Insurance: Mathematics and Economics 104 (2022) 185–199

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

www.elsevier.com/locate/ime

What can we learn from telematics car driving data: A survey

Guangyuan Gao [a], Shengwang Meng [a], Mario V. Wüthrich [b,*]

[a] Center for Applied Statistics and School of Statistics, Renmin University of China, 100872 Beijing, China
[b] ETH Zurich, RiskLab, Department of Mathematics, 8092 Zurich, Switzerland

ARTICLE INFO

Article history:
Available online 1 March 2022

JEL classification:
G22
C38

Keywords:
Telematics car driving data
Heatmaps
Poisson regression models
Convolutional neural networks
Limited fluctuation credibility model

ABSTRACT

We give a survey on the field of telematics car driving data research in actuarial science. We describe and discuss telematics car driving data, we illustrate the difficulties of telematics data cleaning, and we highlight the transparency issue of telematics car driving data resulting in associated privacy concerns. Transparency of telematics data is demonstrated by aiming at correctly allocating different car driving trips to the right drivers. This is achieved rather successfully by a convolutional neural network that manages to discriminate different car drivers by their driving styles. In a last step, we describe two approaches of using telematics data for improving claims frequency prediction, one is based on telematics heatmaps and the other one on time series of individual trips, respectively.

1. Introduction and literature overview

Statistical analysis of telematics car driving data is a vastly growing and exciting field of actuarial science. The purpose of this survey is to give a summary of the current state-of-the-art of this field, and to point out potential research directions. The best way to give an overview of the field is to provide a literature review where we try to identify important contributions to this field of statistical modeling; this will be done in this section. In subsequent sections we will describe the nature of telematics car driving data, the difficulties one faces dealing with telematics data, and we illustrate ways of making telematics data useful for inference and predictive modeling of insurance claims.

Early work on telematics car driving data is not directly related to actuarial problems. The papers of Esteves-Booth et al. (2001), Hung et al. (2007), Wang et al. (2008), Kamble et al. (2009) and Ho et al. (2014) have appeared in the transportation literature after the turn of the millennium. These papers use telematics data to understand vehicular emission, energy consumption and impacts on traffic in different cities of the world. In the transportation literature the corresponding field also goes under the name of driving cycles, and these authors study the cities of Edinburgh, Hongkong, Pune, Singapore and various Chinese cities. Analyzing driving cycles, Hung et al. (2007) use speed-acceleration probability distributions as a selection criterion of the most representative driving cycles. Their analysis shows that driving cycles on different routes (e.g., urban, sub-urban, highway) result in quite different speed-acceleration patterns. Similar to speed-acceleration probability distributions, the speed-acceleration (v-a) heatmap was introduced to the actuarial literature in Wüthrich (2017). It has been successfully used for classifying different driving styles, see Gao and Wüthrich (2018) and Zhu and Wüthrich (2021), and for improving claims frequency prediction, we refer to Gao et al. (2019a,b, 2022).

There is a stream in the transportation literature on driving behavior associated with accidents. Joubert et al. (2016) discretize acceleration into a finite risk space and complement it with speed data to analyze accidents; this approach is similar to v-a heatmaps. Ma et al. (2018) and Hu et al. (2019) consider contextual driving performance measurements such as relative speed to real-time traffic speed on the same road, contextual speeding under various traffic conditions, congestion level, duration on different road types or at different daytimes, etc. They examine the relationship between driving performance and accidents in a generalized linear model. They conclude that peak time driving, hard braking and acceleration relative to speed are important risk factors. A limitation in their study is that the accident history does not match with telematics data observation period. Wahlström et al. (2015) propose a framework for the detection of dangerous vehicle cornering events. They apply an unscented Kalman filter to raw GPS time series data to obtain their driving statistics. This framework is tested in a field study.

Early developments on telematics data that are more closely related to insurance problems are the work of Boucher et al. (2013)

* Corresponding author.
E-mail address: mario.wuethrich@math.ethz.ch (M.V. Wüthrich).

https://doi.org/10.1016/j.insmatheco.2022.02.004

and Paefgen et al. (2014) who study risk modeling based on so-called pay-as-you-drive (PAYD) insurance data. PAYD insurance is also called usage-based insurance (UBI). These first papers try to understand the impact of mileage on the risk of accidents. Boucher et al. (2013) show that the effect of mileage on the risk of an accident is far different from being linear. This may be because drivers with higher mileage are more experienced drivers, driving newer automobiles and driving more often on (safer) highways. Paefgen et al. (2014) consider multivariate exposures by aggregating mileage w.r.t. the daytime, weekday, road type and speed intervals. In a similar spirit, Verbelen et al. (2018) treat such multivariate exposures as compositional variables. Ayuso et al. (2016b) establish a survival model using a Weibull regression for the distance traveled to the first accident at fault. They find that speeding and nighttime driving reduce the distance traveled to the first accident. With the same survival model, Ayuso et al. (2016a) find that gender differences in the risk of accidents are mainly attributable to the intensity of vehicle use, i.e., men drive more frequently than women. Ayuso et al. (2019) further incorporate driving habit data such as percentage of distance traveled at night, above the limit, or in urban areas. Based on a dataset of a Taiwan insurer, Lemaire et al. (2016) find that mileage is the most powerful predictor of the number of claims at fault (using a Wald test), and the information contained in the bonus-malus premium level complements the importance of the mileage variable. Boucher et al. (2017) study the non-linear effects of duration and distance exposure on the risk of an accident in a generalized additive model.

The previous literature mainly considers PAYD features, but, of course, besides driving habits we should also explore driving skills and driving behavior. In insurance, this has motivated to contrast PAYD products to pay-how-you-drive (PHYD) products. Such PHYD products can be designed based on dangerous driving maneuvers, disregarding traffic rules like speeding, smart phone use, etc. There is vastly growing literature in this area of research, and we only mention a small selection of the available literature. Huang and Meng (2019) incorporate travel habits, driving performance and critical incidents into their claims frequency predictive model. Based on thorough feature engineering, they extract 30 telematics variables for each driver. Sun et al. (2020) use brake count and average position of the accelerator pedal as dependent variable measuring driving risk due to lack of accident or claims data. They use driving distance, speed, and revolutions per minute (RPM) as independent variables. So et al. (2021b) use intensity of sudden braking/acceleration/turns and proportions of durations on different road types and daytimes to predict the number of accidents in a cost-sensitive multi-class AdaBoost algorithm. This proposed algorithm can deal with the class imbalance problem of accidents. Denuit et al. (2019) propose a credibility approach to incorporate posterior information of driving experience into insurance pricing. Since public telematics data is not available, So et al. (2021a) generate synthetic data that shares similar features as real telematics data maneuvers, so that our community can at least explore such synthetic data. More on the economic and legal side Geyer et al. (2020) explore the effect of driving behavior on insurance contract design, risk and insurance selection. Eling and Kraft (2020) provide an overview of the usefulness of telematics in automobile insurance, health insurance and household insurance.

Another interesting stream of literature is Guillén et al. (2020, 2021) who study near-misses. Since car insurance claims are rare events (low frequency events), car insurance data suffers from the so-called class imbalance problem which means that the class of accident-free drivers is by far bigger than the class of drivers that suffer accidents. This may pose some challenges in determining important explanatory features for claims prediction. Guillén et al. (2020, 2021) enlarge the class of drivers with accidents by so-called near-misses incidents. A main difficulty is that a good definition of a near-miss event needs to be found, e.g., extreme acceleration and deceleration or excessive use of smart phones during driving may be a suitable definition.

Most of the previous work is based on extracting scores from telematics data, e.g., distances driven at nighttimes, numbers of excessive acceleration, relative amount of speeding, etc. Early work on directly using raw telematics data in the sense of high-frequency data (recording speed, acceleration and change of direction) has been done by Weidner et al. (2016, 2017) who extract covariates directly from time series of telematics location data using discrete Fourier transforms. In a similar vein Gao and Wüthrich (2019) use such time series data to allocate individual car driving trips to the right drivers. The paper of Meng et al. (2022) studies time series of individual trips for claims prediction by identifying more and less safe trips. Interestingly, in a similar spirit, Bayat et al. (2021) use telematics data as a biomarker for preclinical diagnosing of an Alzheimer disease, as driving styles seem to change under this disease. In general, new statistical and machine learning approaches should be able to directly act on raw time series of telematics data through the capability of representation learning, that is, machine learning methods are capable to engineer raw features into a new representation so that they are suitable for predictive modeling, we also refer to Section 7.1 in Wüthrich and Merz (2021) for representation learning.

Finally, we highlight Appendix A of Gao et al. (2019a) and the paper of Duval et al. (2021) which study how much telematics information is needed. Both papers conclude that roughly 3 months of telematics car driving data are sufficient to receive stable telematics information, of course, the caveat being that the car driving behavior is stationary over at least this time period. Thus, for those insurance companies that have not been collecting telematics data, yet, it will not take too much time to catch up on the data side (supposed that the underlying insurance portfolio is sufficiently large and that the right technology is in place).

Organization. In the next section we discuss available explanatory variables for insurance pricing. This includes classical actuarial covariates as well as telematics car driving data. Moreover, we identify the typical size of telematics data, and we conclude that maintaining and using this data can be a challenge because of its big size. In Section 3 we illustrate telematics data and we discuss the issue of data quality. In Section 4 we allocate telematics car driving observations to individual car drivers, which allows us to classify car drivers and it also raises the critical issues of transparency and privacy concern. In Section 5 we indicate two different approaches of making telematics data useful for claims prediction and car insurance pricing. The first approach is based on aggregated telematics data (to reduce the size of telematics data), and the second approach scores individual car driving trips before entering a regression model. Finally, in Section 6 we conclude and give an outlook.

2. What is telematics car driving data?

We discuss the nature of telematics car driving data in this section. We start by describing the classical covariate information that is usually available for car insurance pricing before explaining telematics car driving data.

2.1. Classical covariates for car insurance pricing

Classical car insurance pricing uses roughly 50 potential covariates in its statistical modeling procedure. This information is either received at inception of the insurance contract (i.e., it is available for initial insurance pricing), or, subsequently, complemented by

claims experience at contract renewals. Therefore, one typically distinguishes between prior information and posterior information for insurance pricing, see Denuit et al. (2019).

Available feature information at inception of the contract includes:

- **Car related features:** type of car, car brand, vehicle model, size of car, age of car, weight of car, horse power, type of engine, fuel type, cubic capacity, price of car, additional equipment, driving assistance tools, number of seats, etc.
- **Driver related features:** driver's age, gender, nationality, size of household, marital status, dependent children, occupation, medical conditions, credit record, leasing, type of flat, garage, etc.
- **Insurance contract related features:** type of contract, date of contract, duration of contract, sales channel, deductible, bonus protection, other insurance products, etc.
- **Location related features:** province of living, city of living, zip code, city-rural area, etc.

All this feature information is not directly driving style related. The difficulty is that we would like to know whether we have a good or bad driver, safe or incautious driver, calm or aggressive driver, but since this information is not available (it is latent) we use proxy variables that may characterize driving styles of drivers. For instance, the type of car or the type of insurance product may reveal driving style information, or better, may correlate with driving styles. Therefore, regression models aim at finding systematic effects in this feature information for explaining propensity to claims.

Apart from the above information, there is also driving related information available in classical insurance pricing:

- **Driving related features:** date of driving test, annual mileage, vehicle use, bonus-malus level, claims experience, time since last claim, etc.

Remark that this driving related information is by far smaller than the previously mentioned features. The date of the driving test tells us something about driving experience. The annual mileage is a quantity that is estimated by the policyholder, thus, it is not precise and it can be corrupted to save premium. Also the bonus-malus level can be problematic, in theory, the bonus-malus level directly encodes past claims experience, and, thus, should be very predictive. However, there are several ways to circumvent an unfavorable bonus-malus level, e.g., the initial assessment is not correctly stated, small claims are not reported but covered by the policyholder, or there is bonus protection insurance in place which ensures that the bonus-malus level is not changed in case of an accident.

Summarizing, these covariates serve as proxies for driving habits and driving styles, but most of them are not directly related to driving skills. In general, we avoid talking about causality, here, as mostly we only explore correlations through these proxy variables.

2.2. Telematics car driving data

In contrast to the classical covariate information above, telematics car driving data directly records driving habits and driving styles. Telematics car driving data may have many different formats. In the introduction we have been mentioning the number of dangerous maneuvers, the amount of speeding or mobile use, see also Guillén et al. (2020, 2021) and So et al. (2021a,b). Here, we are mainly going to focus on high-frequency time series of telematics data, sometimes also called raw telematics data. This time series data records a set of variables every second during driving. Consequently, we see telematics data in the field of the so-called internet of things (IoT) where sensors continuously control and record the environment, and the resulting measurements are exchanged and stored, e.g., in a data cloud. Typically, we think of the following information received second by second:

- global positioning signal (GPS) location data, speed, acceleration, braking, intensity of left and right turns,
- engine information like engine revolutions, engine temperature, etc.,
- vehicle sensors and cameras, information of driving assistance tools,
- time stamp (daytime, rush hour, night, etc.), total mileage at different daytimes,
- road type, traffic conditions, weather conditions, etc.
- traffic rules (e.g. speeding), driving and health conditions, etc.
- number of passengers, distraction, smart phone use, etc.

In Table 5 in the appendix we provide a short example of a telematics car driving data time series. This example includes time stamp, GPS location, speed, acceleration, engine revolution and the quality of the GPS signal being in $\{0, 1\}$. We observe that the GPS signal is sometimes missing (empty entries in Table 5) due to signal failure. This difficulty is going to be discussed further in the next section.

We do a small back-of-the-envelop calculation to understand the size of telematics data. If we assume that we collect 100 KB telematics data per driver and per day, this amounts to roughly 40 MB of data per driver every year. Having a comparably small insurance portfolio of 100,000 drivers, results in 4 TB of data every year. Thus, we easily get data volumes that can neither be stored nor be handled on conventional personal computers. This highlights that the volume of telematics data can easily be a challenge. Such data is likely to be stored in a data cloud, and it requires cloud computing to evaluate a whole insurance portfolio of telematics high-frequency time series data. We mention this because it imposes quite some challenges to academics as maintaining such an environment is both costly and time consuming, and we probably need to find ways to collaborate with industry to do research on telematics data.

At this stage, also good data warehousing becomes important as it may not be easy to identify all trips of a given driver in big data where new data is added for multiple drivers every day. For this reason, most of the present research in the literature does not directly explore time series telematics data, but extracts scores that are further processed by statistical models. In our research, see e.g. Gao et al. (2019a), we work with exposures and claims data of roughly 2 to 3 years, i.e., we can follow individual drivers and their claims experience for almost 3 years, however, the telematics information that we use is compiled from a shorter time period. Typically, we extract driving styles from 3 months of telematics data, and we have proved that this volume is sufficient to receive stable and reliable telematics information. Of course, the disadvantage of this shortcut is that the claims experience and the telematics data do not span exactly the same time interval, and claims cannot be allocated to individual trips, but only to the driving styles that have been extracted from the 3 months of telematics data. Moreover, such an approach assumes stationarity over the whole time interval considered, so that the extracted telematics features are relevant for the observed claims history.

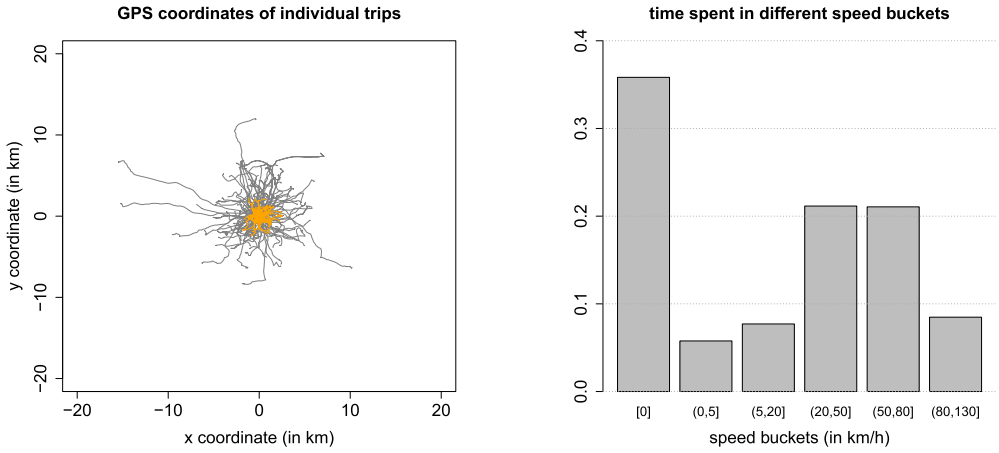

Fig. 1. (lhs) 200 individual trips of a given driver, and (rhs) simple speed statistics. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

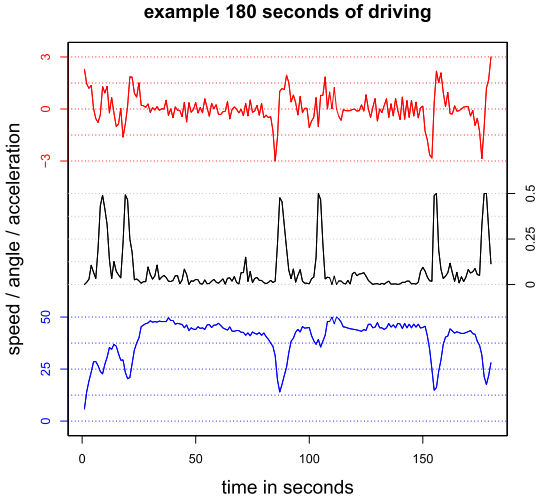

Fig. 2. Speed v_t (blue), acceleration/deceleration a_t (red) and change in direction Δ_t (black) for 180 seconds of driving $1 \le t \le 180$.

3. Illustration of telematics data

In this section we illustrate telematics time series data and we highlight the challenges in data collection and data cleaning.

3.1. Speed, acceleration and change of direction

We start by compiling information similar to Table 5 in the appendix to simple statistics. The left-hand side of Fig. 1 shows the GPS coordinates second by second of 200 individual trips of a given car driver. For illustrative purposes all trips start in coordinate $(0, 0)$, and each trip is rotated by a random angle. The 100 shorter trips are colored yellow, and the 100 longer trips are in gray. Having these GPS coordinates (x_t, y_t) every second $t \ge 0$ we can calculate average speeds v_t, acceleration/deceleration a_t and change in direction Δ_t every second $t \ge 1$. The right-hand side of Fig. 1 shows the relative time spent (over all trips) in the different speed buckets [0], $(0, 5]$, $(5, 20]$, $(20, 50]$, $(50, 80]$, $(80, 130]$; units are km/h. Noticeable is that the idling phase amounts to roughly 35% of the total time spent while running the vehicle engine.

Fig. 2 shows speed v_t, acceleration a_t and change of angle Δ_t of 180 seconds of driving $1 \le t \le 180$. From this graph it is apparent that braking (in red) lowers the speed (in blue), and in many cases this goes along with a change of direction (in black). All this seems rather obvious, and the use of this data seems straightforward, but to get to this point we first need to discuss data quality.

3.2. Difficulties about telematics data

Often data quality is a serious issue in telematics car driving data, and having big data makes data cleaning an even bigger challenge. Data quality can be most easily understood by having different devices measuring the same quantity, but giving different results. In our case, we have different sources of speed and acceleration information; we also refer to Table 5 in the appendix:

- GPS location data gives the position (x_t, y_t) every second t from which speed $v_t^{(\mathrm{xy})}$, acceleration $a_t^{(\mathrm{xy})}$ and change in direction $\Delta_t^{(\mathrm{xy})}$ can be calculated.
- Moreover, the GPS device directly provides speed $v_t^{(\mathrm{gps})}$ and heading of the car. This information is supported by the information of the quality of the signal being in $\{0, 1\}$.
- There is the vehicle instrumental panel that provides speed $v_t^{(\mathrm{vss})}$ from the vehicle speed sensor (VSS).
- There is an accelerometer installed (black box) that measures longitudinal $a^{(\mathrm{acc})}$, lateral and vertical acceleration w.r.t. the car's direction.

The difficulty in practice is that this information is often not in line, information is incomplete, jumps in direction may happen, e.g., between 0° and 360°, and (regular) calibration of the accelerometer seems a general issue.

Fig. 3 shows one single trip of 120 seconds. It illustrates differences between GPS speed $v_t^{(\mathrm{gps})}$ in blue and the vehicle sensor speed $v_t^{(\mathrm{vss})}$ in green on the left-hand side. The right-hand side shows the acceleration $a_t^{(\mathrm{acc})}$ from the accelerometer in red, which after second 80 is a positive constant, though the speed pattern on the left shows that this car is standing still. Thus, the calibration of the accelerometer is not correct, here, and the different speed patterns between $v_t^{(\mathrm{gps})}$ and $v_t^{(\mathrm{vss})}$ may be essential in understanding whether we have a calm or an aggressive driver.

Fig. 4 gives an example of missing signals. The left-hand side of this figure shows the GPS and the vehicle sensor speeds, and the right-hand side shows the quality of these signals. The GPS signal has quality issues around seconds 30 and 200, and the vehicle sensor signal completely fails after 400 seconds. Of course, such a situation will not allow us to successfully calculate the entire acceleration pattern, if the speed entry is missing for part of the trip.

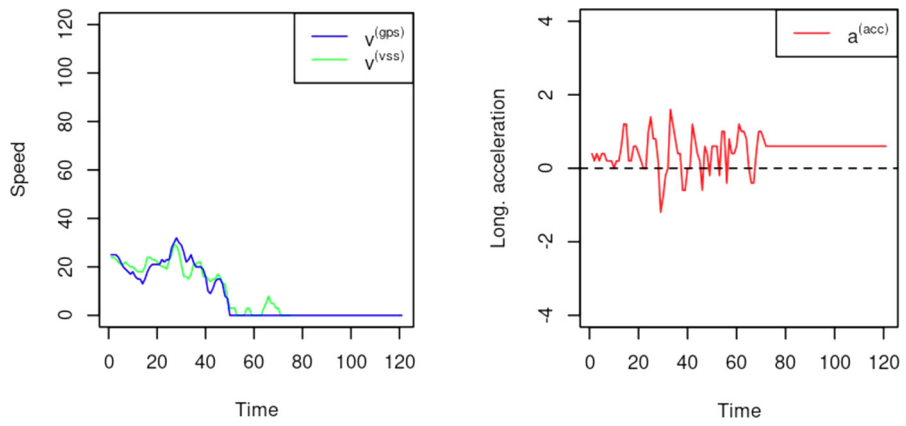

Fig. 3. 120 seconds of driving: both plots show the same trip with GPS speed $v_t^{(\text{gps})}$ in blue, the vehicle sensor speed $v_t^{(\text{vss})}$ in green and the accelerometer acceleration $a_t^{(\text{acc})}$ in red.

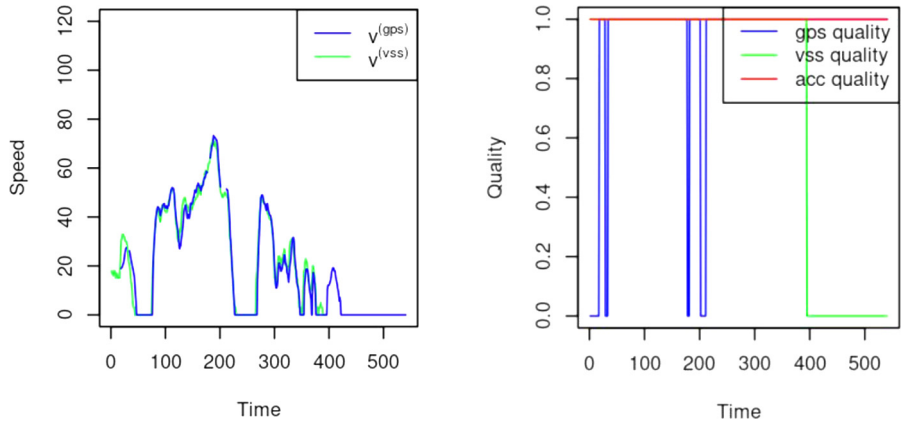

Fig. 4. Missing GPS signals and missing vehicle sensor signals: (lhs) gives the GPS and the vehicle sensor speeds, and (rhs) shows the quality of these signals.

We conclude that we have experienced substantial difficulties with the signal quality of the accelerometer, and also with GPS locations (x_t, y_t) we have experienced issues like position shifts. GPS speeds $v_t^{(\text{gps})}$, GPS direction and vehicle sensor speeds $v_t^{(\text{vss})}$ seem more reliable. Having the quality of these two signals, we can also determine missing data, and in case of missing data we can either disregard the whole trip or impute suitable values. We remark that this data cleaning process takes by far the most time in the whole data modeling process. This can easily add up to more than 90% of the total time of modeling, and the statistical part seems time-wise almost negligible.

4. Transparency of telematics car driving data

We now assume to have successfully cleaned our data and we would like to present a first analysis on this data. This first analysis will indicate how much insight we can gain about individual drivers from telematics trip data second by second.

We select 3 different drivers, we call them A, B and C, and we analyze the speed v_t, acceleration/deceleration a_t and change in direction Δ_t patterns of their individual trips of 180 seconds, $1 \le t \le T = 180$. The question that we try to answer is whether this time series data is sufficient to correctly allocate the individual trips to the right driver. Fig. 5 shows 3 trips of each of these 3 drivers A, B and C, and we explore whether these plots discriminate different drivers. We emphasize that the 180 seconds have been selected at random from the entire (bigger) trips to limit the influence of frequently traveled routes, e.g., the beginning of the trip will likely explore the neighborhood of the living place of the driver. There is also nothing particular in choosing exactly 180 seconds, it has just turned out in our analysis that this amount of data provides reliable classification results. In order to solve this classification question we use the set-up of Gao and Wüthrich (2019).

4.1. Problem setting and data pre-processing

Our goal is to correctly allocate $T = 180$ seconds of telematics car driving experience to the right driver, see Fig. 5. Dealing with time series data, there are different machine learning tools that allow one to discriminate these trips. These are recurrent neural networks (RNNs), convolutional neural networks (CNNs) or attention layers. For our purpose CNNs are the most appropriate tool, as CNNs have certain translation invariance properties, see Wiatowski and Bölcskei (2018). These translation invariance properties allow CNNs to find similar structure in different parts of the time series. This is exactly a necessary property that we would like to have in our analysis: hard braking and fast acceleration can occur at any time during these 180 seconds, but this describes the same driving style. In a nutshell, CNNs have different filters that allow the network to find similar structure in different parts of the time series. These filters act like rolling windows, sliding across the whole time series trying to spot a particular structure in this (rolling) window.

Before formally introducing CNNs, we describe and pre-process our telematics data. We denote the triple speed, acceleration and

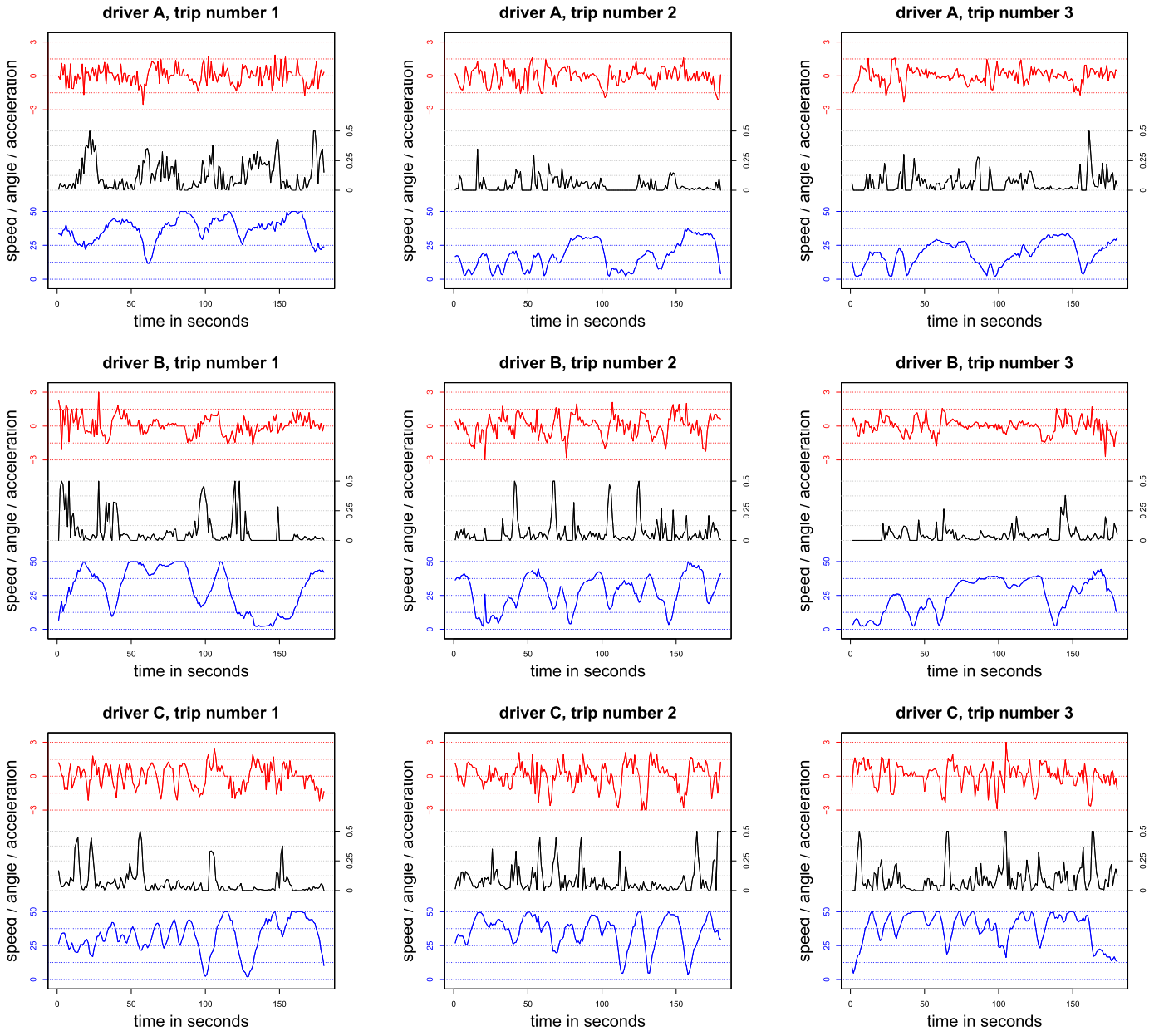

Fig. 5. First 3 trips of 3 selected drivers A, B and C: each trip is $T = 180$ seconds.

change in direction at seconds $1 \leq t \leq T$ by

$$(v_{j,t}, a_{j,t}, \Delta_{j,t})^\top \in [2, 50]\text{km/h} \times [-3, 3]\ \text{m/s}^2 \times [0, 1/2], \quad (4.1)$$

where $1 \leq j \leq J$ labels the individual trips. Speed satisfies $v_{j,t} \in [2, 50]$km/h, we make this choice because we do not want to classify drivers by having a different idling or high speeding behavior. To remain in this speed interval we censor speed, and we concatenate the trips by cutting off the censored part up to 2 seconds, so that the resulting speed pattern is still smooth. Acceleration and deceleration is censored at ± 3 m/s^2 because of scarcity of data outside of this interval, and we take the absolute value of the sine of the change in angle censored at 1/2. Weidner et al. (2017) state that strong acceleration can go up to $+6$ m/s^2 and a maximal deceleration can go down to -8 m/s^2 under good conditions, a wet surface typically changes this value to -7 m/s^2. Remark that strong acceleration and hard deceleration also serve at defining the near-missing events in Guillén et al. (2020).

The triple (4.1) allows us to define the 3-dimensional time series

$$\boldsymbol{z}_j = \Big((v_{j,1}, a_{j,1}, \Delta_{j,1})^\top, \ldots, (v_{j,T}, a_{j,T}, \Delta_{j,T})^\top \Big)^\top \in \mathbb{R}^{T \times 3}. \quad (4.2)$$

The covariate $\boldsymbol{z}_j$ describes 180 seconds of telematics driving experience and we allocate the response $Y_j \in \{\text{A}, \text{B}, \text{C}\}$ to each time series $\boldsymbol{z}_j$ to indicate which of the 3 drivers has been driving the selected trip. Thus, we have observations $(Y_j, \boldsymbol{z}_j)_j$ with a categorical response taking the three values A, B or C.

4.2. Classification with convolutional neural networks

In this section we describe the CNN architecture that we use for our classification problem, this architecture is taken from Gao and Wüthrich (2019). A CNN can be seen as an extension of a multinomial logistic regression model. Our time series feature $\boldsymbol{z} \in$

Listing 1: CNN architecture for individual trip allocation taken from Gao and Wüthrich (2019).

```
model <- keras_model_sequential()
#
model %>%
layer_conv_1d(filters = 12, kernel_size = 5, activation='tanh', input_shape = c(180,3)) %>%
     layer_max_pooling_1d(pool_size = 3) %>%
layer_conv_1d(filters = 10, kernel_size = 5, activation='tanh') %>%
     layer_max_pooling_1d(pool_size = 3) %>%
layer_conv_1d(filters = 8, kernel_size = 5, activation='tanh') %>%
     layer_global_max_pooling_1d() %>%
     layer_dropout(rate = 0.3) %>%
layer_dense(units = 3, activation = 'softmax')
```

$\mathbb{R}^{T\times q_0}$ has length T and $q_0 = 3$ channels. For a CNN layer we have to choose the number of filters $q_1 \in \mathbb{N}$, the band width $b \in \mathbb{N}$ and the filter weights $W_s \in \mathbb{R}^{b\times q_0}$, $1 \le s \le q_1$. A CNN layer is then given by a mapping

$$\begin{aligned}\psi : \mathbb{R}^{T\times q_0} &\to \mathbb{R}^{(T-b+1)\times q_1} \\ \boldsymbol{z} &\mapsto \psi(\boldsymbol{z}) = \big(\psi_1(\boldsymbol{z}), \ldots, \psi_{q_1}(\boldsymbol{z})\big),\end{aligned} \tag{4.3}$$

where each component $\psi_s(\boldsymbol{z})$, $1 \le s \le q_1$, is a new univariate time series of length $T - b + 1$. Broadly speaking, these components are obtained by a convolution operation $*$ of the input $\boldsymbol{z}$ with the filter weights W_s

$$\boldsymbol{z} \mapsto \psi_s(\boldsymbol{z}) = \phi\left(w_s + W_s * \boldsymbol{z}\right) \in \mathbb{R}^{T-b+1}, \tag{4.4}$$

with bias $w_s \in \mathbb{R}$, $\phi : \mathbb{R} \to \mathbb{R}$ is the chosen non-linear activation function, and the length of the time series is reduced by the band width from T to $T - b + 1$. We make a couple of remarks:

- The convolutional operation $*$ in (4.4) is slightly different from the classical mathematical convolution, because indexes are reversed compared to the mathematical version. For a precise definition of (4.4) we refer to formula (9.3) in Wüthrich and Merz (2021).
- A CNN layer involves $q_1(1 + bq_0)$ parameters (biases and filter weights).
- We have used the default stride of 1, which implies sliding the window with step size 1.
- Typically, after a CNN layer one applies a so-called max-pooling layer to reduce the size of the data. A max-pooling layer works very similar to a CNN layer, but the convolutional operation is replaced by considering the maximum in (disjoint) windows of a given band width; for details see Section 9.2.4 in Wüthrich and Merz (2021).
- If the reduction in the length of the time series from T to $T-b+1$ is an unwanted feature, one applies padding, meaning that one fills both ends of the shortened time series with zeros back to the original length T.

To process our time series data $\boldsymbol{z}_j \in \mathbb{R}^{T\times 3}$ given in (4.2) we compose three CNN layers ψ^1, ψ^3, ψ^5 with band width $b = 5$ and filters $q_1 = 12, 10, 8$, respectively. Each of these CNN layers is followed by a max-pooling layer ψ^2, ψ^4, ψ^6. This gives us a CNN architecture that maps the telematics time series data to an eight dimensional vector:

$$\boldsymbol{z}_j \mapsto \psi^{(6:1)}(\boldsymbol{z}_j) = \Big(\psi^6 \circ \psi^5 \circ \psi^4 \circ \psi^3 \circ \psi^2 \circ \psi^1\Big)(\boldsymbol{z}_j) \in \mathbb{R}^8.$$

Listing 1 shows this CNN architecture. The CNN part $\psi^{(6:1)}(\boldsymbol{z}_j) \in \mathbb{R}^8$ uses $192 + 610 + 408 = 1,210$ parameters, see Listing 3 in Gao and Wüthrich (2019). To prevent from overfitting, we apply a drop-out layer to these 8 neurons having a drop-out rate of 30%, see Listing 1. Finally, we apply a fully-connected dense layer that maps the 8 neurons to the 3 categorical outputs, this involves another 27 parameters. We use for this output the softmax activation to ensure that we obtain categorical probabilities adding up to 1, this corresponds to the inverse of the canonical link in the multinomial logistic classification model. This network architecture is fitted to the data using a variant of the gradient descent algorithm and back-propagation.

Table 1
Individual trip allocation out-of-sample results on $\mathcal{T}$.

	true labels		
	driver A	driver B	driver C
predicted label A	**33**	4	0
predicted label B	8	**38**	6
predicted label C	1	5	**36**
% correct	78.6%	80.9%	85.7%

4.3. Fitting and results

We fit the network architecture of Listing 1 to the available telematics data $(Y_j, \boldsymbol{z}_j)_j$ of the three drivers. In total we have 652 individual trips of these 3 car drivers A, B and C. We partition this data at random into a learning data set $\mathcal{L}$ which receives 521 individual trips and a test data set $\mathcal{T}$ that contains the remaining 131 trips. We fit the CNN architecture only on $\mathcal{L}$ and we use $\mathcal{T}$ for the out-of-sample analysis. The learning data $\mathcal{L}$ is further split into a training data set $\mathcal{U}$ and a validation data set $\mathcal{V}$ at ratio 4:1 to fit the architecture on $\mathcal{U}$ and to track over-fitting on $\mathcal{V}$. We use a callback to retrieve the model with the lowest validation loss on $\mathcal{V}$, we refer to Section 7.2.3 and Figure 7.7 in Wüthrich and Merz (2021) for this fitting strategy. The fitting of this architecture takes roughly 40 seconds on a standard personal laptop.

Table 1 gives the out-of-sample results on $\mathcal{T}$ of this classification problem. We observe that roughly 80% of the trips have been correctly allocated; a purely random allocation would be correct in 33.3% of the cases. This result is quite remarkable as we have trained this network architecture on only 521 individual trips of a total length of 180 seconds. Moreover, we did not apply much fine-tuning to the network nor did we pay special attention during the data cleaning. Thus, this network architecture learns structure from the individual telematics time series data that allows to discriminate the drivers A, B and C rather accurately. To verify these results we have performed the same analysis on different sets of drivers, Table 2 shows the results of 3 other drivers called D, E and F, which confirms the results.

Remarks.

- The 180 seconds have been chosen at random from the entire (bigger) trips to ensure that the network does not learn a frequently traveled route (which may always have the same speed pattern). The results are stable w.r.t. different random

Table 2
Individual trip allocation out-of-sample results on $\mathcal{T}$.

	true labels		
	driver D	driver E	driver F
predicted label D	**43**	12	2
predicted label E	5	**64**	5
predicted label F	4	2	**51**
% correct	82.7%	82.1%	87.9%

choices of these 180 seconds, and Table 2 also verifies that we have stability w.r.t. other selected drivers.

- The main drawback of these results (and of networks in general) is that we cannot say why the networks manage to solve this task so successfully. That is, we do not know according to which features the network has learned to discriminate the drivers. Unfortunately, we do not have better telematics data to explore this question, e.g., it would be interesting to know how the vehicle model or different road types influence this classification task.
- We use 180 seconds of driving experience for this classification. This is rather short but already sufficient to classify the trips. For claim frequency prediction we prefer to work with longer observation periods, in fact, we would like to consider maximal information to capture, e.g., as many hard braking and acceleration events as possible.

4.4. What's next?

In the previous section we have shown that we can allocate individual car driving trips rather successfully to the right driver. Of course, from an actuarial viewpoint this is very interesting because it also means that we can distinguish different driving styles. Thus, a natural next extension is to score individual car driving trips according to driving styles, which then directly relates to propensity to claims and to the corresponding insurance prices. In Section 5.5, below, we discuss such an approach and we also describe the difficulties dealing with individual trip scoring.

Up to here, we have considered telematics data through the lenses of a machine learner and a statistician, and we have become very excited about how accurately we can perform prediction based on telematics data. However, at this stage we should also put on our professional actuarial glasses, as this transparency in data will raise privacy concerns of policyholders. It seems that it is comparably easy to identify people, their state of driving and health, as well as their daily routine using this telematics data. E.g., this data will display whether someone regularly visits a pub in the evening by car, as the driving style will slightly change on the way home; it will show which family member drives a particular trip in the family car; or such data may be used to preclinical diagnose an Alzheimer disease, see Bayat et al. (2021). Therefore, at this stage, we have to have clear legal rules to deal with associated privacy concerns of customers, and the ownership of the data needs to be discussed and clarified. These are general questions that concern our society and are beyond statistical modeling, but need a broader political and legal discussion and consensus.

5. Using telematics data for claims frequency prediction

The purpose of this section is to indicate the predictive power of telematics car driving data to forecast claim frequency. As mentioned in Section 2, classical actuarial covariates may serve as proxies for driving habits and driving styles. These classical covariates are now complemented with telematics data which directly records driving behavior. Telematics data can be incorporated into a claims frequency model via various formats such as summary statistics, scores of dangerous maneuvers, v-a heatmaps, time series, etc. In this section we mainly focus on an automated feature engineering of heatmaps, i.e., representation learning on telematics data, and we briefly discuss risk scoring of time series of individual trips. This section describes similar models to those presented in Gao et al. (2022) and Meng et al. (2022). Remark that the results in these two papers are not directly comparable since they are based on different volumes of telematics data. Here, we make the results comparable by using the same telematics data as in Meng et al. (2022). Moreover, we also incorporate change in direction Δ, that is, we also consider v-Δ heatmaps.

5.1. Available claims data and Poisson claims count regression modeling

We first summarize the available data used in Meng et al. (2022). The data comes from $n = 1,847$ motor third-party liability (MTPL) insurance policies with a risk exposure from 01/01/2014 to 19/06/2017. We denote the time exposure of policy $1 \le i \le n$ by $v_i > 0$ and the (observed) number of claims on that policy by N_i. The total exposure over all policies is $\sum_{i=1}^n v_i = 4,214.65$ years-at-risk, thus, these policies have an average exposure of 2.28 years. The average observed annual claim frequency is $\bar{\mu} = \sum_{i=1}^n N_i / \sum_{i=1}^n v_i = 0.22$, i.e., one out of four car drivers suffers an accident within a given year. Similar to Section 4.3, this data is then partitioned into a learning data set $\mathcal{L}$ and a test data set $\mathcal{T}$. Model fitting only takes place on the learning data $\mathcal{L}$, and the test data set $\mathcal{T}$ is used for an out-of-sample generalization analysis. We use the same partition as in Table 1 of Meng et al. (2022).

Each insurance policy $1 \le i \le n$ is further supported by covariates $(\boldsymbol{x}_i, \boldsymbol{z}_i)$, where $\boldsymbol{x}_i$ describes the classical actuarial covariates in tabular form, and $\boldsymbol{z}_i$ describes the telematics information, we refer to Section 2. The goal is to model the number of claims N_i of each of these car drivers, given covariate information $(\boldsymbol{x}_i, \boldsymbol{z}_i)$. We therefore assume existence of a regression function that captures the systematic effects in the claims

$$(\boldsymbol{x}, \boldsymbol{z}) \;\mapsto\; \mu(\boldsymbol{x}, \boldsymbol{z}) > 0,$$

such that for all car drivers $1 \le i \le n$ we have an expected number of claims

$$\mathbb{E}\left[N_i \,|\, \boldsymbol{x}_i, \boldsymbol{z}_i\right] = \mu(\boldsymbol{x}_i, \boldsymbol{z}_i) v_i.$$

A common assumption then is that the numbers of claims N_i are independent across all car drivers and they can be modeled by a Poisson regression model

$$N_i \overset{\text{ind.}}{\sim} \text{Poi}\left(\mu(\boldsymbol{x}_i, \boldsymbol{z}_i) v_i\right), \qquad \text{for } 1 \le i \le n \text{ and exposure } v_i > 0. \tag{5.1}$$

The case of only using classical actuarial covariates $\mu(\boldsymbol{x}, \boldsymbol{z}) = \mu(\boldsymbol{x})$ has been studied extensively in the actuarial literature. Typical approaches use a GLM, a generalized additive model (GAM), a neural network or a tree boosting approach. Our main focus, here, is on integrating telematics data $\boldsymbol{z}$ that is not necessarily in tabular form.

For an out-of-sample model evaluation, i.e., a generalization analysis of an estimated model $\widehat{\mu}$, which has been received on the learning data $\mathcal{L}$, we consider the Poisson deviance loss (scoring function) on the test data $\mathcal{T}$ that takes the following form

$$L\left(\widehat{\mu}; \mathcal{T}\right) = \frac{2}{|\mathcal{T}|} \sum_{i \in \mathcal{T}} N_i \left(\frac{\widehat{\mu}(\boldsymbol{x}_i, \boldsymbol{z}_i) v_i}{N_i} - 1 - \log\left(\frac{\widehat{\mu}(\boldsymbol{x}_i, \boldsymbol{z}_i) v_i}{N_i} \right) \right) \ge 0, \tag{5.2}$$

Table 3
Generalization analysis: out-of-sample test error on $\mathcal{T}$ of the models studied.

Error	Null	Classical	C+VA	C+VA+VD	VA	VA+VD	C+T
	(5.4)	(5.3)	(5.7)	(5.10)	(5.9)	(5.11)	(5.15)
Test error	1.1003	1.0306	1.0128	**0.9982**	1.0616	1.0381	1.0286
Reduction	–	0.0697	0.0875	0.1021	0.0387	0.0622	0.0717

where we set the terms with $N_i = 0$ equal to $2\widehat{\mu}(\boldsymbol{x}_i, \boldsymbol{z}_i)v_i$. Typically, the model with the smallest out-of-sample loss $L(\widehat{\mu}; \mathcal{T})$ is preferred; for a decision-theoretic approach of model validation we refer to Gneiting and Raftery (2007), Gneiting (2011), Krüger and Ziegel (2021) and Section 4.1 of Wüthrich and Merz (2021); deviance loss (5.2) gives a proper scoring rule according to Gneiting and Raftery (2007).

5.2. Benchmark: Poisson GLM with classical actuarial covariates

As base case we fit a Poisson GLM on the classical actuarial covariates $\boldsymbol{x}$ by setting expected frequency $\lambda(\boldsymbol{x}) := \mu(\boldsymbol{x}) = \mu(\boldsymbol{x}, \boldsymbol{z})$, i.e., we drop telematics information $\boldsymbol{z}$. We pre-process the available categorical covariates of '`region`', '`car_brand`' and '`gender`', and the continuous covariates of '`driver_age`', '`car_age`', '`seat_count`', '`car_price`' and '`av_daily_dist`' (average daily distance), so that they fit the regression structure of a GLM. In particular, we merge regions with small exposures, categorize the 66 different car brands into 6 different vehicle classes according to the country of origin, called '`car_made`' in the sequel. We check the log-linearity of continuous variables using marginal generalized additive models. We remove the non-significant covariates of '`seat_count`' and '`car_price`', keep the log-linear covariates of '`car_age`' and '`av_daily_dist`'. We categorize '`driver_age`' into '`age_group`' because driver's age does not have a monotone influence on the expected claims frequency. We also refer to Chapter 5 in Wüthrich and Merz (2021) on covariate engineering within GLMs. The pre-processed covariates then provide us with feature information

$$\boldsymbol{x} = (1, \texttt{region}, \texttt{car\_made}, \texttt{gender}, \texttt{age\_group}, \texttt{car\_age}, \texttt{av\_daily\_dist}) \in \mathcal{X} \subset \{1\} \times \mathbb{R}^q.$$

Under the general setting (5.1), the regression function of a classical Poisson GLM has the following structural form

$$(\boldsymbol{x}, \boldsymbol{z}) \;\mapsto\; \mu(\boldsymbol{x}, \boldsymbol{z}) = \lambda(\boldsymbol{x}) = \exp\langle \boldsymbol{\beta}, \boldsymbol{x} \rangle, \tag{5.3}$$

where we choose the log-link (the canonical link of the Poisson model) with regression parameter $\boldsymbol{\beta} \in \mathbb{R}^{q+1}$ and covariates $\boldsymbol{x} \in \mathcal{X}$. This model is fitted on the learning data $\mathcal{L}$. The resulting GLM has an out-of-sample Poisson deviance loss on $\mathcal{T}$ of 1.0306, see Table 3. Furthermore, we could analyze this fitted model, e.g., by the Wald test checking for possible variable reduction, etc. We benchmark the GLM with the null model that does not consider any covariates

$$\bar{\mu} = \frac{\sum_{i \in \mathcal{L}} N_i}{\sum_{i \in \mathcal{L}} v_i} = 0.22. \tag{5.4}$$

The null model has an out-of-sample loss of 1.1003. Thus, the GLM has a clearly better out-of-sample predictive performance, this justifies the inclusion of classical actuarial covariates $\boldsymbol{x}$.

Remark. We work in a low-frequency problem, here, where the event of a claim is by far more rare than not observing any claim. As a result, loss figures are mainly driven by the pure randomness of the random variables (irreducible risk), and model improvements are of a smaller magnitude in loss evaluations. Considering the Poisson deviance losses, probably slightly more than 1 should be allocated to the irreducible risk, and 0.1 should be allocated to model error in the null model in Table 3. The Poisson deviance loss can also be used to estimate a dispersion parameter (after adjustment for the degrees of freedom), in our case this Poisson deviance loss is slightly bigger than 1, which indicates slight over-dispersion. This might be implied by model error or insufficient covariate information.

5.3. A representation of telematics data: speed-acceleration heatmaps

In a second step we enrich GLM (5.3) by telematics car driving information $\boldsymbol{z}$. The telematics data used is collected by the black box devices installed in the selected cars. We have different volumes of telematics data for different cars since the black box device can be installed anytime during the exposure period from 01/01/2014 to 19/06/2017. We use the same telematics data as in Meng et al. (2022). An assumption made here is that the driving behavior does not change during the observation period. This assumption seems reasonable since our portfolio contains more mature drivers. If one only studies less experienced (young) drivers, this assumption will be violated. Because telematics data is not in tabular form, we start by pre-processing telematics information so that we can use it for regression modeling. A first approach is the speed-acceleration (v-a) heatmap, it has its name from the graphical illustrations in Wüthrich (2017). In general, we can compress any amount of individual telematics car driving data into a v-a heatmap with fixed size. This v-a heatmap describes how a driver accelerates and decelerates at different speeds. A second more granular approach will be described in Section 5.5.

We briefly discuss how to construct a v-a heatmap. In a nutshell, it is similar to a two-dimensional histogram. In a v-a rectangle $R = [10, 60]$ km/h$\times[-4, 4]$ m/s^2, the acceleration interval is divided into $1 \le j \le 9$ sub-intervals $[-4, -3.5)$, $[-3.5, -2.5)$, $[-2.5, -1.5), \ldots$, $[1.5, 2.5)$, $[2.5, 3.5)$, $[3.5, 4]$, and the speed interval is partitioned into $1 \le k \le 5$ equally spaced sub-intervals $[10, 20)$, $[20, 30), \ldots, [50, 60]$. Note that Meng et al. (2022) truncate speed within $[10, 60]$ km/h and censor the acceleration within $[-4, 4]$ m/s^2.

We define the acceleration pattern of driver $1 \le i \le n$ in speed sub-interval $1 \le k \le 5$ by

$$z_{i,j,k} = \frac{t_{i,j,k}}{\sum_{j=1}^{9} t_{i,j,k}} \ge 0, \tag{5.5}$$

where $t_{i,j,k} \ge 0$ is the total amount of time spent in speed sub-interval $1 \le k \le 5$ and acceleration sub-interval $1 \le j \le 9$. The resulting vector $(z_{i,1,k}, \ldots, z_{i,9,k})^\top$ defines a discrete distribution for fixed i and k. These discrete distributions are summarized for each driver i in the following 9×5 matrix, called v-a heatmap,

$$\boldsymbol{z}_i = (z_{i,j,k})_{1 \le j \le 9,\ 1 \le k \le 5} \in [0, 1]^{9 \times 5}. \tag{5.6}$$

Thus, $\boldsymbol{z}_i$ describes a spatial object that can be interpreted as an image. In image recognition such an object $\boldsymbol{z}_i$ is represented by a three-dimensional array (tensor) $\boldsymbol{z}_i \in [0, 1]^{9 \times 5 \times 1}$. The first two components correspond to the (j, k) location in the image and the third component to the channels. This is similar to Section 4.2, where, here, we extend the time-series object to a spatial object, and where we have only one channel. Similarly, we can construct a v-Δ heatmap denoted by $\boldsymbol{u} \in [0, 1]^{9 \times 5 \times 1}$, where the v-Δ rectangle $[10, 60]$ km/h$\times[-45, 45]^\circ$ is divided into 45 equal area sub-rectangles.

Fig. 6 gives these v-a heatmaps (5.6) for two selected drivers. We observe an obvious contrast between these two drivers, the

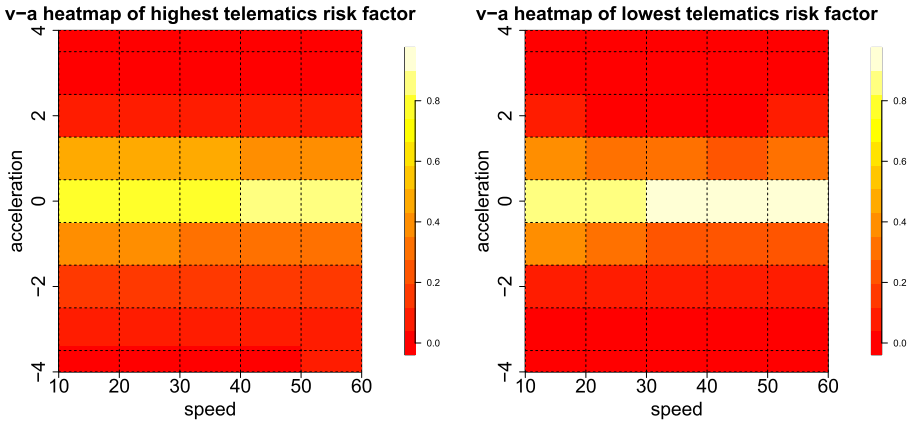

Fig. 6. v-a heatmaps of the two drivers with the highest and lowest telematics risk factors $\rho(\boldsymbol{z})$ in equation (5.7) in the test data; the colors reflect the magnitudes of $z_{i,j,k} \in [0,1]$.

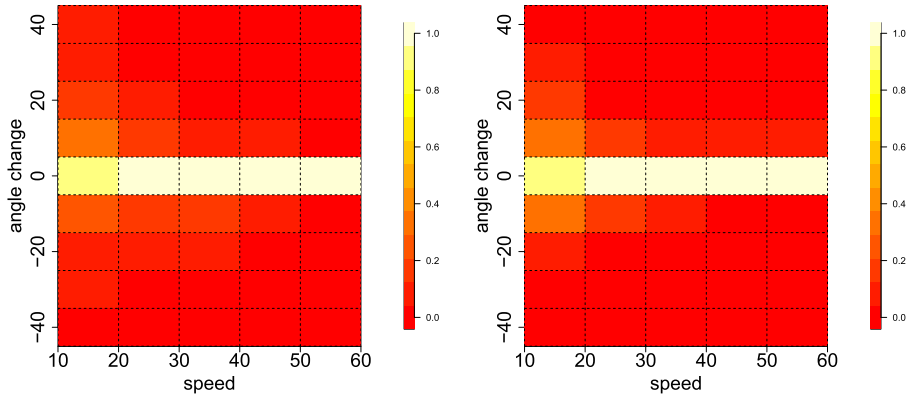

Fig. 7. The corresponding v-Δ heatmaps for the two selected drivers with the v-a heatmaps in Fig. 6; the colors reflect the magnitudes of $u_{i,j,k} \in [0,1]$.

first driver seems to be an aggressive driver having hard acceleration and deceleration, whereas the second driver seems more moderate, especially in the high speed region; the color scale reflects the magnitudes of $z_{i,j,k} \in [0,1]$. Later in Section 5.4, we show that these two drivers have the highest and lowest telematics risk factors, respectively. Fig. 7 gives the corresponding v-Δ heatmaps for the same drivers as in Fig. 6. For the change of direction pattern we do not observe such a strong contrast as in Fig. 6.

5.4. Boosting the GLM with heatmaps

We extend GLM (5.3) by telematics information. We make the following multiplicative assumption

$$\mathbb{E}\left[N_i \middle| \boldsymbol{x}_i, \boldsymbol{z}_i\right] = \mu(\boldsymbol{x}_i, \boldsymbol{z}_i) v_i = \lambda(\boldsymbol{x}_i) v_i \rho(\boldsymbol{z}_i) = \exp\langle \boldsymbol{\beta}, \boldsymbol{x}_i \rangle v_i \rho(\boldsymbol{z}_i),$$

thus, we multiply the GLM term $\lambda(\boldsymbol{x}_i)v_i$ considering classical covariates $\boldsymbol{x}_i$ and exposure v_i with a term $\rho(\boldsymbol{z}_i)$ collecting the telematics information $\boldsymbol{z}_i$. We postulate this multiplicative structure, let us comment on this. Ideally, the classical actuarial covariates $\boldsymbol{x}$ and the telematics covariates $\boldsymbol{z}$ interact in a more sophisticated way than in the multiplicative structure as above. The difficulty here is that we only have data of $n = 1,847$ car drivers, and this limits the choices of regression functions because more complex regression functions cannot be estimated reliably.

In fact, we restrict even more. Namely, we use a two-step fitting strategy by first fitting the GLM (5.3) neglecting telematics information. This gives us an estimated GLM regression parameter $\widehat{\boldsymbol{\beta}} \in \mathbb{R}^{q+1}$. In a second step, we boost this fitted GLM by telematics data $\boldsymbol{z}$, and letting $\widehat{\boldsymbol{\beta}}$ be fixed, thus, we consider regression function in the second step

$$(\boldsymbol{x}, \boldsymbol{z}) \mapsto \mu(\boldsymbol{x}, \boldsymbol{z}) = \widehat{\lambda}(\boldsymbol{x}) \rho(\boldsymbol{z}) = \exp\langle \widehat{\boldsymbol{\beta}}, \boldsymbol{x} \rangle \rho(\boldsymbol{z}), \tag{5.7}$$

where we fit the telematics risk factor $\rho(\boldsymbol{z})$ in this second step by a CNN exploring the v-a heatmap information $\boldsymbol{z}$, and using $\exp\langle \widehat{\boldsymbol{\beta}}, \boldsymbol{x} \rangle$ as a non-trainable offset. The chosen CNN architecture is shown in Listing 2, consisting of two CNN layers ψ^1 and ψ^2, one flatten layer ψ^3 and one fully-connected dense layer ψ^4 providing us with a mapping from the v-a heatmap $\boldsymbol{z}$ to the telematics risk factor

$$\rho : [0,1]^{9\times 5\times 1} \to \mathbb{R}_+, \quad \boldsymbol{z} \mapsto \rho(\boldsymbol{z}) = \left(\psi^4 \circ \psi^3 \circ \psi^2 \circ \psi^1\right)(\boldsymbol{z}); \tag{5.8}$$

we also refer to Section 4.2. We use hyper-parameters $q1 = 8$ and $q2 = 2$ for this architecture. It reduces the $9 \times 5 \times 1$ dimensional input tensor to the one-dimensional telematics risk factor $\rho(\boldsymbol{z}) \in \mathbb{R}_+$.

This model is then calibrated to the available learning data $\mathcal{L}$ where we further split this data to a training data set $\mathcal{U}$ and a validation data set $\mathcal{V}$ to track over-fitting. The fitting is rather fast, taking a couple of seconds on a personal computer. The out-of-sample analysis on the test data $\mathcal{T}$ shows that the telematics risk factor $\rho(\boldsymbol{z})$ improves the generalization error of the GLM from 1.0306 to 1.0128, see model 'C+VA' in Table 3. Thus, these v-a heatmaps $\boldsymbol{z}_i$ clearly contain information beyond the classical covariates $\boldsymbol{x}_i$ that improves the predictive performance of our model.

Listing 2: CNN architecture for boosting the GLM with v-a heatmaps (5.7).

```
build_model_cnn<-function(height,width,q1,q2){
  ### input layer
  heatmap<-layer_input(shape=c(height,width,1),
                       dtype = "float32", name="heatmap")
  vol<-layer_input(shape=c(1),dtype = "float32",name="vol")

  ### convolutional neural network
  Heatmap_Network = heatmap %>%
    layer_conv_2d(filters=q1, kernel_size=c(height,1),
                  activation="tanh", strides = 1, name="heatmap_conv1") %>%
    layer_conv_2d(filters=q2, kernel_size=c(1,1),
                  activation="tanh", strides = 1, name="heatmap_conv2") %>%
    layer_flatten(name="heatmap_flat") %>%
    layer_dense(units=1, activation="exponential", name="heatmap_factor",
                weights=list(array(c(0), dim=c(q2*width,1)), array(0, dim=c(1))))

  ### response
  Response = list(Heatmap_Network,vol) %>%
    layer_multiply (name="response",trainable = F)

  ### compile model
  model<-keras_model(inputs=c(heatmap,vol),outputs=c(Response))
  model %>% compile(optimizer=optimizer_adam(), loss="poisson")
  model
}
```

In a separate analysis we have seen that such a model combining $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$ is better than a model only considering $\boldsymbol{z}_i$ (model 'VA' in Table 3):

$$\boldsymbol{z} \mapsto \mu(\boldsymbol{z}) = \bar{\mu}\rho(\boldsymbol{z}), \tag{5.9}$$

where $\bar{\mu}$ is the estimated claims frequency from the null model (5.4). The reason may be that classical covariate information can explain under which circumstances the telematics data has been collected, e.g., the same driving style may more likely cause an accident downtown than in a rural region.

We interpret the telematics risk factor $\rho(\boldsymbol{z})$. Firstly, the two-step approach of first fitting a GLM and then building the telematics risk factor around this GLM corresponds to the combined actuarial neural network (CANN) model proposed by Wüthrich and Merz (2019). We have used this here for stability reasons, but, beyond that, it allows us to isolate the information from the telematics v-a heatmap $\boldsymbol{z}$. Fig. 8 shows this telematics risk factor $\rho(\boldsymbol{z})$ ranging from 0.6 to 1.6. This decreases or increases the estimated expected frequencies in the range of -40% to $+60\%$ compared to only considering the classical actuarial covariates in a GLM. Of course, this can be seen as experience rating, as we correct initial information (like age and gender) with posterior information summarizing driving experience. Unlike densely connected neural networks, the convolutional neural network is not a black box, and there are many visualization tools to understand how it works, e.g., the class activation map of Selvaraju et al. (2017) can be used. Gao et al. (2022) interpret it by studying the network weights and find that hard braking in low speeds contributes most to a high telematics risk factor.

We further boost model (5.7) with v-Δ heatmaps in a similar fashion, i.e., in a three-step strategy. With estimated $\hat{\lambda}$ and $\hat{\rho}$ in the first two steps, we consider regression function in the third step

$$(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}) \mapsto \mu(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}) = \widehat{\lambda}(\boldsymbol{x})\widehat{\rho}(\boldsymbol{z})\varphi(\boldsymbol{u}), \tag{5.10}$$

where we model the second telematics risk factor $\varphi(\boldsymbol{u})$ by the same CNN architecture as shown in Listing 2. As shown in Table 3 the test error of 0.9982 ('C+VA+VD') is the lowest among the models studied. From this we conclude that the v-Δ heatmaps complements the v-a heatmaps. Finally, we compare with the model which only considers the two telematics heatmaps $\boldsymbol{z}$ and $\boldsymbol{u}$:

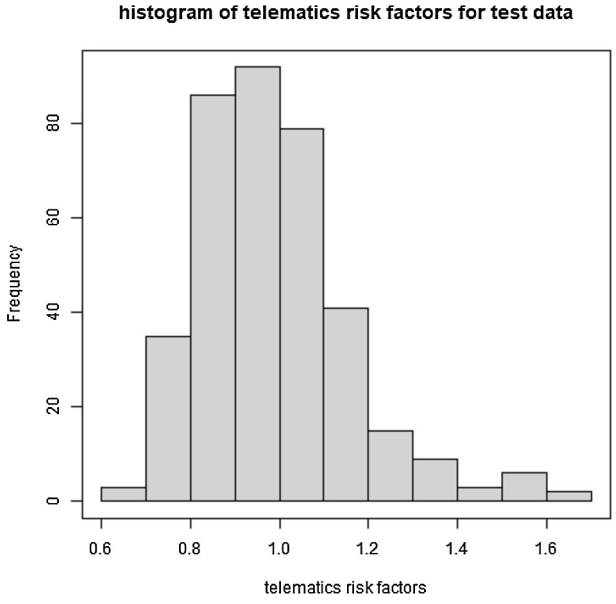

Fig. 8. The distribution of $\rho(\boldsymbol{z})$ on the test data set $\mathcal{T}$.

$$(\boldsymbol{z}, \boldsymbol{u}) \mapsto \mu(\boldsymbol{z}, \boldsymbol{u}) = \bar{\mu}\widehat{\rho}(\boldsymbol{z})\varphi(\boldsymbol{u}). \tag{5.11}$$

The test error of 1.0381 (larger than 0.9982) shown in Table 3 ('VA+VD') indicates that classical actuarial covariates complement telematics heatmaps, and we need both information.

Because our portfolio is very small, we perform a sensitivity analysis. We alternatively calculate the test error for 5 mutual exclusive test data sets, i.e., we perform a sort of $K = 5$ cross-validation. The results are listed in Table 4. Our previous statements and conclusions are supported by this sensitivity analysis. Another observation is that it seems that the improvement due to the v-Δ heatmap is rather weak in 3 out of 5 test data sets; see columns 'VA' and 'VA+VD'. We believe that the partition in v-Δ might not be suitable. A better way is to first investigate their continuous joint distribution and then determine an appropriate discretization.

Table 4
Sensitivity analysis: out-of-sample test error on $\mathcal{T}$ of the models studied.

test index	Null	Classical	C+VA	C+VA+VD	VA	VA+VD	C+T
	(5.4)	(5.3)	(5.7)	(5.10)	(5.9)	(5.11)	(5.15)
1	1.1095	1.0981	1.0981	1.0966	1.1040	1.1040	**1.0918**
2	1.1003	1.0306	1.0128	**0.9982**	1.0616	1.0381	1.0286
3	1.0949	1.0641	1.0431	**1.0330**	1.0625	1.0465	1.0439
4	1.0952	1.0721	1.0675	1.0675	1.0666	1.0666	**1.0654**
5	1.0996	1.0318	**1.0266**	**1.0266**	1.0719	1.0718	1.0268

Remark. We do not incorporate contextual information such as road types, weather, regions, driving times in this predictive model. These predictors can be included in a multivariate compositional form, i.e., the proportion of duration/distance on different road types/weather/regions/daytimes, we refer to Paefgen et al. (2014) and Verbelen et al. (2018) for a thorough analysis. One limitation of heatmaps is that we need sufficient telematics data to obtain a stable representation, and the modeling procedure cannot be easily adjusted for insufficient telematics data. Moreover, we need to assume a stationary situation, because heatmaps react more slowly to changes as they reflect data in aggregated form.

5.5. Other ways of using telematics data

With heatmaps, we aggregate telematics data of multiple trips into a two dimensional distribution. Sometimes, it is desirable to evaluate driving risk associated with each trip, e.g., how phone usage interacts with driving behavior and how driving behavior changes with time. This motivates our next approach where we use time series of individual trips for claims frequency prediction. For a more detailed treatment, we refer to Meng et al. (2022). One major concern with individual trips is that trips are not labeled as risky or safe. Inspired by Section 4, we select 10 archetypal risky drivers $\mathcal{L}_r \subset \mathcal{L}$ who caused the most claims in our learning portfolio, and we select 10 archetypal safe drivers $\mathcal{L}_s \subset \mathcal{L}$ who have the longest exposures without any claims. We label their trips as potentially risky (coded as 1) and potentially safe (coded as 0). More specifically, the j-th trip of archetypal driver $i \in \mathcal{L}_r \cup \mathcal{L}_s$ is a multivariate time series of length $T = 300$ seconds, denoted by

$$\boldsymbol{z}_{i,j} = \Big((v_{i,j,1}, a_{i,j,1}, \Delta_{i,j,1}, a^2_{i,j,1}, \Delta^2_{i,j,1})^\top, \dots, (v_{i,j,T}, a_{i,j,T}, \Delta_{i,j,T}, a^2_{i,j,T}, \Delta^2_{i,j,T})^\top \Big)^\top \in \mathbb{R}^{T\times 5},$$

for $j = 1, \dots, J_i$. These trips are labeled with responses $Y_{i,j} \in \{0, 1\}$ for $i \in \mathcal{L}_r \cup \mathcal{L}_s$ and $j = 1, \dots, J_i$. Compared to (4.2), we have extended from 180 to 300 seconds, and we have added the squares of acceleration $a^2_{i,j,t}$ and change in direction $\Delta^2_{i,j,t}$ to improve model performance; from the universality approximation theorem point of view this may not be necessary, but it improves network fitting on finite samples.

After normalizing the time series $\boldsymbol{z}_{i,j}$ into $[-1, 1]$ with the MinMaxScaler, a one-dimensional CNN is calibrated to classify those binary labeled trips

$$\psi : [-1, 1]^{T\times 5} \to (0, 1), \quad \boldsymbol{z} \mapsto \psi(\boldsymbol{z}). \tag{5.12}$$

The calibrated CNN $\widehat{\psi}$ is then employed to evaluate risk scores of the J_i individual trips for each driver $i = 1, \dots, n$. For computational reasons, the number of evaluated individual trips is constrained to $J_i \le 500$. Based on this, we calculate the average risk score of each driver i

$$\bar{\psi}_i = \frac{1}{J_i} \sum_{j=1}^{J_i} \widehat{\psi}(\boldsymbol{z}_{i,j}), \tag{5.13}$$

which is an estimate of a true risk score ν_i. To reduce volatility, we apply the limited fluctuation credibility model, see Section 17.2 of Klugman et al. (2012). We quantify the credibility of the average risk scores (5.13). The standard of full credibility (in terms of numbers of individual trips) for average risk score $\bar{\psi}_i$ is computed as

$$J_i^{(f)}(\alpha, r) = \left(\frac{\Phi^{-1}(1-\alpha/2)}{r} \right)^2 \left(\frac{\sigma_i}{\nu_i} \right)^2, \tag{5.14}$$

where Φ^{-1} is the inverse of standard normal distribution function. We select significance level $\alpha = 10\%$ and fluctuation level $r = 10\%$, and we estimate the standard deviation σ_i and the mean ν_i by the sample deviation and sample mean of $(\widehat{\psi}(\boldsymbol{z}_{i,j}))_{j=1:J_i}$, respectively. If the number of individual trips J_i for driver i does not meet this minimal standard of full credibility (5.14), i.e., $J_i < J_i^{(f)}$, we use the following credibility average risk score instead

$$\widetilde{\psi}_i = Z_i \bar{\psi}_i + (1 - Z_i)\hat{\nu},$$

where

$$Z_i = \min\left(1, \sqrt{\frac{J_i}{J_i^{(f)}}}\right)$$

is the partial credibility, see Section 17.4 of Klugman et al. (2012), and $\hat{\nu}$ is the estimated overall risk score $\hat{\nu} = 1/n \sum_{i=1}^n \bar{\psi}_i$. We draw the histogram of the standard of full credibility $(J_i^{(f)})_{i=1:n}$ and the histograms of the partial credibility Z_i in Fig. 9, which indicates that most drivers meet the standard of full credibility $J_i^{(f)}$. Indeed, there are $1,816$ out of $n = 1,847$ drivers satisfying the standard of full credibility. The averaged standard of full credibility is around 200 individual trips that is comparable to the required minimal volume of three months' telematics data for stable v-a heatmaps (in average two trips per day); see Gao et al. (2019a). Remark that by this credibility approach we do not require a minimal volume of telematics data.

We select for driver i telematics information $\boldsymbol{z}_i = (\boldsymbol{z}_{i,j})_{1\le j\le J_i}$. We incorporate the credibility average risk score into the Poisson regression model (5.1) with the regression function given by

$$(\boldsymbol{x}_i, \boldsymbol{z}_i) \;\mapsto\; \mu(\boldsymbol{x}_i, \boldsymbol{z}_i) = \widehat{\lambda}(\boldsymbol{x}_i)\rho(\boldsymbol{z}_i) = \exp\langle \widehat{\boldsymbol{\beta}}, \boldsymbol{x}_i \rangle \rho(\boldsymbol{z}_i), \tag{5.15}$$

where $\widehat{\boldsymbol{\beta}}$ is the GLM estimate in (5.3), and the telematics risk factor $\rho(\boldsymbol{z}_i)$ is related to the credibility average risk score defined by

$$\log \rho(\boldsymbol{z}_i) = \alpha_0 + \alpha_1 \tilde{\phi}_i. \tag{5.16}$$

The parameters α_0 and α_1 are estimated in a Poisson GLM with offset $v_i \widehat{\lambda}(\boldsymbol{x}_i)$ on the learning data set $\mathcal{L}$. The resulting out-of-sample Poisson deviance loss is 1.0286; see column 'C+T' of Table 3. It is better than the GLM only considering classical covariates (1.0306), but worse than in the model based on the v-a heatmaps (1.0128). Remark that, although we consider one additional telematics variable of direction change, we get a worse out-of-sample prediction than the model based on the v-a heatmap. However, this comparison is not conclusive since we have a rather small portfolio, and because selection of archetypal drivers needs more exploration. In Table 4 on the sensitivity analysis, 3 out of 5 test errors for model (5.15) are comparable to model (5.10) considering both classical covariates and two heatmaps, indicating that the time series approach is not necessarily worse than the heatmap approach.

Remark. We have investigated how individual trip risk score changes over time for each archetypal driver, and there is no obvious change or trend for any of the selected archetypal drivers. Another interesting metric is the standard deviation of individual trip risk scores, and we do not observe any change or trend in their standard deviations either. This may be due to a relatively short period. Also most drivers are experienced mature drivers,

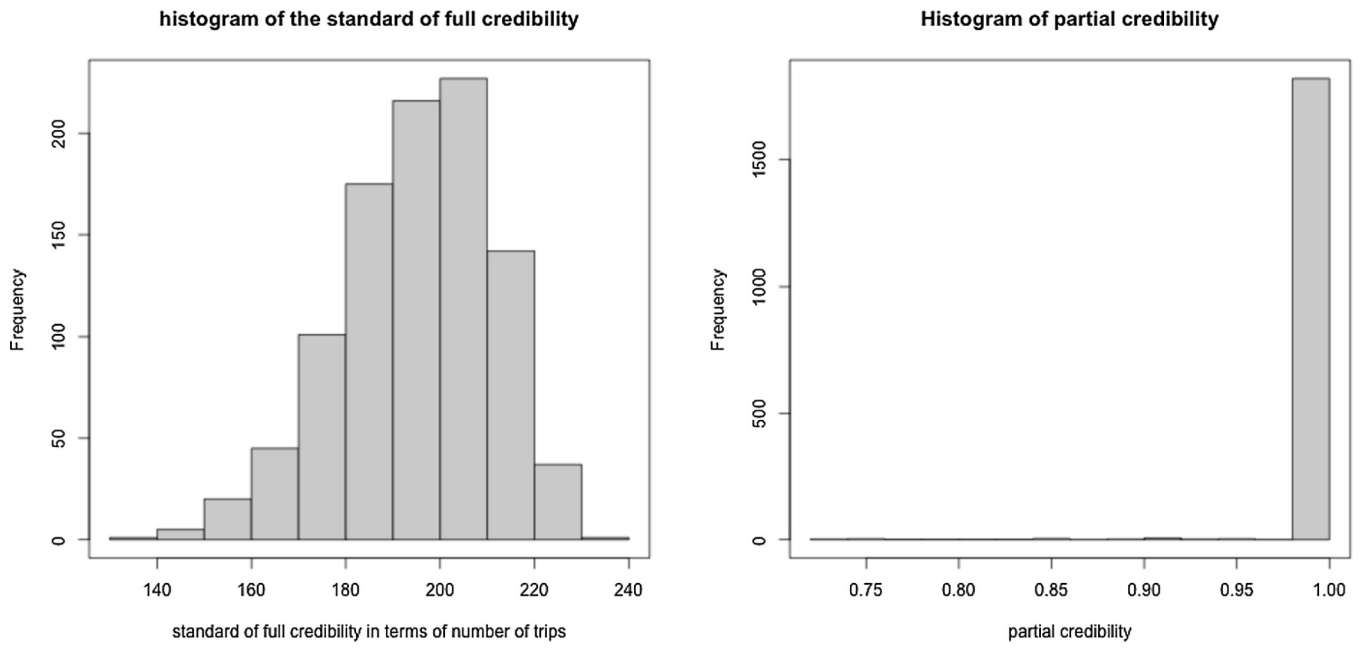

Fig. 9. The standard of full credibility $(J_i)_{i=1:n}^{(f)}$ and the histogram of partial credibility Z_i.

and their driving skills may be already stabilized. We do not have contextual information such as phone usage, weather, traffic flow, road type for our data. If this information was available (e.g. in smart phone-based telematics data), one could investigate their interaction with driving behavior.

5.6. Summary

We have presented two different ways of making telematics time series data useful for claims prediction. In the first approach we have aggregated telematics time series data into telematics heatmaps, and statistical modeling has been exploring these aggregated statistics. In the second approach we have scored each trip individually, and statistical modeling has been exploring these individual scores for claims prediction. Both ways aim at reducing the complexity and size of the telematics data in a first step, before entering the regression model. Since this first step of information extraction may not be optimal, the selected statistics may focus on wrong properties in the data. Therefore, we can think of a third way of directly working on the raw telematics data and letting the machine learning methods perform representation learning. This sounds very appealing, however, it may be computationally too demanding because this will require that many TBs of data are processed simultaneously by the machine learning model. This is still out of scope at the current stage of model development and computational power, for this reason, we rely on the first two proposals.

6. Outlook

We have discussed two aspects of telematics data: transparency and its usefulness in claims frequency prediction. We mention possible extensions. One can construct other aggregate statistics, and use them to improve claims frequency prediction. Such approaches are based on first aggregating telematics information, and then working on the aggregated data, benefiting from the law of large numbers during aggregation. If we work with granular telematics data, we could also use recurrent neural networks for scoring individual trips of varying lengths. The law of large numbers then only comes into play in the next step where scores are aggregated in a regression model for claims frequency prediction.

In our approach we have paid attention to explainability by introducing a telematics risk factor $\rho(\boldsymbol{z})$ that can be interpreted and explained to management and customers. In recent machine learning research quite some efforts have been made in making machine learning solutions explainable. A different approach that has recently been taken by Richman and Wüthrich (2021) is to select an explainable network architecture in the first place. Besides explainability, this architecture also supports variable selection. Based on such a transparent network architecture it will become feasible to identify different driving features and behaviors that may help to improve driving styles. Moreover, we did not study any temporal component, e.g., how driving styles of young drivers change when gaining more driving experience. In this context, also driving assistance tools play an increasingly important role in driving safety, potentially making telematics data non-stationary which, of course, provides a bigger challenge on the statistical side.

Our telematics data is received by the devices installed in cars. In practice, many insurance companies prefer to use smart phone-based telematics data considering the expense of data collection. On the one hand, this will impose more challenges on the data cleaning side. On the other hand, smart phone-based data contains more information such as road and weather condition, smart phone use, etc. Although not all of these variables can be used for insure pricing, they are related to accidents. Similar to Section 5.5, we could score each trip, extracting all the associated risk factors, and giving instant feedback to drivers through smart phones. This would promote driving safety and ultimately reduce the claim payments.

Declaration of competing interest

There is no competing interest.

Acknowledgement

Guangyuan Gao gratefully acknowledges financial support from the National Natural Science Foundation of China (71901207).

Appendix A. An excerpt of telematics car driving data

Table 5
Excerpt of telematics car driving data in the first seconds; for privacy reasons, we have removed the vehicle identification number and normalized GPS coordinates.

Time_Stamp	GPS_Latitude	GPS_Longitude	GPS_Heading	GPS_Speed	Positional_Quality	VSS_Speed	Engine_RPM	Accel_Lateral	Accel_Longitudinal	Accel_Vertical
1435622463					0	12	1419	-0.1	-0.2	9.6
1435622465					0	12	1419	-0.1	-1	9.6
1435622466					0	10	1152	-0.3	-0.7	9.5
1435622467					0	8	936	-0.1	-0.3	9.5
1435622468					0	8	936	-0.2	-0.6	9.7
1435622469					0	7	853	-0.2	0	9.6
1435622470					0	9	1287	-0.3	0.5	10.3
1435622471	0	0	170.6	7.3	1	15	1908	0.1	0.4	9.7
1435622472	-3.2e-05	8e-06	167.69	9.8	1	15	1908	-0.5	-0.7	9.5
1435622473	0.000304	-0.00104	149.19	14.3	1	14	1049	-0.1	0.3	8
1435622474					0	14	1049	-0.1	0.3	8
1435622475					0	15	1111	0	0.6	8.5
1435622476					0	17	1196	0.2	-0.6	9.6
1435622477					0	15	1079	0.1	-0.9	10
1435622478					0	15	1079	0	-0.6	9.7
1435622479	-9.6e-05	-0.00088	153.19	15.3	1	14	1007	0.5	-0.6	9.8
1435622480	-0.0001344	-0.00084	155.5	14.3	1	11	822	0.8	-0.4	9.4
1435622481	-0.0001856	-0.00084	159	12.5	1	11	822	1.5	0	9.6
1435622482	-0.0002688	-0.000584	176.3	11.1	1	13	1006	2.3	0.5	9.7
1435622483	-0.0002976	-6e-04	197.8	12.5	1	19	1389	1.1	0.5	9.6
1435622484	-0.00032	-0.000632	219.19	14.1	1	19	1389	0.6	0.7	9.6
1435622485	-0.0003456	-0.00068	228	16.3	1	24	1798	0	0.5	9.7
1435622486	-0.0003744	-0.000736	233.1	20.5	1	27	1811	0	-0.3	9.4
1435622487	-0.0005504	-0.000448	241.1	24.1	1	27	1811	0	0.1	9.7
1435622488	-0.0005792	-0.000528	241.89	27.1	1	30	1477	0	0.1	9.7
1435622489	-0.0006336	-0.000576	243.5	30.3	1	30	1477	0	0.7	9.7
1435622490	-0.0006848	-0.000648	243.6	32.5	1	32	1623	0	-0.3	9.6
1435622491	-0.0007232	-0.000744	243.3	34	1	35	1678	-0.1	-0.6	9.6
1435622492	-0.0008032	-0.000776	242.89	34	1	35	1678	0.1	-0.6	9.7
1435622493	-0.0008544	-0.000856	243.69	34.4	1	35	1672	0	-0.5	9.6

References

Ayuso, M., Guillén, M., Nielsen, J.P., 2019. Improving automobile insurance ratemaking using telematics: incorporating mileage and driver behaviour data. Transportation 46, 735–752.

Ayuso, M., Guillén, M., Pérez-Marín, A.M., 2016a. Telematics and gender discrimination: some usage-based evidence on whether men's risk of accidents differs from women's. Risks 4 (2), 10.

Ayuso, M., Guillén, M., Pérez-Marín, A.M., 2016b. Using GPS data to analyse the distance traveled to the first accident at fault in pay-as-you-drive insurance. Transportation Research. Part C, Emerging Technologies 68, 160–167.

Bayat, S., Babulal, G.M., Schindler, S.E., Fagan, A.M., Morris, J.C., Mihailidis, A., Roe, C.M., 2021. GPS driving: a digital biomarker for preclinical Alzheimer disease. Alzheimer's Research & Therapy 13, 115.

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

Boucher, J.-P., Pérez-Marín, A.M., Santolino, M., 2013. Pay-as-you-drive insurance: the effect of the kilometers on the risk of accident. Anales del Instituto de Actuarios Espanoles 19, 135–154.

Denuit, M., Guillén, M., Trufin, J., 2019. Multivariate credibility modelling for usage-based motor insurance pricing with behavioural data. Annals of Actuarial Science 13 (2), 378–399.

Duval, F., Boucher, J.-P., Pigeon, M., 2021. How much telematics information do insurers need for claim classification. arXiv:2105.14055v1.

Eling, M., Kraft, M., 2020. The impact of telematics on the insurability of risks. The Journal of Risk Finance 21 (2), 77–109.

Esteves-Booth, A., Muneer, T., Kirby, H., Kubie, J., Hunter, J., 2001. The measurement of vehicular driving cycle within the city of Edinburgh. Transportation Research. Part D, Transport and Environment 6 (3), 209–220.

Gao, G., Meng, S., Wüthrich, M.V., 2019a. Claims frequency modeling using telematics car driving data. Scandinavian Actuarial Journal 2019 (2), 143–162.

Gao, G., Wang, H., Wüthrich, M.V., 2022. Boosting Poisson regression models with telematics car driving data. Machine Learning 111 (1), 243–272.

Gao, G., Wüthrich, M.V., 2018. Feature extraction from telematics car driving heatmaps. European Actuarial Journal 8 (2), 383–406.

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

Gao, G., Wüthrich, M.V., Yang, H., 2019b. Evaluation of driving risk at different speeds. Insurance. Mathematics & Economics 88, 108–119.

Geyer, A., Kremslehner, D., Muermann, A., 2020. Asymmetric information in automobile insurance: evidence from driving behavior. The Journal of Risk and Insurance 87 (4), 969–995.

Gneiting, T., 2011. Making and evaluating point forecasts. Journal of the American Statistical Association 106 (494), 746–762.

Gneiting, T., Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association 102 (477), 359–378.

Guillén, M., Nielsen, J.P., Pérez-Marín, A.M., Elpidorou, V., 2020. Can automobile insurance telematics predict the risk of near-miss events? North American Actuarial Journal 24 (1), 22–34.

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

Ho, S.-H., Wong, Y.-D., Chang, V.W.-C., 2014. Developing Singapore driving cycle for passenger cars to estimate fuel consumption and vehicular emissions. Atmospheric Environment 97, 353–362.

Hu, X., Zhu, X., Ma, Y.L., Chiu, Y.C., Tang, Q., 2019. Advancing usage-based insurance – a contextual driving risk modelling and analysis approach. IET Intelligent Transport Systems 13 (3), 453–460.

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

Hung, W.T., Tong, H.Y., Lee, C.P., Ha, K., Pao, L.Y., 2007. Development of practical driving cycle construction methodology: a case study in Hong Kong. Transportation Research. Part D, Transport and Environment 12 (2), 115–128.

Joubert, J.W., De Beer, D., De Koker, N., 2016. Combining accelerometer data and contextual variables to evaluate the risk of driver behaviour. Transportation Research. Part F, Traffic Psychology and Behaviour 41, 80–96.

Kamble, S.H., Mathew, T.V., Sharma, G.K., 2009. Development of real-world driving cycle: case study of Pune, India. Transportation Research. Part D, Transport and Environment 14 (2), 132–140.

Klugman, S.A., Panjer, H.H., Willmot, G.E., 2012. Loss Models: From Data to Decisions. John Wiley & Sons.

Krüger, F., Ziegel, J.F., 2021. Generic conditions for forecast dominance. Journal of Business & Economics Statistics 39 (4), 972–983.

Lemaire, J., Park, S.C., Wang, K., 2016. The use of annual mileage as a rating variable. ASTIN Bulletin 46 (1), 39–69.

Ma, Y.L., Zhu, X., Hu, X., Chiu, Y.C., 2018. The use of context-sensitive insurance telematics data in auto insurance rate making. Transportation Research. Part A, Policy and Practice 113, 243–258.

Meng, S., Wang, H., Shi, Y., Gao, G., 2022. Improving automobile insurance claims frequency prediction with telematics car driving data. ASTIN Bulletin: The Journal of the IAA, 1–29. https://doi.org/10.1017/asb.2021.35.

Paefgen, J., Staake, T., Fleisch, E., 2014. Multivariate exposure modeling of accident risk: insights from pay-as-you-drive insurance data. Transportation Research. Part A, Policy and Practice 61, 27–40.

Richman, R., Wüthrich, M.V., 2021. LocalGLMnet: interpretable deep learning for tabular data. arXiv:2107.11059.

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad-cam: visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 618–626.

So, B., Boucher, J.-P., Valdez, E.A., 2021a. Synthetic dataset generation of driver telematics. Risks 9 (4), 58.

So, B., Boucher, J.-P., Valdez, E.A., 2021b. Cost-sensitive multi-class AdaBoost for understanding behavior based on telematics. ASTIN Bulletin 51, 719–751.

Sun, S., Bi, J., Guillén, M., Pérez-Marín, A.M., 2020. Assessing driving risk using internet of vehicles data: an analysis based on generalized linear models. Sensors 20 (9), 2712.

Verbelen, R., Antonio, K., Claeskens, G., 2018. Unraveling the predictive power of telematics data in car insurance pricing. Journal of the Royal Statistical Society. Series C. Applied Statistics 67, 1275–1304.

Wahlström, J., Skog, I., Händel, P., 2015. Detection of dangerous cornering in GNSS-data-driven insurance telematics. IEEE Transactions on Intelligent Transportation Systems 16 (6), 3073–3083.

Wang, Q., Huo, H., He, K., Yao, Z., Zhang, Q., 2008. Characterization of vehicle driving patterns and development of driving cycles in Chinese cities. Transportation Research. Part D, Transport and Environment 13 (5), 289–297.
Weidner, W., Transchel, F.W.G., Weidner, R., 2016. Classification of scale-sensitive telematic observables for riskindividual pricing. European Actuarial Journal 6 (1), 3–24.
Weidner, W., Transchel, F.W.G., Weidner, R., 2017. Telematic driving profile classification in car insurance pricing. Annals of Actuarial Science 11 (2), 213–236.
Wiatowski, T., Bölcskei, H., 2018. A mathematical theory of deep convolutional neural networks for feature extraction. IEEE Transactions on Information Theory 64 (3), 1845–1866.
Wüthrich, M.V., 2017. Covariate selection from telematics car driving data. European Actuarial Journal 7 (1), 89–108.
Wüthrich, M.V., Merz, M., 2019. Editorial: yes, we CANN! ASTIN Bulletin 49 (1), 1–3.
Wüthrich, M.V., Merz, M., 2021. Statistical foundations of actuarial learning and its applications. SSRN, 3822407.
Zhu, R., Wüthrich, M.V., 2021. Clustering driving styles via image processing. Annals of Actuarial Science 15 (2), 276–290.