
	(I)	5585.30	5559.23	**5445.21**	5577.84	5552.90	**5443.12**	5496.15	5486.46	**5432.51**			
	(II)	5492.67	5484.23	**5429.77**	5489.48	5481.23	**5428.16**	5479.20	5472.44	**5428.16**	5436.82	5464.24	**5412.74**
	(III)	5394.66	5393.90	**5387.52**	5394.10	5393.36	**5387.16**	5387.64	5387.81	**5386.24**	5345.51	5348.09	**5342.65**
In-S. MSE*	**(0)**†	5.0890	5.0890	**5.0890**	5.0890	5.0890	**5.0890**	5.1876	5.1876	**5.1876**			
	(I)	5.1527	5.1421	**5.0902**	5.1541	5.1432	**5.0903**	5.4962	5.4112	**5.1686**			
	(II)	5.1061	5.1035	**5.0824**	5.1062	5.1035	**5.0824**	5.2764	5.2475	**5.1309**	5.0889	5.0981	**5.0777**
	(III)	5.0627	5.0621	**5.0583**	5.0627	5.0622	**5.0584**	5.1886	5.1722	**5.0999**	**5.0387**	5.0467	5.0401
Out-of-S. MSE*	**(0)**†	5.0890	5.0890	**5.0890**	5.0890	5.0890	**5.0890**	5.1876	5.1876	**5.1876**			
	(I)	4.9730	4.9639	**4.9203**	4.9743	4.9648	**4.9204**	5.3007	5.2216	**4.9989**			
	(II)	4.9319	4.9293	**4.9116**	4.9320	4.9294	**4.9117**	5.1154	5.0858	**4.9636**	4.9203	4.9259	**4.9058**
	(III)	4.8956	4.8946	**4.8899**	4.8956	4.8947	**4.8900**	5.0371	5.0184	**4.9370**	4.8852	4.8902	**4.8801**

† For model (0), exposure measures (Km, NOx or CO) are not included, therefore the performance measures remain constant.

*multiplied by 100.

The results in Table 7 show that when comparing exposure measures, all three indicators yield similar results. However, CO is generally the most effective indicator (results highlighted in bold), albeit not significantly improving model performance. While NOx outperforms kilometers in models (**I**) and (**II**). In model (**III**), which includes all telematics covariates, kilometers, and NOx become extremely comparable both in terms of the in-sample and out-of-sample MSEs.

The choice of distribution produces mixed outcomes with regard to model performance depending on the evaluation criteria. Specifically, the ZIP distribution provides better results in terms of the AIC but its out-of-sample MSE is significantly higher than those of its two counterparts, hinting at potential overfitting. Yet, overall, the preferred model is the CANN since it provides the best results across all three statistical measures. Moreover, all the models benefit from including more information in the training process, as models (**III**) outperform the rest.

All the aforementioned analyses are based on Spain's 2017 emission estimates. However, to ensure the robustness of these findings, a sensitivity analysis was conducted by alternating the geographical emission profiles, specifically utilizing the 2018 Spanish emission factors and the 2017 UK factors (see 5). This cross-year, cross-country swap allows us to assess the stability of our analyses. The Appendix showcases Tables 13 and 14, which contain the results under these hypotheses. We note that our comments would remain unchanged, signaling the robustness of our results.

We conclude that employing certain pollutants – particularly CO – as exposure measures can provide comparable quality of models estimating road accident risk, as those that employ traditional exposure measures, such as distance driven. Therefore, overall, our findings suggest that emission-based measures can serve as viable alternatives for the modeling of road accident risk.

4.3.4 *Out-of-sample Simulation Study*

Finally, we conduct an out-of-sample simulation study to evaluate the performance of different exposure measures in predicting road accidents. To do so, using models trained on 2017 data, we simulate total claims counts in 2018 across 100,000 iterations. Specifically, we consider two distinct frameworks: (Model (**I**)), which includes only the exposure measures (as offsets) and excludes other telematics covariates, and (Model (**III**)), which incorporates all telematics variables, including the exposure measures, as covariates. For this analysis, we select the negative binomial distribution, as an example. Table 8 summarizes the key results, which are further illustrated for models (**I**) and (**III**) in Figures 6 and 7, respectively.

Table 8: Out-of-sample distribution of total claim counts in 2018 for negative binomial models

Model Tag	Dist	k	Mean	SD	Quantile								
					1%	5%	10%	25%	50%	75%	90%	95%	99%
(**I**)	NB	KM	779.91	28.69	714	733	743	760	780	799	817	827	848
		NOx	778.94	28.67	713	732	742	760	779	798	816	827	846
		CO	776.33	28.30	711	730	740	757	776	795	813	823	843
(**III**)	NB	KM	791.56	28.41	726	745	755	772	791	811	828	838	858
		NOx	791.32	28.38	726	745	755	772	791	810	828	838	859
		CO	788.49	28.27	724	742	752	769	788	808	825	835	855
Observed			843										

The simulation results presented in Table 8 suggest that all models underpredict the total claim counts. At the 95% Value-at-Risk (VaR), only one of the six models yields a VaR equal to or higher than the observed claim count. When comparing models of the same type (i.e., Models (**I**) or (**III**)), the choice of exposure measure has minimal impact on the total claim count distribution: the mean, standard deviation, and quantiles are nearly identical, resulting in similar distributions regardless of the exposure measure used (see Figures 6 and 7).

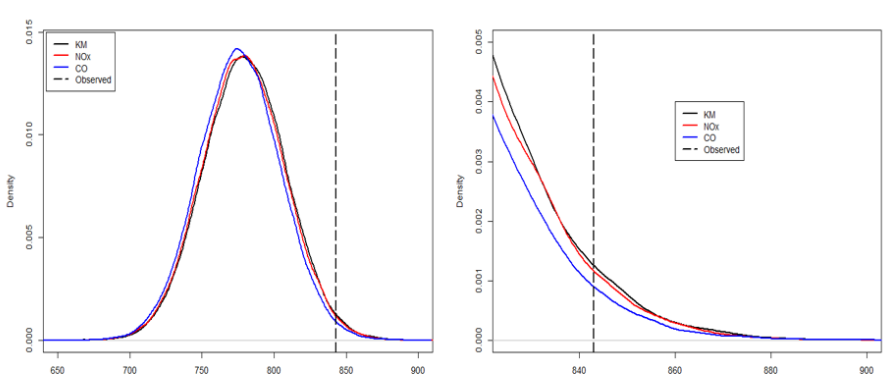

Figure 6: Negative Binomial model (**I**): Distribution of the total claim counts in 2018

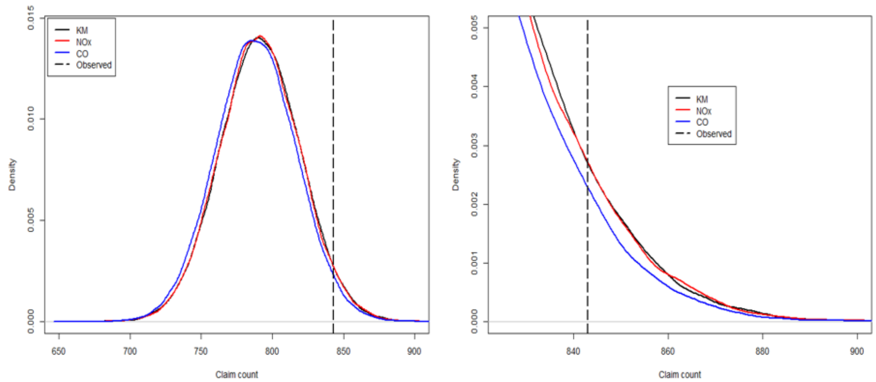

Figure 7: Negative Binomial model (**III**): Distribution of the total claim counts in 2018

Based on this analysis, we conclude that for large datasets of drivers and claims, road accident risk can be effectively estimated using models based on either traditional distance-driven exposure measures or pollutant-based measures, without compromising predictive quality. Moreover, models that incorporate pollution-based exposure measures offer a unique opportunity to promote road safety by exploiting environmental arguments to influence driver behavior.

5 Discussion

This paper introduces emissions as an alternative exposure measure in individual-level accident risk modeling within an actuarial and telematics-based framework. Unlike traditional measures such as distance or driving time, emissions have not previously been explicitly used in this context. Our contribution is therefore primarily conceptual: we propose a reinterpretation of exposure that integrates environmental externalities with established risk modeling practices, rather than introducing new estimation techniques.

A key insight of our approach is that emissions should not be viewed merely as a transformation of distance. Although they are constructed from distance and road type, emissions embed additional information related to driving context, such as congestion, stop-and-go traffic, and road environment. As a result, they capture aspects of exposure quality that are not fully reflected in standard measures. Two drivers covering identical distances may generate different emissions due to differences in location and traffic conditions, highlighting the added informational content of this metric.

At the same time, we acknowledge that many relevant driver, vehicle, and behavioral characteristics are already included in the model through telematics covariates. Emissions are not intended to proxy for omitted heterogeneity, but rather to complement these variables by providing a more integrated measure of exposure. Even when controlling for rich telematics information, our results indicate that emissions retain some explanatory power, albeit with modest incremental predictive gains.

Importantly, the value of emissions should not be assessed solely in terms of predictive performance. The empirical improvements over traditional exposure measures are moderate, emissions offer a conceptually distinct perspective that aligns risk modeling with broader societal concerns, particularly environmental sustainability. In this sense, emissions extend the scope of actuarial modeling beyond purely financial risk to incorporate externalities that are increasingly relevant in regulatory and market contexts.

From a practical standpoint, emissions also offer advantages in terms of data aggregation and privacy. When information on distance, location, and driving conditions is consolidated into a single metric, emissions may reduce the need to collect and process highly granular telematics data, such as detailed location tracking. This has the potential to make risk modeling less intrusive while preserving key dimensions of exposure, an aspect that may be particularly valuable in settings with privacy constraints.

We also recognize limitations related to measurement in our case study. Emissions are not directly observed but estimated using standardized factors, which may introduce noise and reduce granularity. However, such measurement error is likely to attenuate estimated effects, suggesting that our findings are conservative. Moreover, existing evidence supports the use of standardized emissions models across similar contexts, and technological advances are expected to further improve measurement accuracy and availability over time.

From a behavioral perspective, the willingness to mitigate air pollution is intrinsically linked to psychological distance—the perceived gap between the immediate 'here and now' and distant environmental consequences, as discussed by Cheng *et al.* (2022) and Mir *et al.* (2016). By linking a driver's environmental footprint to their insurance premiums through exposure metrics, this psychological distance can be narrowed, making abstract environmental harm feel more tangible and immediate. This presents opportunities to design incentives that foster sustainable behavior; as noted by Riley *et al.* (2021) and Helbig *et al.* (2021), communicating pollution data to drivers via smartphones or in-vehicle devices enables flexible, targeted engagement, even though Haddad & de Nazelle (2018) have shown that relying exclusively on communication has limited impact on behavioral change. In this regard, the disclosure of emission levels is reinforced by monetary incentives in the form of premium reductions. This allows telematics data to influence driver behavior in a manner analogous to road pricing (Gibson & Carnovale , 2015).

Finally, regarding implications for insurance markets, we emphasize that emissions are not intended to replace existing telematics-based pricing models but to complement them. Our results do not show large changes in premiums or loss ratios. Hence, emissions-based exposure can support the development of new contract features, such as sustainability-linked pricing or other green projects, as insurance companies increasingly invest in environmentally friendly initiatives (Khovrak , 2020). Labini *et al.* (2025) show that American insurance companies are characterized by quality governance that is committed to environmental, social and governance sustainability policies which is linked to good financial performance (see also Khovrak , 2020). These new eco-designed products applications are forward-looking but align with ongoing trends in regulation and consumer preferences, suggesting that the relevance of emissions in actuarial practice may increase over time.

In essence, emissions-based exposure can support the design of pricing and underwriting schemes that incorporate environmental considerations alongside risk. Even in cases where predictive gains are modest, such measures can enable insurers to differentiate contracts, introduce sustainability-linked incentives, and align pricing structures with emerging regulatory and societal priorities.

6 Conclusion

In this paper, we have demonstrated the feasibility of assessing road crash risk by incorporating pollutant emissions as exposure measures, thus replacing traditional metrics such as observation periods and distance driven. We evaluated this approach using various statistical frameworks and three distributions: Poisson, negative binomial, and zero-inflated Poisson, to assess the impact of using CO and NOx emissions as exposure measures.We show that CO and NOx emissions provided comparable predictive power to distance driven (as measured by AIC and MSE). The slight edge provided by CO be attributed to the fact that more CO is emitted in urban areas, where traffic accidents are more frequent.

Indeed, emissions are mechanically related to distance, yet they are not a simple transformation of it. Emissions depend on a combination of speed profiles, acceleration patterns, congestion levels, and road context. This implies that two drivers with identical distance exposure—and even similar telematics summaries—may still generate different emissions due to differences in driving environments (e.g., urban versus highway in our case). In this sense, emissions capture aspects of exposure quality that are not fully spanned by distance or standard telematics variables alone. Our empirical analysis, although showing moderate gains, confirms that emissions retain explanatory content even after controlling for these factors.

To further illustrate our methodology, we simulated annual crash counts for 17,405 drivers, testing models with distinct covariate structures. By comparing the effect of different exposure measures on the total claim count distribution, we found that for a large number of vehicles, as in our case study, the distribution of road crash counts remains virtually indistinguishable whether traditional mileage metrics or pollutant-based exposure measures are used.

While our macroscopic approach is well-suited to large groups of vehicles, its reliance on aggregated telematics data limits its level of detail. Future research could usefully integrate microscopic data about specific trip characteristics and individual driver behavior. Combining extensive datasets, as in our study, with more granular telematics data could significantly enhance the precision and applicability of this modeling framework. Such an extension would allow for the development of risk and emission measures specific to individual drivers and trips, offering new insights into both accident and environmental risk.

Our study introduces a novel modeling approach to road crash risk assessment that integrates environmental considerations. By providing evidence that reducing pollutant emissions may also mitigate accident risk, we highlight a dual benefit for public health and safety. This approach can be used to transform the way classical insurance rates consider exposure to risk, and ease the introduction of green taxes that penalize high-emission driving. Moreover, it can help drivers minimize their environmental footprint while contributing to accident prevention and benefiting society as a whole.

Acknowledgments.

Competing interests.

References

Abdel-Aty, M., Uddin, N., Pande, A., Abdalla, M. F., & Hsia, L. (2004). Predicting freeway crashes from loop detector data by matched case-control logistic regression. *Transportation Research Record, 1897*(1), 88-95.

Ahn, K., Rakha, H., Trani, A., & Van Aerde, M. (2002). Estimating vehicle fuel consumption and emissions based on instantaneous speed and acceleration levels. *Journal of Transportation Engineering, 128*(2), 182-190.

Anenberg, S.C., Belova, A., Brandt, J., Fann, N., Greco, S., Guttikunda, S., Heroux, M.-E., Hurley, F., Krzyzanowski, M., Medina, S., Miller, B., Pandey, K., Roos, J. & Van Dingenen, R. (2016). Survey of ambient air pollution health risk assessment tools. *Risk analysis, 36*(9), 1718-1736.

André, M. (2004). The ARTEMIS European driving cycles for measuring car pollutant emissions. *Science of the Total Environment, 334*, 73-84.

Ayuso, M., Guillen, M. & Pérez-Marín, A. M. (2016). Using GPS data to analyse the distance travelled to the first accident at fault in pay-as-you-drive insurance. *Transportation Research Part C: Emerging Technologies, 68*, 160-167.

Ayuso, M., Guillen, M. & Pérez-Marín, A. M. (2014). Time and distance to first accident and driving patterns of young drivers with pay-as-you-drive insurance. *Accident Analysis & Prevention, 73*, 125-131.

Bauer, D., Schmit, J., & Sydnor, J. (2024). The economics of emerging insurance technologies: Theory and early evidence. *Journal of Risk and Insurance, 91*(4), 809-812.

Becker, N., Rust, H. W., & Ulbrich, U. (2022). Weather impacts on various types of road crashes: a quantitative analysis using generalized additive models. *European Transport Research Review, 14*(1), 37.

Boucher, J. P., Côté, S. & Guillen, M. (2017). Exposure as duration and distance in telematics motor insurance using generalized additive models. *Risks, 5*(4), 54.

Chacón, J. E., & Duong, T. (2018). *Multivariate kernel smoothing and its applications*. Chapman and Hall/CRC.

Cheng, Y., Ao, C., Mao, B., & Xu, L. (2022). Influential factors of environmental behavior to reduce air pollution: integrating theories of planned behavior and psychological distance. *Journal of Environmental Planning and Management*, *65*(13), 2490–2510.

Chipman, M. L., MacGregor, C. G., Smiley, A. M., & Lee-Gosselin, M. (1992). Time vs. distance as measures of exposure in driving surveys. *Accident Analysis & Prevention, 24*(6), 679-684.

Chipman, M. L., MacGregor, C. G., Smiley, A. M., & Lee-Gosselin, M. (1993). The role of exposure in comparisons of crash risk among different drivers and driving environments. *Accident Analysis & Prevention, 25*(2), 207-211.

Department for Transport (2019). *Road traffic estimates: Great Britain 2018.* https://www.gov.uk/government/statistics/road-traffic-estimates-in-great-britain-2018.

Department for Transport (2018). *Road traffic estimates: Great Britain 2017.* https://www.gov.uk/government/statistics/road-traffic-estimates-in-great-britain-2017.

Davison, J., Rose, R. A., Farren, N. J., Wagner, R. L., Murrells, T. P., & Carslaw, D. C. (2021). Verification of a national emission inventory and influence of on-road vehicle manufacturer-level emissions. *Environmental Science & Technology, 55*(8), 4452-4461.

De Haan, P., & Keller, M. (2004). Modelling fuel consumption and pollutant emissions based on real-world driving patterns: the HBEFA approach. *International Journal of Environment and Pollution, 22*(3), 240-258.

Duarte, G. O., Gonçalves, G. A., & Farias, T. L. (2016). Analysis of fuel consumption and pollutant emissions of regulated and alternative driving cycles based on real-world measurements. *Transportation Research Part D: Transport and Environment, 44*, 43-54.

Duval, F., Boucher, J. P., & Pigeon, M. (2023). Enhancing claim classification with feature extraction from anomaly-detection-derived routine and peculiarity profiles. *Journal of Risk and Insurance, 90*(2), 421-458.

Duval, F., Boucher, J. P., & Pigeon, M. (2024). Telematics combined actuarial neural networks for cross-sectional and longitudinal claim count data. *ASTIN Bulletin: The Journal of the IAA, 54*(2), 239-262.

El-Basyouny, K., Barua, S., & Islam, M. T. (2014). Investigation of time and weather effects on crash types using full Bayesian multivariate Poisson lognormal models. *Accident Analysis & Prevention, 73*, 91-99.

European Commission (2024). *Annual statistical report on road safety in the EU 2024.* European Commission. https://road-safety.transport.ec.europa.eu/european-road-safety-observatory/data-and-analysis/annual-statistical-report_en

European Environmental Agency (2024). *New registrations of electric vehicles in Europe.* European Environmental Agency. https://www.eea.europa.eu/en/analysis/indicators/new-registrations-of-electric-vehicles/new-registration-of-electric-cars-eu-27

European Environmental Agency (2024). *EEA greenhouse gases — data viewer.* European Environmental Agency. https://www.eea.europa.eu/en/analysis/maps-and-charts/greenhouse-gases-viewer-data-viewers

European Environmental Agency (2021). *Air quality in Europe 2021 Europe's air quality status 2021- update.* European Environmental Agency. https://www.eea.europa.eu/publications/air-quality-in-europe-2021/air-quality-status-briefing-2021

Fafoutellis, P., Mantouka, E. G., & Vlahogianni, E. I. (2020). Eco-driving and its impacts on fuel efficiency: An overview of technologies and data-driven methods. *Sustainability, 13*(1), 226.

Fomunung, I., Washington, S., Guensler, R., & Bachman, W. (2000). Validation of the MEASURE automobile emissions model: a statistical analysis. *Journal of Transportation and Statistics, 3*, 65-84.

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), 1-20.

Gao, G. & Wüthrich, M. V. (2019). Convolutional neural network classification of telematics car driving data. *Risks, 7*(1), 6.

Gatzert, N., Reichel, P., & Zitzmann, A. (2020). Sustainability risks & opportunities in the insurance industry. *Zeitschrift für die gesamte Versicherungswissenschaft*, *109*(5), 311–331.

Gibson, M., & Carnovale, M. (2015). The effects of road pricing on driver behavior and air pollution. *Journal of Urban Economics*, *89*, 62–73.

Guillen, M., Nielsen, J. P., Ayuso, M., & Pérez-Marín, A. M. (2019). The use of telematics devices to improve automobile insurance rates. *Risk analysis, 39*(3), 662-672.

Guillen, M., Nielsen, J. P., & Pérez-Marín, A. M. (2021). Near-miss telematics in motor insurance. *Journal of Risk and Insurance, 88*(3), 569-589.

Guo, M., Zhao, X., Yao, Y., Yan, P., Su, Y., Bi, C., & Wu, D. (2021). A study of freeway crash risk prediction and interpretation based on risky driving behavior and traffic flow data. *Accident Analysis & Prevention, 160*, 106328.

Haddad, H., & de Nazelle, A. (2018). The role of personal air pollution sensors and smartphone technology in changing travel behaviour. *Journal of Transport & Health*, *11*, 230–243.

Hausberger, S., Rexeis, M., & Luz, R. (2013). PHEM Passenger car and Heavy duty Emissions Model User guide for Version 11, Graz University of Technology. *Institute for Internal Combustion Engines and Thermodynamic.*

Helbig, C., Ueberham, M., Becker, A. M., Marquart, H., & Schlink, U. (2021). Wearable sensors for human environmental exposure in urban settings. *Current Pollution Reports*, *7*(3), 417–433.

Holzapfel, J., Peter, R., & Richter, A. (2024). Mitigating moral hazard with usage-based insurance. *Journal of Risk and Insurance, 91*(4), 813-839.

Huang, Y. & Meng, S. (2019). Automobile insurance classification ratemaking based on telematics driving data. *Decision Support Systems, 127*, 113156.

Huboyo, H. S., Handayani, W., & Samadikun, B. P. (2017, June). Potential air pollutant emission from private vehicles based on vehicle route. In *IOP Conference Series: Earth and Environmental Science* (Vol. 70, No. 1, p. 012013). IOP Publishing.

Khovrak, I. (2020). ESG-driven approach to managing insurance companies' sustainable development. *Insurance Markets and Companies*, *11*(1), 42–52.

Labini, S. S., di Biase, P., & D'Apolito, E. (2025). Sustainability strategy and financial performance in the insurance company. *International Review of Economics & Finance*, *98*, 103924.

Liu, H., Chen, X., Wang, Y., & Han, S. (2013). Vehicle emission and near-road air quality modeling for shanghai, china: Based on global positioning system data from taxis and revised moves emission inventory. *Transportation Research Record, 2340*(1), 38-48.

Ma, Y. L., Zhu, X., Hu, X. & Chiu, Y. C. (2018). The use of context-sensitive insurance telematics data in auto insurance ratemaking. *Transportation Research Part A: Policy and Practice, 113*, 243-258.

Malin, F., Norros, I., & Innamaa, S. (2019). Accident risk of road and weather conditions on different road types. *Accident Analysis & Prevention, 122*, 181-188.

McDonnell, K., Sheehan, B., Murphy, F., & Guillen, M. (2024). Are electric vehicles riskier? A comparative study of driving behaviour and insurance claims for internal combustion engine, hybrid and electric vehicles. *Accident Analysis & Prevention, 207*, 107761.

Meseguer, J. E., Toh, C. K., Calafate, C. T., Cano, J. C., & Manzoni, P. (2017). Drivingstyles: A mobile platform for driving styles and fuel consumption characterization. *Journal of Communications and Networks, 19*(2), 162-168.

Mir, H. M., Behrang, K., Isaai, M. T., & Nejat, P. (2016). The impact of outcome framing and psychological distance of air pollution consequences on transportation mode choice. *Transportation Research Part D: Transport and Environment*, *46*, 328–338.

Ministerio para la Transición Ecológica y el Reto Demográfico. (2020). *Metodologías de estimación de emisiones de contaminantes a la atmósfera de origen móvil: Transporte por carretera.* `https://www.miteco.gob.es/content/dam/miteco/es/calidad-y-evaluacion-ambiental/temas/sistema-espanol-de-inventario-sei-/07_tpte_carretera_combustion.pdf`.

Naik, B., Tung, L. W., Zhao, S., & Khattak, A. J. (2016). Weather impacts on single-vehicle truck crash injury severity. *Journal of Safety Research, 58*, 57-65.

Negrenti, E. (1996). TEE: The ENEA traffic emissions and energetics model micro-scale applications. *Science of the Total Environment, 189*, 167-174.

Nejadkoorki, F., Nicholson, K., Lake, I., & Davies, T. (2008). An approach for modelling CO2 emissions from road traffic in urban areas. *Science of the Total Environment, 406*(1-2), 269-278.

Nocera, S., Basso, M., & Cavallaro, F. (2017). Micro and Macro modelling approaches for the evaluation of the carbon impacts of transportation. *Transportation Research Procedia, 24*, 146-154.

Ntziachristos, L., Gkatzoflias, D., Kouridis, C., & Samaras, Z. (2009). COPERT: a European road transport emission inventory model. In *Information Technologies in Environmental Engineering: Proceedings of the 4th International ICSC Symposium Thessaloniki, Greece, May 28-29, 2009* (pp. 491-504). Springer Berlin Heidelberg.

Ohlsson, E., & Johansson, B. (2010). *Non-life insurance pricing with generalized linear models* (Vol. 174). Berlin: Springer.

Pawar, N. M., Khanuja, R. K., Choudhary, P.,& Velaga, N. R. (2020). Modelling braking behaviour and accident probability of drivers under increasing time pressure conditions. *Accident Analysis & Prevention, 136*, 105401.

Pei, X., Wong, S. C., & Sze, N. N. (2012). The roles of exposure and speed in road safety analysis. *Accident Analysis & Prevention, 48*, 464-471.

Pirayre, A., Michel, P., Rodriguez, S. S., & Chasse, A. (2022). Driving behavior identification and real-world fuel consumption estimation with crowdsensing data. *IEEE Transactions on Intelligent Transportation Systems, 23*(10), 18378-18391.

Qin, X., Ivan, J. N., Ravishanker, N., Liu, J., & Tepas, D. (2006). Bayesian estimation of hourly exposure functions by crash type and time of day. *Accident Analysis & Prevention, 38*(6), 1071-1080.

Quddus, M. A., Noland, R. B. & Chin, H. C. (2002). An analysis of motorcycle injury and vehicle damage severity using ordered probit models. *Journal of Safety Research, 33*(4), 445-462.

Reiter, M. S., & Kockelman, K. M. (2016). The problem of cold starts: A closer look at mobile source emissions levels. *Transportation Research Part D: Transport and Environment, 43*, 123-132.

Rakha, H., Ahn, K., & Trani, A. (2004). Development of VT-Micro model for estimating hot stabilized light duty vehicle and truck emissions. *Transportation Research Part D: Transport and Environment, 9*(1), 49-74.

Retallack, A. E., & Ostendorf, B. (2019). Current understanding of the effects of congestion on traffic accidents. *International Journal of Environmental Research and Public Health, 16*(18), 3400.

Riley, R., de Preux, L., Capella, P., Mejia, C., Kajikawa, Y., & de Nazelle, A. (2021). How do we effectively communicate air pollution to change public attitudes and behaviours? A review. *Sustainability Science*, *16*(6), 2027–2047.

Rodriguez, S. S., Alix, G., Cheimariotis, I., & Dégeilh, P. (2023). Monitoring the real-driving emissions considering the fleet and the driving behavior by geographical area. *Transportation Research Procedia, 72*, 2692-2699.

Scora, G., & Barth, M. (2006). Comprehensive modal emissions model (cmem), version 3.01. User guide. *Centre for Environmental Research and Technology. University of California, Riverside, 1070*, 79.

Sentoff, K. M., Aultman-Hall, L., & Holmén, B. A. (2015). Implications of driving style and road grade for accurate vehicle activity data and emissions estimates. *Transportation Research Part D: Transport and Environment, 35*, 175-188.

Segui-Gomez, M., Lopez-Valdes, F. J., Guillen-Grima, F., Smyth, E., Llorca, J., & de Irala, J. (2011). Exposure to traffic and risk of hospitalization due to injuries. *Risk Analysis: An International Journal, 31*(3), 466-474.

Shen, S., Benedetti, M. H., Zhao, S., Wei, L., & Zhu, M. (2020). Comparing distance and time as driving exposure measures to evaluate fatal crash risk ratios. *Accident Analysis & Prevention, 142*, 105576.

Smit, R., Ntziachristos, L., & Boulter, P. (2010). Validation of road vehicle and traffic emission models–A review and meta-analysis. *Atmospheric Environment, 44*(25), 2943-2953.

Smit, R., Brown, A. L., & Chan, Y. C. (2008a). Do air pollution emissions and fuel consumption models for roadways include the effects of congestion in the roadway traffic flow?. *Environmental Modelling & Software, 23*(10-11), 1262-1270.

Smit, R., Poelman, M., & Schrijver, J. (2008). Improved road traffic emission inventories by adding mean speed distributions. *Atmospheric Environment, 42*(5), 916-926.

Smit, R., Smokers, R., & Rabé, E. (2007). A new modelling approach for road traffic emissions: VERSIT+. *Transportation Research Part D: Transport and Environment, 12*(6), 414-422.

Soleymanian, M., Weinberg, C. B., & Zhu, T. (2025). Insurtech, sensor data, and changes in customers' coverage choices: Evidence from usage-based automobile insurance. *Journal of Risk and Insurance, 92*(1), 227-256.

Steg, L., & van Brussel, A. (2009). Accidents, aberrant behaviours, and speeding of young moped riders. *Transportation Research Part F: Traffic Psychology and Behaviour, 12*(6), 503-511.

Stipancic, J., Miranda-Moreno, L. & Saunier, N. (2018). Vehicle manoeuvers as surrogate safety measures: Extracting data from the gps-enabled smartphones of regular drivers. *Accident Analysis & Prevention, 115*, 160-169.

Tang, Y., Xiang, L., & Zhu, Z. (2014). Risk factor selection in rate making: EM adaptive LASSO for zero-inflated Poisson regression models. *Risk Analysis, 34*(6), 1112-1127.

Thibault, L., Degeilh, P., Lepreux, O., Voise, L., Alix, G., & Corde, G. (2016). A new GPS-based method to estimate real driving emissions. *In 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)* (pp. 1628-1633). IEEE.

Van den Bossche, F., Wets, G., & Brijs, T. (2005). Role of exposure in analysis of road accidents: a Belgian case study. *Transportation research record, 1908*(1), 96-103.

Wang, C., Quddus, M. A., & Ison, S. G. (2013). The effect of traffic and road characteristics on road safety: A review and future research direction. *Safety Science, 57*, 264-275.

Wang, X., Yue, X., Huang, J., & Li, S. (2025). Integrating Traffic Dynamics and Emissions Modeling: From Classical Approaches to Data-Driven Futures. *Atmosphere, 16*(6), 695.

World Health Organization (2023). *Global status report on road safety 2023*. World Health Organization. https://www.who.int/teams/social-determinants-of-health/safety-and-mobility/global-status-report-on-road-safety-2023

World Health Organization (2021). *What are the WHO Air quality guidelines?*. World Health Organization. https://www.who.int/news-room/feature-stories/detail/what-are-the-who-air-quality-guidelines

Wüthrich, M. V., & Merz, M. (2019). Yes, we CANN!. *ASTIN Bulletin: The Journal of the IAA, 49*(1), 1-3.

Wüthrich, M. V. (2017). Covariate selection from telematics car driving data. *European Actuarial Journal, 7*(1), 89–108.

Yao, Y., Zhao, X., Liu, C., Rong, J., Zhang, Y., Dong, Z., & Su, Y. (2020). Vehicle fuel consumption prediction method based on driving behavior data collected from smartphones. *Journal of Advanced Transportation, 2020*(1), 9263605.

Yu, R., & Abdel-Aty, M. (2013). Multi-level Bayesian analyses for single-and multi-vehicle freeway crashes. *Accident Analysis & Prevention, 58*, 97-105.

7 Appendix: Additional Tables

Table 9: Vehicle Distance Estimates (Millions of KM) and Percentages

		Year	2017			2018		
Country	Measure	Type	Petrol	Diesel	Total	Petrol	Diesel	Total
Spain	Total KM*	Urban	24,974	92,635	117,609	26,091	93,969	120,060
		Rural	11,542	43,383	54,925	12,008	43,826	55,834
		Motorway	26,876	116,941	143,817	28,277	119,464	147,740
		Total	63,392	252,958	316,350	66,376	257,258	323,634
	%KM	Urban	0.07894	0.29282	0.37177	0.08062	0.29036	0.37097
		Rural	0.03649	0.13714	0.17362	0.03711	0.13542	0.17252
		Motorway	0.08496	0.36966	0.45461	0.08737	0.36913	0.45650
		Total	0.20039	0.79961	1	0.20510	0.79491	1
UK	Total KM*	Urban	83,343	64,530	147,873	85,769	62,233	148,003
		Rural	90,913	92,519	183,432	91,521	93,169	184,690
		Motorway	30,375	47,738	78,112	29,899	47,791	77,690
		Total	204,631	204,787	409,417	207,190	203,193	410,383
	%KM	Urban	0.20357	0.15761	0.36118	0.20900	0.15165	0.36065
		Rural	0.22205	0.22598	0.44803	0.22301	0.22703	0.45004
		Motorway	0.07419	0.11660	0.19079	0.07286	0.11645	0.18931
		Total	0.49981	0.50019	1	0.50487	0.49513	1

* Values expressed in millions of kilometers (Mkm).

Table 10: Descriptive statistics for the risk exposure measures (mean yearly exposure by driver)

Descriptive	Risk exposure measure		
Statistics	Kilometres	NOx*	CO*
Mean	10233	4575	2680
1st Quartile	5473	2504	1603
Median	9103	4095	2479
2nd Quartile	13692	6103	3514
Standard Deviation	6463	2818	1501

* in grams

Table 11: Poisson model with static covariates, telematics covariates, and exposure measures as covariates

	Risk exposure measures							
	Years		Km		NOx		CO	
	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value
(μ intercept)	-3.312	< 0.001	-5.388	< 0.001	-9.943	< 0.001	-9.460	< 0.001
Driver Age	-0.018	0.031	-0.014	0.089	-0.014	0.087	-0.014	0.084
Gender Male	-0.015	0.839	-0.059	0.415	-0.060	0.408	-0.064	0.380
Vehicle Power	0.003	0.002	0.003	0.017	0.003	0.017	0.003	0.020
Km Night	0.018	< 0.001	0.013	0.001	0.013	0.001	0.012	0.002
Km Overspeed	-0.002	0.665	-0.008	0.145	-0.009	0.134	-0.008	0.153
Km Urban	0.010	< 0.001	0.026	< 0.001	0.024	< 0.001	0.014	< 0.001
$\log\left(T_{i,j}^{(k)}\right)$			0.751	< 0.001	0.751	< 0.001	0.769	< 0.001

Table 12: Zero-Inflated Poisson model with static covariates, telematics covariates, and exposure measures as covariates

	Risk exposure measures							
	Years		Km		NOx		CO	
	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value
(μ intercept)	-2.938	< 0.001	-4.159	< 0.001	-6.432	< 0.001	-6.937	< 0.001
Driver Age	-0.017	0.035	-0.014	0.094	-0.014	0.091	-0.014	0.084
Gender Male	-0.015	0.842	-0.053	0.475	-0.053	0.469	-0.058	0.432
Vehicle Power	0.003	0.002	0.003	0.016	0.003	0.017	0.003	0.019
Km Night	0.018	< 0.001	0.013	0.002	0.013	0.002	0.013	0.002
Km Overspeed	-0.002	0.672	-0.008	0.187	-0.008	0.178	-0.008	0.198
Km Urban	0.010	< 0.001	0.025	< 0.001	0.024	< 0.001	0.014	< 0.001
$\log\left(T_{i,j}^{(k)}\right)$			0.372	0.010	0.377	0.011	0.485	0.002
(σ intercept)	0.156	< 0.001	1.676	0.012	8.948	< 0.001	7.685	0.007
$\log\left(T_{i,j}^{(k)}\right)$			-1.182	< 0.001	-1.186	< 0.001	-1.131	0.006

Table 13: Akaike information criterion (AIC) and mean squared error (MSE) (in-sample and out-of-sample) for selected models with various exposure measures. Bold figures indicate the best exposure measure for each model. These results use Spain's emission factors from 2018.

		Poisson			Negative Binomial			Zero-Inflated Poisson			CANN		
Criteria	Model	Km	NOx	CO	Km	NOx	CO	Km	NOx	CO	Km	NOx	CO
AIC	**(0)**†	5552.28	5552.28	5552.28	5547.71	5547.71	5547.71	5548.88	5548.88	5548.88			
	(I)	5585.30	5556.84	**5455.36**	5577.84	5550.61	**5452.96**	5496.15	5485.63	**5439.00**			
	(II)	5492.67	5483.47	**5436.27**	5489.48	5480.49	**5434.49**	5479.20	5471.87	**5433.87**	5436.82	5463.62	**5418.51**
	(III)	5394.66	5393.79	**5388.46**	5394.10	5393.26	**5388.07**	5387.64	5387.82	**5386.80**	5345.51	5345.86	**5341.58**
In-S. MSE*	**(0)**†	5.1227	5.1227	5.1227	5.1228	5.1228	5.1228	5.2030	5.2030	5.2030			
	(I)	5.1527	5.1411	**5.0949**	5.1541	5.1421	**5.0951**	5.4962	5.4041	**5.1846**			
	(II)	5.1061	5.1032	**5.0852**	5.1062	5.1033	**5.0853**	5.2764	5.2444	**5.1388**	5.0889	5.0976	**5.0805**
	(III)	5.0627	5.0621	**5.0588**	5.0627	5.0621	**5.0588**	5.1886	5.1703	**5.1049**	5.0387	5.0406	**5.0378**
Out-of-S. MSE*	**(0)**†	4.9539	4.9539	4.9539	4.9540	4.9540	4.9540	5.1876	5.1876	5.1876			
	(I)	4.9730	4.9630	**4.9241**	4.9743	4.9639	**4.9242**	5.3007	5.2149	**5.0132**			
	(II)	4.9319	4.9291	**4.9138**	4.9320	4.9291	**4.9138**	5.1154	5.0827	**4.9716**	4.9203	4.9269	**4.9110**
	(III)	4.8956	4.8946	**4.8904**	4.8956	4.8946	**4.8905**	5.0371	5.0162	**4.9426**	4.8852	4.8876	**4.8791**

† For model (0), exposure measures (Km, NOx, or CO) are not included; the performance measures remain constant.

*multiplied by 100.

Table 14: Akaike information criterion (AIC) and mean squared error (MSE) (in-sample and out-of-sample) for selected models with various exposure measures. Bold figures indicate the best exposure measure for each model. These results use the UK's emission factors from 2017.

		Poisson			Negative Binomial			Zero-Inflated Poisson			CANN		
Criteria	Model	Km	NOx	CO	Km	NOx	CO	Km	NOx	CO	Km	NOx	CO
AIC	**(0)**†	5552.28	5552.28	5552.28	5547.71	5547.71	5547.71	5548.88	5548.88	5548.88			
	(I)	5585.30	5577.26	**5482.14**	5577.84	5570.10	**5478.86**	5496.15	5492.60	**5454.58**			
	(II)	5492.67	5489.60	**5452.15**	5489.48	5486.48	**5449.96**	5479.20	5476.59	**5447.50**	5436.82	5472.40	**5433.68**
	(III)	5394.66	5395.20	**5390.19**	5394.10	5394.64	**5389.75**	5387.64	5388.38	**5387.60**	5345.51	5347.61	**5343.49**
In-S. MSE*	**(0)**†	5.1227	5.1227	5.1227	5.1228	5.1228	5.1228	5.2030	5.2030	5.2030			
	(I)	5.1527	5.1500	**5.1074**	5.1541	5.1513	**5.1076**	5.4962	5.4682	**5.2296**			
	(II)	5.1061	5.1052	**5.0919**	5.1062	5.1052	**5.0919**	5.2764	5.2663	**5.1625**	5.0889	5.1008	**5.0870**
	(III)	5.0627	5.0627	**5.0599**	5.0627	5.0627	**5.0600**	5.1886	5.1838	**5.1191**	**5.0387**	5.0457	5.0425
Out-of-S. MSE*	**(0)**†	4.9539	4.9539	4.9539	4.9540	4.9540	4.9540	5.1876	5.1876	5.1876			
	(I)	4.9730	4.9718	**4.9341**	4.9743	4.9730	**4.9344**	5.3007	5.2783	**5.0535**			
	(II)	4.9319	4.9312	**4.9190**	4.9320	4.9313	**4.9190**	5.1154	5.1050	**4.9965**	4.9203	4.9292	**4.9128**
	(III)	4.8956	4.8958	**4.8916**	4.8956	4.8959	**4.8917**	5.0371	5.0315	**4.9589**	4.8852	4.8902	**4.8825**

† For model (0), exposure measures (Km, NOx, or CO) are not included; the performance measures remain constant.

*multiplied by 100.