IEEE ITS Italian Chapter - PhD Dissertation Award 2022

A conflict-based approach for road safety analysis

Federico Orsini

Department of Civil, Environmental and Architectural Engineering – University of Padova, Italy

Università degli Studi di Padova

Introduction

Traditionally, road safety analysis relies on the use of **crash data**. However, **several issues** may affect these data: lack of availability, lack of spatial/temporal precision, under-reporting, misclassification; moreover, crashes are relatively rare events, so data must be collected for several years and/or in several different locations to obtain enough data. An **alternative approach** consists of analyzing **traffic conflicts.**

Emerging ITS and sensing technologies allow the collection of large amounts of high-quality traffic data, from which it is possible to reliably identify and quantify traffic conflicts.

What is a traffic conflict?

- Intuitively, it is a **near-crash**
- Formally, "an observable situation in which two or more road users approach each other in space and time to such an extent that there is a **risk of collision if their movements remain unchanged**" (Güttinger, 1984)
- Traffic conflicts are "**precursors of crashes** and not alternative outcomes" (Tarko, 2019)

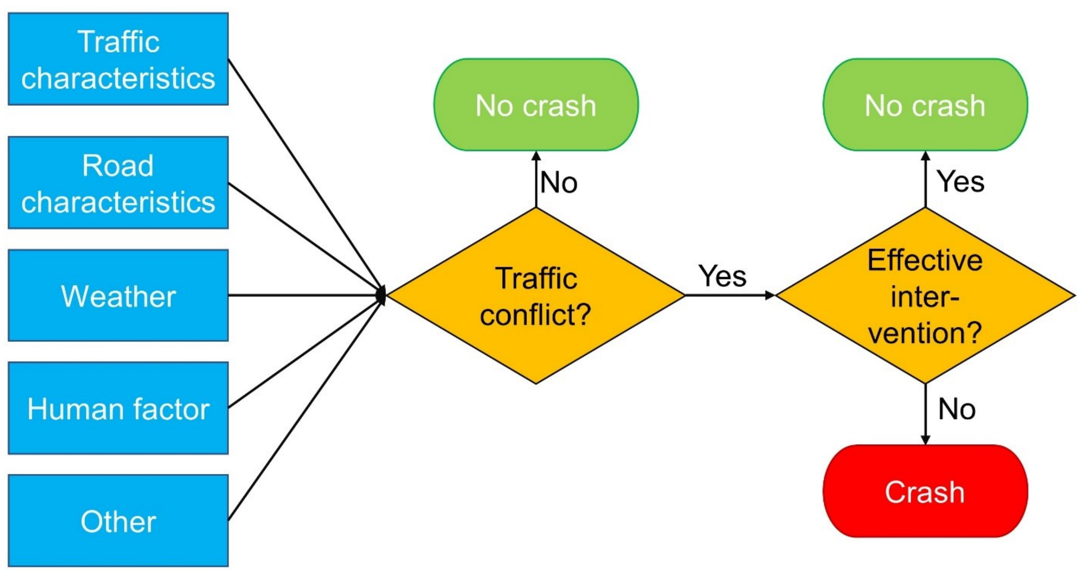

Figure 1. Causality model of traffic conflicts and crashes (adapted from Tarko, 2019)

How to identify and quantify a traffic conflict?

- *Qualitative approach*: observing evasive maneuvers (but it could be rather subjective, it is difficult to assess the severity)
- *Quantitative approach*: **surrogate measures of safety** (i.e., proximity indicators)

What is the relationship between conflicts and crashes?

- ~~Count models~~
- Probabilistic theory

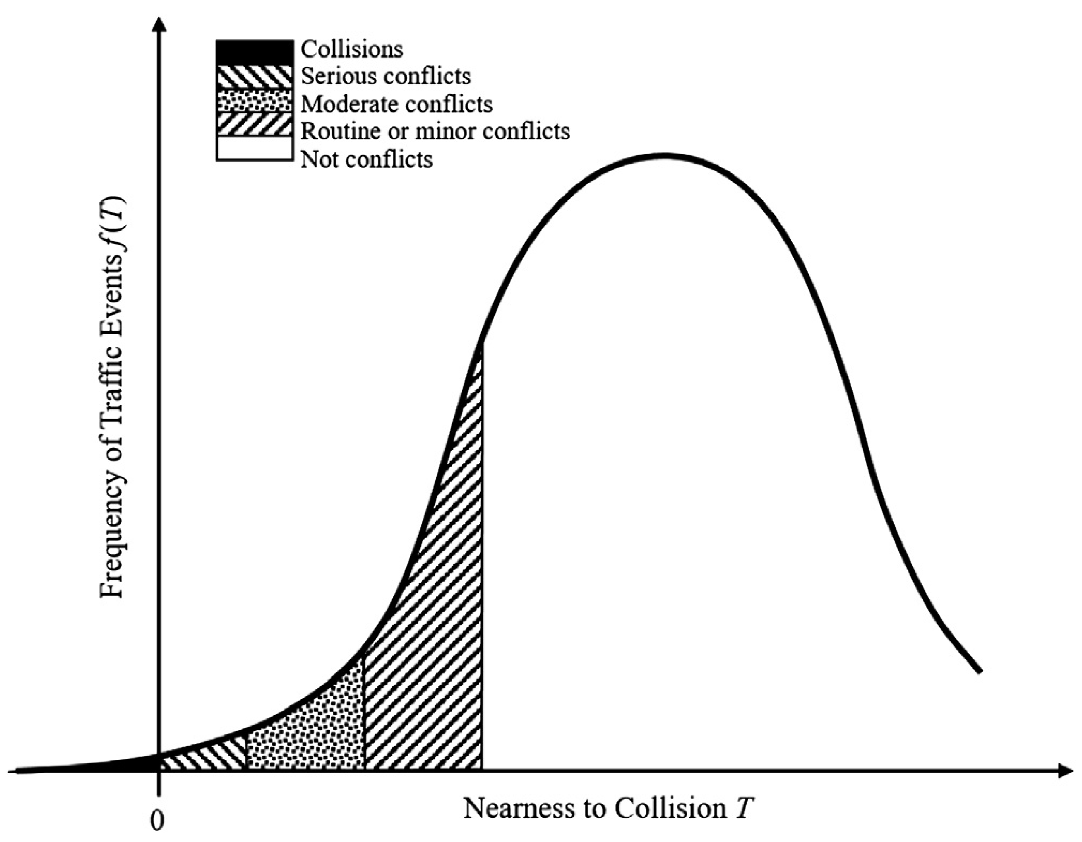

Figure 2. Illustration of the concept of continuous distribution of crash nearness. Source: Tarko (2019), adapted from Glauz & Migletz (1980).

Research objectives

Despite an ever-growing interest toward traffic conflicts in transportation research, there are still several open questions that this research aims to answer, for both **long-term** and **real-time** road safety applications.

1) *How to model the probabilistic relationship between traffic conflicts and crashes?*
2) *Is it possible to predict crashes in real-time with a conflict-based approach?*

Long-term road safety analysis

The first part of the dissertation aims to provide insights on the probabilistic relationship between traffic conflicts and crashes, by applying **extreme value theory**, for long-term road safety analysis (i.e., prediction of annual crash rates in selected infrastructures).

Extreme Value Theory

Extreme value theory (EVT) is a branch of statistics which deals with the extreme deviations from the median of probability distributions. It seeks to assess, from a given ordered sample of a given random variable, **the probability of events that are more extreme than any previously observed.** There are two main approaches in EVT:

BLOCK MAXIMA (BM)

- Observations are aggregated into homogeneous time or space intervals (the blocks)
- The highest value in each block is considered an extreme event and is sampled
- The **block maxima** are used to estimate Generalized Extreme Value (GEV) distribution parameters

PEAK-OVER-THRESHOLD (POT)

- A threshold is defined
- Every observation whose value is higher than the threshold is sampled
- The **exceedances over the threshold** are used to estimate Generalized Pareto (GP) distribution parameters

REAL-WORLD CASE STUDY

Data collection – setup

- Data collected on a 150km-long 3-lane Italian motorway
- For one year, from January 1st to December 31st, 2013.
- 19 cross-sections equipped with micro-wave Doppler radars.

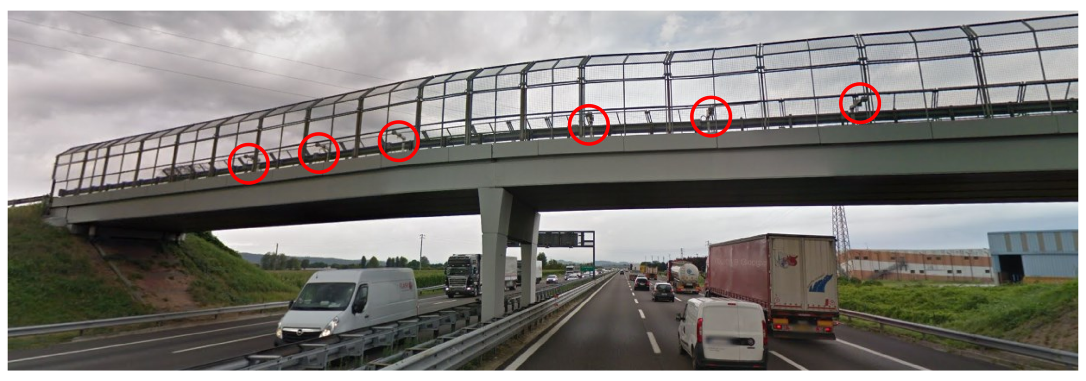

Figure 4. Data collection setup. One radar was positioned above each lane of the cross-section

Data collection – data for model estimation & validation

Vehicle-by-vehicle data collected: speed, time gap, time stamp, vehicle length.

With these data it is possible to calculate a surrogate measure of safety, Time-To-Collision (TTC), between each couple of vehicles consecutively detected on the same lane, with the classic definition:

$$TTC = \frac{R}{RR} = \frac{v_L * t_{GAP}}{v_F - v_L}$$

Crash counts collected for 5 years (2011-2016, except 2013) were available for model validation

Results

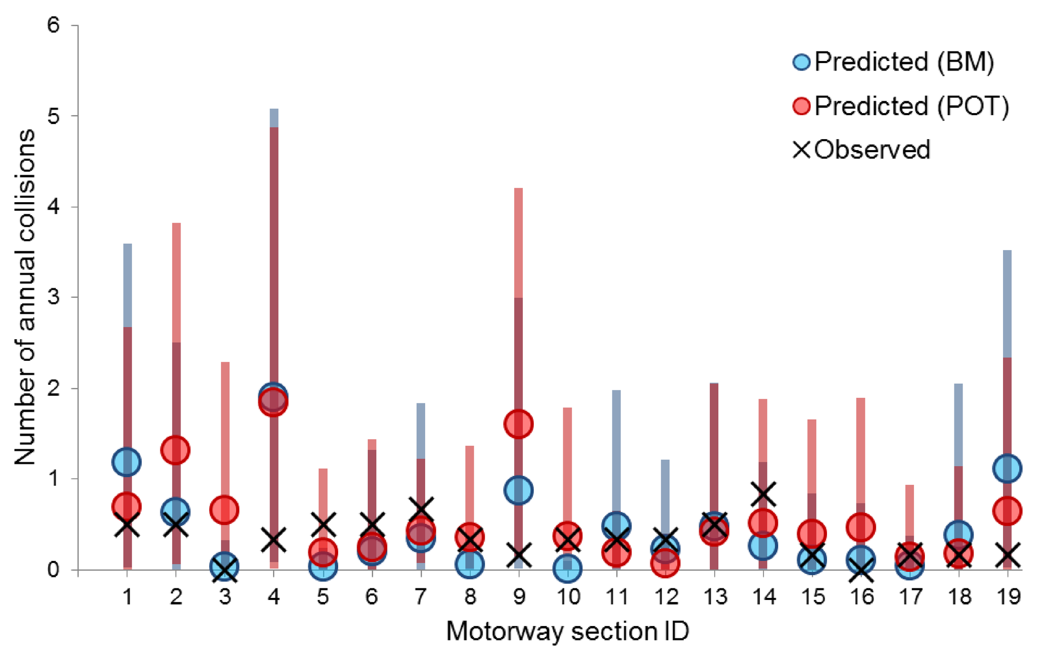

Figure 5. Comparison between predicted and observed (within a 500-meter segment) annual number of collisions, for each motorway section. Blue and red bars represent respectively predicted BM and POT 95% confidence intervals.

- Compared to previous works, significantly better performance
- Observed number of crash is within confidence interval in almost all sections
- In some section BM and POT produce different results, but the overall performance is similar
- POT is able to extract information from more data; this can be crucial for shorter observation periods

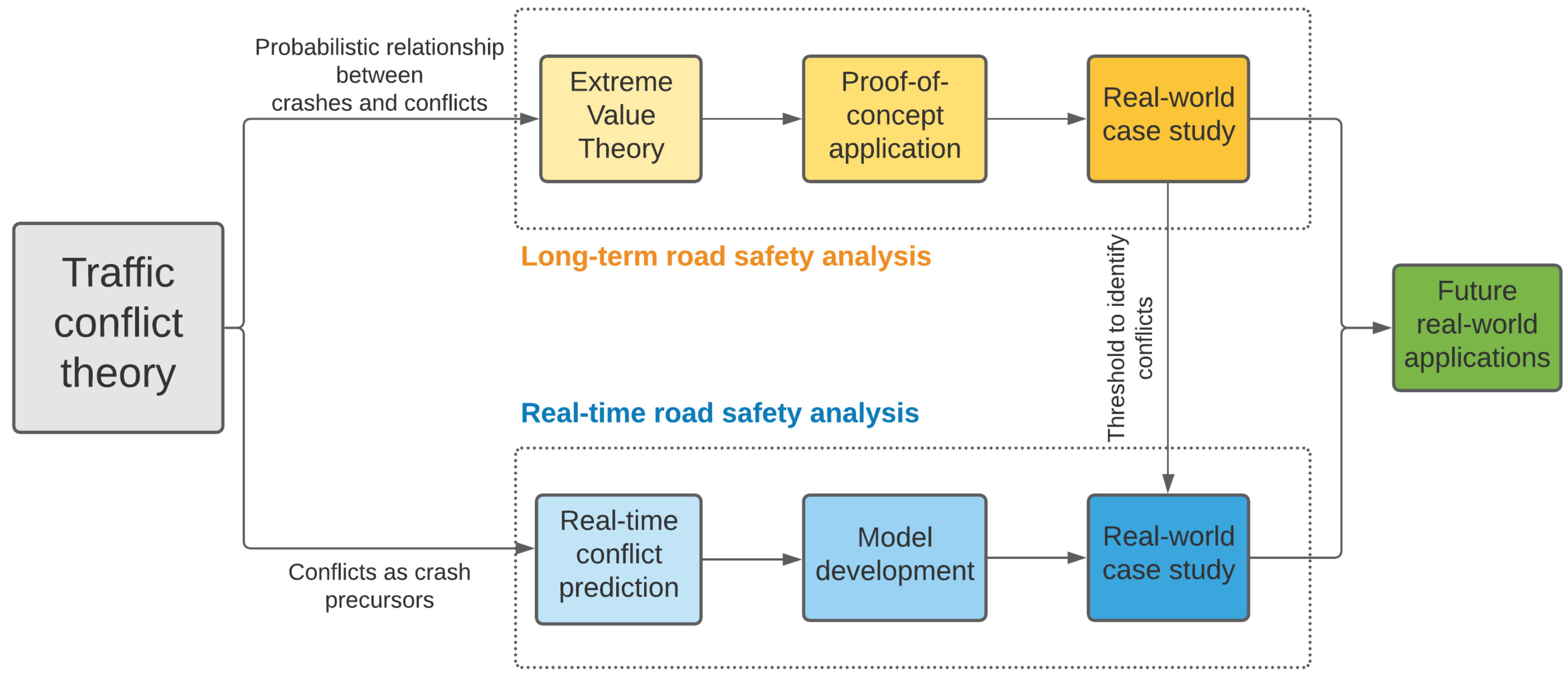

Figure 3. Outline of the dissertation

Real-time road safety analysis

The second part focuses on developing a conflict-based approach for real-time road safety analysis. A **real-time conflict prediction model (RTConfPM)** is structured similarly to a real-time crash prediction model (RTCPM):

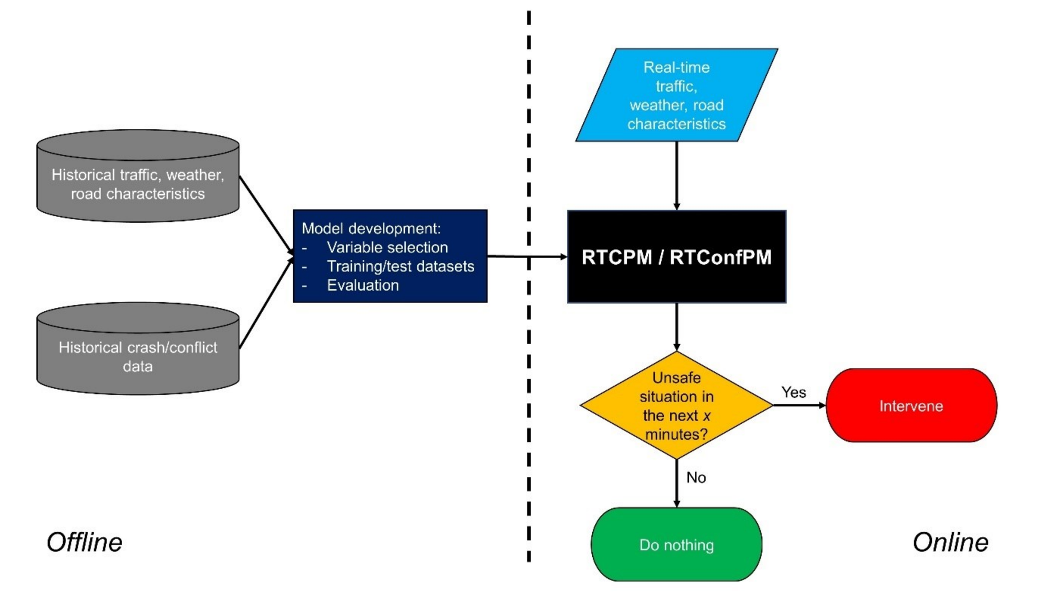

Figure 6. Flowchart representing the offline development of an RTCPM/RTConfPM (left) and the real-time application of an RTCPM/RTConfPM.

REAL-WORLD CASE STUDY

Data collection – predictor variables

The same highway dataset was used. For each lane, vehicle-by-vehicle **data were aggregated in 5-minute intervals**, providing the following predictor variables: traffic volume, heavy-duty traffic volume, speed, speed variance; a total of 12 candidate predictors

Data collection – response variable: safe vs. unsafe

Surrogate measure of safety: TTC.

In order to define a time interval as "unsafe":

- A **TTC threshold *u*** is chosen to identify conflicts
  → Extreme value theory application, *u* = 0.78 s
- A **minimum number of conflicts *n*** within the interval in chosen
  → sensitivity analysis on predictive performance, *n* = 3

Model development

1. A variable selection method is applied, in order to identify potentially irrelevant variables that can be discarded by the model (**Random Forest**).
2. The training dataset is formed taking into account unbalanced classification issues (**SMOTE**).
3. A classifier is trained and tested (**Support vector machine**)
4. Several performance indicators are used to evaluate the classifier **(Recall, Specificity, AUC**).

Results

RTConfPM (trained with conflict data, to predict unsafe situations)

Full-test dataset

Kernel function	Input variables	Accuracy	Recall	Specificity	False alarm rate	AUC
	Set #1	0.934	0.983	0.933	0.067	0.948
Linear	Set #2	0.933	0.981	0.933	0.067	0.947
	Set #3	0.600	0.978	0.600	0.400	0.942
	Set #1	0.958	0.945	0.958	0.042	0.972
Radial basis	Set #2	0.959	0.947	0.959	0.041	0.973
	Set #3	0.639	0.904	0.639	0.361	0.966

RTCPM (trained with crash data, to predict crashes)

Full-test dataset

Kernel function	Input variables	Accuracy	Recall	Specificity	False alarm rate	AUC
	Set #1	0.788	0.395	0.788	0.212	0.612
Linear	Set #2	0.795	0.529	0.795	0.205	0.682
	Set #3	0.175	0.824	0.175	0.825	0.583
	Set #1	0.899	0.219	0.899	0.101	0.615
Radial basis	Set #2	0.875	0.177	0.875	0.125	0.569
	Set #3	0.485	0.353	0.485	0.515	0.537

Direct comparison: RTConfPM to predict crashes
Recall +11.8%
Specificity +13.8%
AUC +9.3%

Conclusion

The application of conflict-based approaches has the potential to provide a positive broader impact on road safety. In particular, the possibility to avoid the use of crash data in practical applications can:

1. allow to **apply statistical methods to new scenarios** in which crash data are unavailable or unreliable (e.g., rural roads, third-world countries, new infrastructures);
2. **proactively analyse safety**, avoiding the ethical dilemma of crash-based approaches, in which injuries and fatalities are needed in order to correctly identify unsafe behaviors and locations;
3. provide **faster road-safety evaluations**, since traffic conflicts are more frequent than crashes; for the same reason, provide more flexible and resilient road safety models, which can be based on more recent data.

Finally, the real-time predictive models developed in this work can be a crucial starting point for the application of **ITS-based intervention strategies** aimed at prevent crashes.