Accident Analysis and Prevention 123 (2019) 356–364

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Coherence assessment of accident database kinematic data

Dario Vangi, Michelangelo-Santo Gulino*, Carlo Cialdai

Università degli Studi di Firenze, Department of Industrial Engineering, Via di Santa Marta 3, 50139, Florence, Italy

ARTICLE INFO

Keywords:
In-depth accident database
Momentum conservation
Velocity triangles
Deformation energy
Impact severity
Injury risk

ABSTRACT

The analysis and research of accidents aimed at improving the safety of vehicles and infrastructures are typically based on the retrospective investigation of data that are collected in in-depth accident databases. In particular, kinematic data related to accidents (impact velocity, velocity change of the vehicles, etc.) make possible the identification of correlations between impact severity and injury risk (IR), as well as assessing the effectiveness of vehicle protection systems. The necessary condition to conduct suitable and significant analyses is to utilise data which are correct and representative of national statistics, i.e., congruent with physical laws governing the accident phenomena. Whereas representativeness can generally be retrospectively verified, the checks on kinematic data coherence during codification are rarely performed.

The present work describes a procedure to verify the internal coherence of kinematic data collected in in-depth accident databases. The introduced checks allow the identification of parameters, which are not internally coherent because the accident reconstruction model employed is inappropriate or improperly used. These checks pertain to physical laws on which road accident reconstruction is based, i.e., momentum conservation, compatibility of velocity triangles, and energy conservation. Moreover, they can be modified and expanded to consider other parameters, making the methodology virtually applicable to any database.

In the case of vehicle-to-vehicle collisions, the application of the procedure to detect incongruent data inside two real databases demonstrates how their number is often not negligible. Furthermore, consequences can be substantial for both direct and secondary analysis, i.e., determining IR curves (for example, logistic regression on input data) and identifying IR associated to an accident. Accordingly, the application of checks is particularly recommended during both analysis and collection phases to confirm the congruence of collected data; consequently, the quality of investigation is enhanced.

1. Introduction

Actions undertaken to increase road safety can be schematically included in the following categories: accident number reduction, impact severity reduction, and injury risk (IR) reduction (Kullgren, 2008). For example, the present trend towards autonomous driving (and the consequent decrease in the human factor influence) promises to drastically reduce the number of accidents and their severity, all at once, through the activation of advanced driving functions, such as autonomous emergency braking (Cicchino, 2017). On the other hand, at equal crash severity, the development of passive safety systems reduces IR, which protects vehicle occupants (Mendez et al., 2010). Thus, in general, the development of integrated safety systems enhances road safety (Burnett et al., 2004). An exact quantification of such improvements requires a retrospective analysis of real-world accident data (Flannagan et al., 2018), whereas a prospective data analysis is frequently required to perfect the advanced safety system design.

Thus, the use of accident data is of primary importance for the enhancement of road safety. A typical requirement in the significant use of databases is that the contained data must be representative of the national statistics (McDonald et al., 2014); that is, the database must reflect the real proportion of accidents based on their characteristics (type, site, time, drivers, passengers, vehicles, etc.). Eventually, this is generally verified during data codification and input into the database. Nevertheless, representativeness is not a strict requirement if a specific analysis aimed at a specific risk factor identification or IR curve definition (e.g., vehicle–bicycle impacts, vehicle–into barrier impacts, etc.) is to be conducted (Yan et al., 2011). The most useful databases for implementing this process are those that are ‘in-depth’—substantial amounts of data are collected and codified for each accident. These data can be schematically divided in categories based on their type and acquisition method, as follows.

- Objective data: These are incontrovertible data, e.g., type of vehicles

* Corresponding author.
E-mail address: michelangelo.gulino@unifi.it (M.-S. Gulino).

https://doi.org/10.1016/j.aap.2018.12.004
Received 11 April 2018; Received in revised form 5 December 2018; Accepted 5 December 2018

involved, occupants (gender, age, number), and airbag deployment.

- Measured data: Some examples of these data are the length of skid marks, and rest positions of vehicles. Uncertainties are typically associated with these data categories, which depend on measuring instruments and method (Brach, 1994).
- Calculated (or estimated) data: These data are calculated by means of the reconstruction of accident dynamics, such as impact velocities, velocity change (ΔV) experienced by vehicles, kinetic energy loss in pre and post-crash phases, and deformation energy of vehicles. Calculated data are not only affected by the uncertainty propagation from measured inputs (Vangi, 2005), but also by the use of simplified physical models to describe accident dynamics and by approximations because of the estimation of parameters (such as the coefficient of friction). For example, if the impact phase between vehicles is reconstructed through impulsive models (e.g., momentum conservation), some coefficients and parameters must be assumed (coefficient of restitution, plane of impact, pre and post-impact directions, etc. (Brach and Brach, 2005)). Literature shows that non-negligible errors (e.g., a difference of as high as 20% between calculated ΔV and that collected by event data recorders) are associated with kinematic parameters derived by special-purpose software (Niehoff and Gabler, 2006; Gabler et al., 2004). On the other hand, a more recent study (Pride et al., 2013) demonstrates that uncertainties related to the estimation of deformation energies (starting from vehicle pictures) propagate on ΔV output values and produce errors that reach as high as 7.7%. Injury-related data can be also grouped in this data category and codified in scales (such as the abbreviated injury scale (AIS) (Gennarelli and Wodzin, 2008), the maximum AIS (MAIS) or the injury-severity score (Baker et al., 1974)); these are susceptible to estimation and codification errors.

Uncertainties and their propagation in calculation, jointly with the estimate and approximation performed, can lead to a non-realistic accident reconstruction. In this case, several kinematic data that are collected in an accident database can be misrepresented.

To guarantee the reliability of data contained in a database, the implementation of some checks should be required during the collection phase, or even better, during the accident reconstruction phase. In this case, calculated values can be verified by comparing them with accident evidences and information. Calculated kinematic data will also have to satisfy the energy conservation principle. Hence, verifying the equality between kinetic energy loss and deformation energy is an effective check. The latter is evaluated by an analysis of vehicle deformations (Vangi, 2009), which constitute an objective datum of an accident. Therefore, during the accident reconstruction phase, it is possible to verify results based on objective inputs, which guarantee that data are correct to a proper degree (i.e., the accident occurred according to reconstructed modes and kinematic values).

As for previously stored data in the database, it is not always possible to perform a complete verification unless the event is reconstructed again. However, this is often impossible because of the lack of necessary information in the database; this explains why the absolute correctness of reconstruction cannot be assessed. Nonetheless, it is possible to define some checks on the coherence of previously collected data. Kinematic data must, in fact, satisfy physical laws depending on the accident type. Based on this, correlations between these data could be defined. Accordingly, the verification of these correlations, although insufficient to guarantee the correctness of results, can ensure that the physical model and assumed parameters have been appropriately applied to the case. These coherence checks reduce analysis errors, whether statistical or not, depending, for example, on the correlations between impact severity and IR (Kononen et al., 2011).

Based on the type of data considered, different types of checks can be performed. However, not all of them are applicable to various databases; for example, a check cannot be implemented because of the lack of some necessary data. Furthermore, not all checks can be conducted without initially obtaining the required data through a case-by-case investigation. In fact, in several cases, the checks can be performed only by extracting information from sketches, such as pre and post-impact directions, and exact geometrical impact configuration between vehicles. For instance, checks pertaining to correlations between the variation of translational and angular velocities can be performed only by knowing the centre of impact (Brach and Brach, 2005), i.e., the point on the vehicle where the resultant of contact forces ideally acted during the impact. This information can be deduced and estimated only by an analysis of the deformed shapes of vehicles involved. The same applies in the estimation of the resultant direction.

The aim of the present work is to propose a number of checks for verifying the internal coherence of kinematic parameters; these checks can be automatically performed through simple routines, which can be implemented in any electronic spreadsheet. The proposed checks are applicable to most of accident databases if proper modifications and adaptation based on data are actually available. Furthermore, because of its generality, the approach can be used in data collection and codification processes, as a tool to verify the correctness of a reconstructed accident.

In this study, analyses are conducted on two accident databases to demonstrate the results of applying the proposed checks. Moreover, these highlight how data errors affect the identification of correlations between IR and impact severity (typically described by ΔV) or the estimate of IR starting from a certain value of impact severity.

2. Method

The procedure is based on three different checks: check on momentum conservation, check on compatibility between pre and post-impact velocities, and check on energy conservation. The physical laws defining these three constitute the classical impulse–momentum theory (Brach and Brach, 2005).

2.1. Check 1—momentum conservation

Applied to the impact between two vehicles, the momentum conservation is expressed by the vector relationship given by the following:

$$m_A \boldsymbol{\Delta V_A} = -\ m_B \boldsymbol{\Delta V_B}, \tag{1}$$

where m_A and m_B, and $\mathbf{\Delta V_A}$ and $\mathbf{\Delta V_B}$ are the masses and velocity change vectors of vehicles A and B, respectively. Because the two $\mathbf{\Delta V}$ vectors are equal and opposite, i.e., they lie on the same line, only the modulus can be considered. It can be written as follows:

$$m_A/m_B = \Delta V_B/\Delta V_A. \tag{2}$$

Eq. (2) applies to all types of crashes regardless of velocities and impact configurations involved (full impacts, sliding impacts, etc.) (Brach and Brach, 2005; Tolouei et al., 2013), only if the forces exchanged between vehicles and road surface are negligible (isolated system). This assumption is generally acceptable when impact forces are prevalent (Mastandrea and Vangi, 2005), i.e., when the crash occurs at a sufficiently high closing speed V_r ($V_r > 10$ km/h).

The first check on data is based on the application of Eq. (2), with a certain tolerance interval. To determine the width of this interval, it must be considered that the actual masses of vehicles may be different from those inputted into the database. In fact, masses recorded in the database are generally kerb weights (m_{kerb}) and not the real masses at the instance of the accident; hence, masses of passengers and loads must also be accounted for.

The kerb weight, as defined by US regulations (CFR, 2018), is equal to the 'weight in operational status… and weight of fuel at nominal tank capacity'. The nominal tank capacity has an average value of 40 L (Pride et al., 2013), but its real value varies as a function of mass in the operational status; a linear relationship between m_{kerb} and m_{fuel} (mass of the fuel; all masses in kg) is assumed:

$$m_{fuel} = 0.02 \cdot m_{kerb} + 20. \tag{3}$$

Eq. (3) considers a capacity of 40 L for a 1000-kg vehicle and 80 L for a 3000-kg vehicle.

The number of occupants, N, involved in the accident is a known variable inside databases, whereas occupant mass is often unknown. Hypothesising a minimum mass, m_{occ_min}, of 40 kg for each occupant (this accounts for the possible presence of children) and a maximum mass, m_{occ_max}, of 100 kg, it is possible to obtain the minimum and maximum total masses of occupants (M_{min} and M_{max}, respectively) in a single vehicle:

$$M_{\min} = m_{occ\_\min} \cdot N; \quad M_{\max} = m_{occ\_\max} \cdot N. \tag{4}$$

The minimum mass that can be associated with vehicle i (i = A, B) is equal to the kerb weight with an empty tank plus the minimum total mass of occupants:

$$m_{i\_\min} = m_{kerb} - m_{fuel} + M_{\min}. \tag{5}$$

The maximum mass of the vehicle is equal to the sum of kerb weight (with the tank at a nominal capacity) and maximum mass of occupants (the mass of the possible cargo is generally unknown):

$$m_{i\_\max} = m_{kerb} + M_{\max}. \tag{6}$$

The effective vehicle mass can range between these two values. Thus, for the check on data congruence, Eq. (2) can be expressed as follows:

$$\frac{m_{B\_\min}}{m_{A\_\max}} \le \frac{\Delta V_{A(database)}}{\Delta V_{B(database)}} \le \frac{m_{B\_\max}}{m_{A\_\min}} \tag{7}$$

where $\Delta V_{i(database)}$ is the value of ΔV collected in the database for vehicle i. The check based on Eq. (7) defines data congruence by the momentum conservation principle. If Eq. (7) is not satisfied, then data are not congruent. However, the incongruity does not depend on the masses; it may be a consequence of using an incorrect accident reconstruction model or procedure.

2.2. Check 2—compatibility between pre and post-impact velocities

A methodology notably used in the accident reconstruction field is based on the kinetic energy that is lost in the vehicle deformation, E_d, from which the values of ΔV and V_r are derived. This approach is the basis of several software, such as WinSmash and EDCrash (Sharma et al., 2007; Day and Hargens, 1987). The calculation relates only to the modulus; thus, other information regarding the vectors are not available. For a generic vehicle, velocity change, initial velocity, and final velocity vectors (**ΔV**, **V**, and $\bar{V}$, respectively) form a closed triangle, because the vector sum, $V_i + \Delta V_i = \bar{V}_i$, must apply. Thus, a second check for the data coherence pertains to the fact that the three vectors compose a triangle. However, in most accident databases, directions and/or moduli of all three velocity vectors are not explicitly reported. For example, in the Initiative for the Global Harmonization of Accident Data (IGLAD, 2018) database, the moduli of vectors **ΔV** and **V**, as well as the Δα angle between **V** and $\bar{V}$, are recorded; however, in the National Automotive Sampling System Crashworthiness Data System (NASS/CDS, 2018) database, such data are not completely included.

In this case, if the threshold value of ΔV allows a closed triangle to exist, then two geometrical conditions can be derived for vehicle i; these conditions link $\Delta V_{i(database)}$ and the pre-impact speed value reported in database $V_{i(database)}$:

$$\text{If } |\Delta\alpha_i| < 90^\circ \Rightarrow d_i = V_{i(database)} \cdot \sin(|\Delta\alpha_i|) \le \Delta V_{i\,(database)}, \tag{8}$$

$$\text{If } 90^\circ \le |\Delta\alpha_i| \le 270^\circ \Rightarrow d_i = V_{i(database)} \le \Delta V_{i\,(database)}. \tag{9}$$

An example of the inconsistency in the velocity triangle is shown in Fig. 1, in which all parameters of interest are depicted. In the application of Eqs. (8)–(9), an uncertainty in $\Delta V_{i(database)}$ associated with the uncertainty in the masses must be considered, as shown above (where check 1 was defined). For example, for $\Delta V_{i(database)}$ of vehicle A (and correspondingly for vehicle B) in Eqs. (8)–(9), the following minimum value can be considered:

$$\Delta V_{A\_\min} = \Delta V_{B(database)} \cdot m_{B\_\min} / m_{A\_\max} \tag{10}$$

Furthermore, a few uncertainties in the directions of velocity vectors exist and produce an uncertainty on Δα; this uncertainty on Δα can be assumed as ± 3°. This value is typically used in commercial software packages (e.g., Pro-Impact, whose calculation logic is described in (Vangi et al., 2018a)) as the uncertainty angle among the directions of post-impact velocities (the pre-impact velocity is considered known with a particular degree of certainty) for performing simulations with the Monte Carlo method. Thus, an angle equal to $t_\alpha = \Delta\alpha - 3^\circ$ can be used in Eqs. (8) and (9).

If Eqs. (8) and (9) are not satisfied, then data are not congruent. This may be because of the incorrect application of the physical model employed to reconstruct the accident. The assumptions made or the data employed (for example, in pre or post-impact directions) may have also been incorrect.

2.3. Check 3—energy conservation

Consider a case in which both checks 1 and 2 are satisfied, i.e., the ratio between ΔV of vehicles A and B is congruent with their mass ratio, and the vectors form a closed triangle (Fig. 2(a)). For the same accident, Fig. 2(b) presents identical kinematic parameters scaled at a certain factor. Even in this case, checks 1 and 2 are satisfied. Hence, data exhibit coherence and could similarly be as substantial as those shown in Fig. 2(a). However, it is evident that in terms of the absolute impact velocity of the two vehicles, the results are different between the two cases; analogously, in terms of the impact severity and IR of vehicle occupants the consequences will be also different (Weaver et al., 2015). Thus, it is necessary to impose an additional check on the coherence of kinematic data and those related to the accident (e.g., the consequences of impact on vehicles).

Consider the example shown in Fig. 3, in which two identical vehicles (equal masses) collide at equal velocities. Considering the crash as perfectly plastic (an assumption completely acceptable for impacts where V_r is greater than 40 km/h (Antonetti, 1998)), the final velocities are null. The application of Eq. (1) yields the relationship $V_A = V_B = V$, which can be satisfied by infinite combinations of kinematic parameters determined as congruent through checks 1 and 2. Different velocities imply different permanent deformations. A check on the deformation energy, E_d, establishes whether the accident has been reconstructed in a way that is also congruent with accident results (Brach and Brach, 2005). For the energy conservation law (neglecting the eventual energy associated to vehicle rotation), the following relationship must hold:

$$\Delta E_c = E_c - \overline{E_c} = \frac{1}{2}(m_A V_A^2 + m_B V_B^2) - \frac{1}{2}(m_A \bar{V}_A^2 + m_B \bar{V}_B^2) = E_d, \tag{11}$$

where E_c and $\bar{E}_c$ represent the initial and final kinetic energies of the system, respectively. If the characteristics of occupants are unknown, an average mass of 70 kg for each occupant can be assumed in order to compute the actual mass of each vehicle.

To apply Eq. (11), it is necessary to know the pre and post-impact velocities of the two vehicles, as well as the deformation energy. Assuming that the kinematic parameters used in check 2 are available, two possible configurations for closed triangles 1 and 2 are shown in Fig. 4. In Fig. 4a, the circumference (of a circle with radius $\Delta V_{i(database)}$ and centre at the tip of the vector $\mathbf{V_{i(database)}}$), intersects the $\bar{V}$ direction at point P_1; in Fig. 4b, the intersection is at P_2. The angle $\Delta\gamma_{ij}$ between $\mathbf{\Delta V_{i(database)}}$ and $\bar{V}_{ij}$ for the jth configuration (j = 1,2) is obtainable by the law of sines based on the values of $\Delta V_{i(database)}$, $\Delta\alpha_i$, and $V_{i(database)}$; noting that $\Delta\beta_{ij} = \pi - \Delta\gamma_{ij} - \Delta\alpha_i$, $\bar{V}_{ij}$ can be analogously derived:

$$\bar{V}_{ij} = \frac{\Delta V_{i(database)}}{\sin\Delta\alpha_i}\sin(\Delta\beta_{ij}). \tag{12}$$

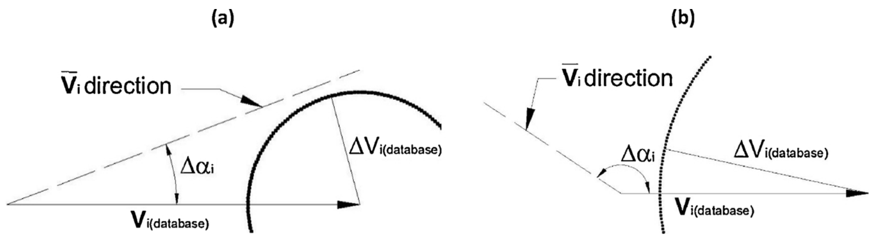

Fig. 1. Conditions of velocity triangle inconsistency: (a) $|\Delta\alpha| < 90°$; (b) $90° \le |\Delta\alpha| \le 270°$.

The two post-impact velocities involve two different values in the kinetic energy of each vehicle. The sum of kinetic energies in the various combinations calculated by Eq. (11) is expressed as ΔE_{c_kl}, with k = 1,2 and l = 1,2; the foregoing indicates the configuration from which $\bar{V}$ is obtained for vehicles A and B. Negative values of $\bar{V}_{ij}$ are not acceptable because they imply a closed triangle only in the case where $\Delta\alpha_i$ has an opposite sign (or is complementary) to that reported. Furthermore, negative values of the kinetic variations of ΔE_{c_kl} must be discarded because of the absence of any physical meaning. For all cases in which these requirements are satisfied, the check on kinetic energy can be expressed as follows:

$$E_{d\_\min} \le \Delta E_{c\_kl} \le E_{d\_\max} \tag{13}$$

where E_{d_min} and E_{d_max} are the threshold deformation energies.

As previously stated, ΔV of the two vehicles must act along the same direction. Therefore, only two of the four possible velocity triangles are valid, i.e., those characterised by the same direction of **ΔV**; this facilitates the verification of energy conservation. However, to do this, the directions of pre-impact velocities must be known in a fixed reference system. These directions can be deduced from eventual sketches collected in the database; however, the calculation cannot be automated. The proposed procedure, even if it considers four possible solutions for ΔE_c, is in fact applicable starting only from numerical data typically present in databases; therefore, it can be automated.

As for the deformation energy in general, Eq. (14) has a parameter that is often reported in databases. This parameter links E_d and the energy equivalent speed (EES) of the two vehicles:

$$E_d = \frac{1}{2} m_A EES_A^2 + \frac{1}{2} m_B EES_B^2 \tag{14}$$

If it is to be assumed that a road accident expert commits an error, t_e = 3 km/h (Vangi, 2009) in estimating the EES by using approximate methods, then the following relationship can be deduced:

$$E_{d\_\min} = \tfrac{1}{2} m_A (EES_A - t_e)^2 + \tfrac{1}{2} m_B (EES_B - t_e)^2,$$
$$E_{d\_\max} = \tfrac{1}{2} m_A (EES_A + t_e)^2 + \tfrac{1}{2} m_B (EES_B + t_e)^2. \tag{15}$$

If none of the four ΔE_{c_kl} satisfies the check in Eq. (13), a coherence error between kinematic data and impact consequences exists in the case.

2.4. Example of procedure application to a single case

To clarify the concepts introduced above, the procedure is applied to a single real accident case found in the IGLAD database; Table 1 summarises the data required for process implementation, all of which are stored in IGLAD database.

Considering the relationships expressed in Eqs. (3)–(6), derived data are summarised in Table 2. Based on Eq. (7), the inequality 0.88 ≤ 1.04 ≤ 1.16 holds, and check 1 is satisfied.

The data listed in Table 3 are calculated using Eqs. (8)–(10). The inequalities of Eqs. (8) and (9) hold, and consequently, check 2 is satisfied.

Table 4 summarises the parameters which must be derived to apply check 3 (defined in Section 2.3). Multiple values of the listed kinematic parameters derive from various configurations of velocity triangles, depicted in Fig. 4. As mentioned above, negative values of both $\bar{V}$ and ΔE_c can be omitted from the analysis; among the different possibilities, a ΔE_c value that satisfies Eq. (13) exists. Overall, the internal data congruence is verified for the analysed case.

For more application examples (including the use of incoherent data), the reader is referred to a previous work of the authors (Vangi et al., 2018b).

3. Procedure application to complete databases

To demonstrate the applicability of the method in the complete evaluation of accident databases, IGLAD and NASS/CDS, are analysed. Only collisions between two vehicles, which fall under the following categories are considered: passenger cars, multi-purpose vehicles, vans, and light trucks. The errors highlighted by the application of each check on analysed databases can be defined for a single vehicle, *i*, as follows:

$$\underline{\text{Check1}} \quad \varepsilon_m = \min(|\Delta V_{i(database)} - \Delta V_{i\_\max}|; |\Delta V_{i(database)} - \Delta V_{i\_\min}|), \tag{16}$$

$$\underline{\text{Check 2}} \quad \varepsilon_{\Delta\alpha} = d_i - \Delta V_{i(database)}, \tag{17}$$

$$\underline{\text{Check3}} \quad \varepsilon_{EES} = \min(|\Delta E_c - E_{d\_\max}|; |\Delta E_c - E_{d\_\min}|), \tag{18}$$

where ΔE_c is the kinetic energy solution that approximates the condition of congruence among other possible solutions (Eq. (13)). Assuming that $\Delta V_{(database)}$ is correct for the other vehicle involved in the crash, the maximum and minimum values of velocity changes are equal to the following:

$$\Delta V_{A\_\min} = \Delta V_{B(database)} \frac{m_{B\_\min}}{m_{A\_\max}}, \quad \Delta V_{A\_\max} = \Delta V_{B(database)} \frac{m_{B\_\max}}{m_{A\_\min}},$$
$$\Delta V_{B\_\min} = \Delta V_{A(database)} \frac{m_{A\_\min}}{m_{B\_\max}}, \quad \Delta V_{B\_\max} = \Delta V_{A(database)} \frac{m_{A\_\max}}{m_{B\_\min}}. \tag{19}$$

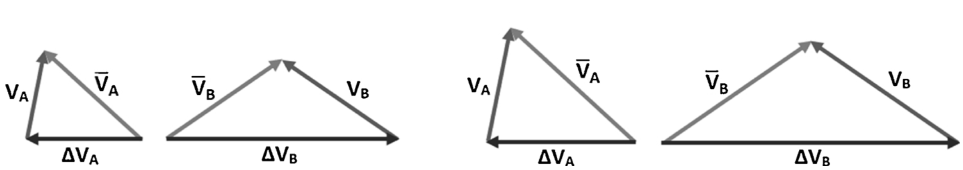

Fig. 2. Velocity triangles satisfying checks 1 and 2, but with a different scale factor.

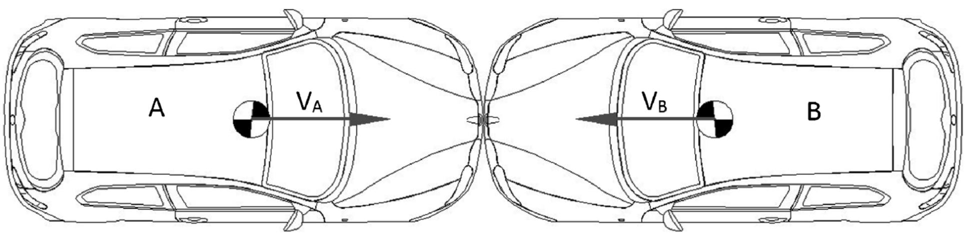

Fig. 3. Collision between two identical vehicles in which only the extent of deformation can provide adequate indications of reconstruction correctness.

3.1. The IGLAD database

The IGLAD database has been created by the collaboration between vehicle manufacturers and research institutes. The purpose is to make a wide array of accident cases available, on which investigations to improve road safety could be based. The database is updated yearly with hundreds of cases from various parts of the world regarding collisions between vehicles, vehicles and infrastructures, vehicles and pedestrians, etc. It is divided into several phases based on the year of data collection. In the present work, all cases currently available are analysed (that is, all accidents that occurred between 2013 and 2016).

The IGLAD database (including the German In-Depth Accident Study (GIDAS) database (BASt, 2018)) belongs to accident databases in which impulsive models are typically employed to reconstruct accidents. For this reason, the information regarding pre and post-impact velocities are known, and all checks presented above can be applied to the database (Vangi et al., 2018b). The parameters used for both vehicles involved in the collision are thus m_{kerb}, $\Delta V_{(database)}$, N, $V_{(database)}$, $\Delta\alpha$, and EES. The results of the application of the checks to a total of 1170 crash cases involving two vehicles are reported in Fig. 5. It distinguishes between cases to which the single check cannot be applied, and cases that do and do not satisfy the criteria. Overall, the number of cases in which all three checks are applicable is 272, of which 91 satisfy at least one of the criteria and 181 satisfy all checks.

The error distribution of cases not satisfying the proposed checks (Eqs. (16)–(18)) is represented in Fig. 6. The entity of errors in check 2 appears to be spread over a wider range of values with respect to distributions corresponding to checks 1 and 3. It is worth noting that the number of cases affected by the error in Fig. 6c is not equal to the number of cases that do not satisfy check 3 in Fig. 5. This discrepancy is explained by the presence of cases compatible with negative values of ΔE_c only (values that lack physical meaning).

Table 1
Data of a real accident case stored in the IGLAD database.

	M_{kerb} (kg)	N	$\Delta V_{(database)}$ (km/h)	$V_{(database)}$ (km/h)	$\Delta\alpha$ (°)	EES (km/h)
Vehicle A	865	1	70	94	29	54
Vehicle B	810	2	67	45	144	78

Table 2
Data calculated by Eqs. (3)–(6) for the considered real accident case.

	M_{min} (kg)	M_{max} (kg)	m_{fuel} (kg)	m_{min} (kg)	m_{max} (kg)
Vehicle A	40	100	37	868	965
Vehicle B	80	200	36	854	1010

Table 3
Data calculated by Eqs. (8)–(10) for the considered real accident case.

	t_α (°)	d(km/h)	ΔV_{min} (km/h)
Vehicle A	26	41	59
Vehicle B	141	45	60

Table 4
Data calculated by relationships introduced in Section 2.3 for the considered real accident case.

	$\Delta\gamma$ (°)	$\Delta\beta$ (°)	$\bar{V}$ (km/h)	ΔE_c (kJ)	E_{d_min} (kJ)	E_{d_max} (kJ)
Vehicle A	41	110	136	−300	300	376
	139	12	30	−1253		
Vehicle B	23	13	26	335		
	157	−121	−98	15		

3.2. The NASS/CDS database

Over 1000 accidents between two vehicles in different states of the USA are collected yearly and inputted into the NASS/CDS database. Cataloguing follows accident reconstruction, which is generally conducted using software packages based on energy algorithms, such as CRASH3 or WinSMASH (Sharma et al., 2007).

Because there are variables stored in the NASS/CDS database that differ with those of IGLAD, it is not possible to apply all checks to the former database. Check 1 can be applied because the mass (m_{kerb}), number of occupants (N), and $\Delta V_{(database)}$ of the vehicles are reported in practically all cases and can therefore be employed. On the other hand, check 2 is not applicable (particularly with the use of automation) because of the lack of information regarding pre and post-impact velocity directions (note that although there is a database field dedicated to the pre-impact velocity modulus, it is rarely filled). As for check 3, it is impossible to apply the criterion because the EES is not known, although the deformation energies of vehicles are. Analogous to the analysis of the IGLAD database (Fig. 5), Fig. 7 shows results of the application of check 1 to NASS/CDS cases that involve collisions of two

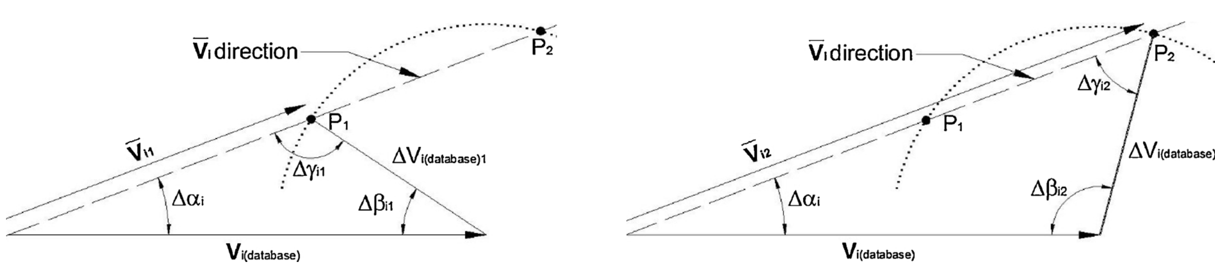

Fig. 4. Two possible configurations in which the velocity triangle can be closed with the same known kinematic parameters: (a) intersection at P_1 and (b) intersection at P_2.

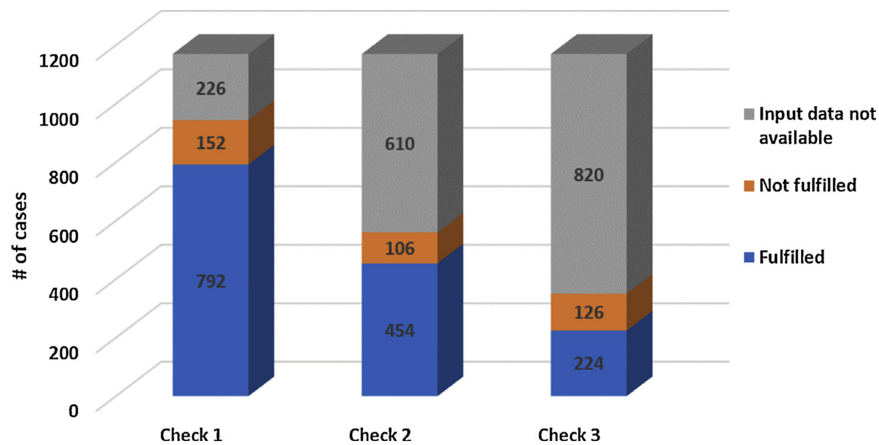

Fig. 5. Results of the application of three checks to data stored in IGLAD database.

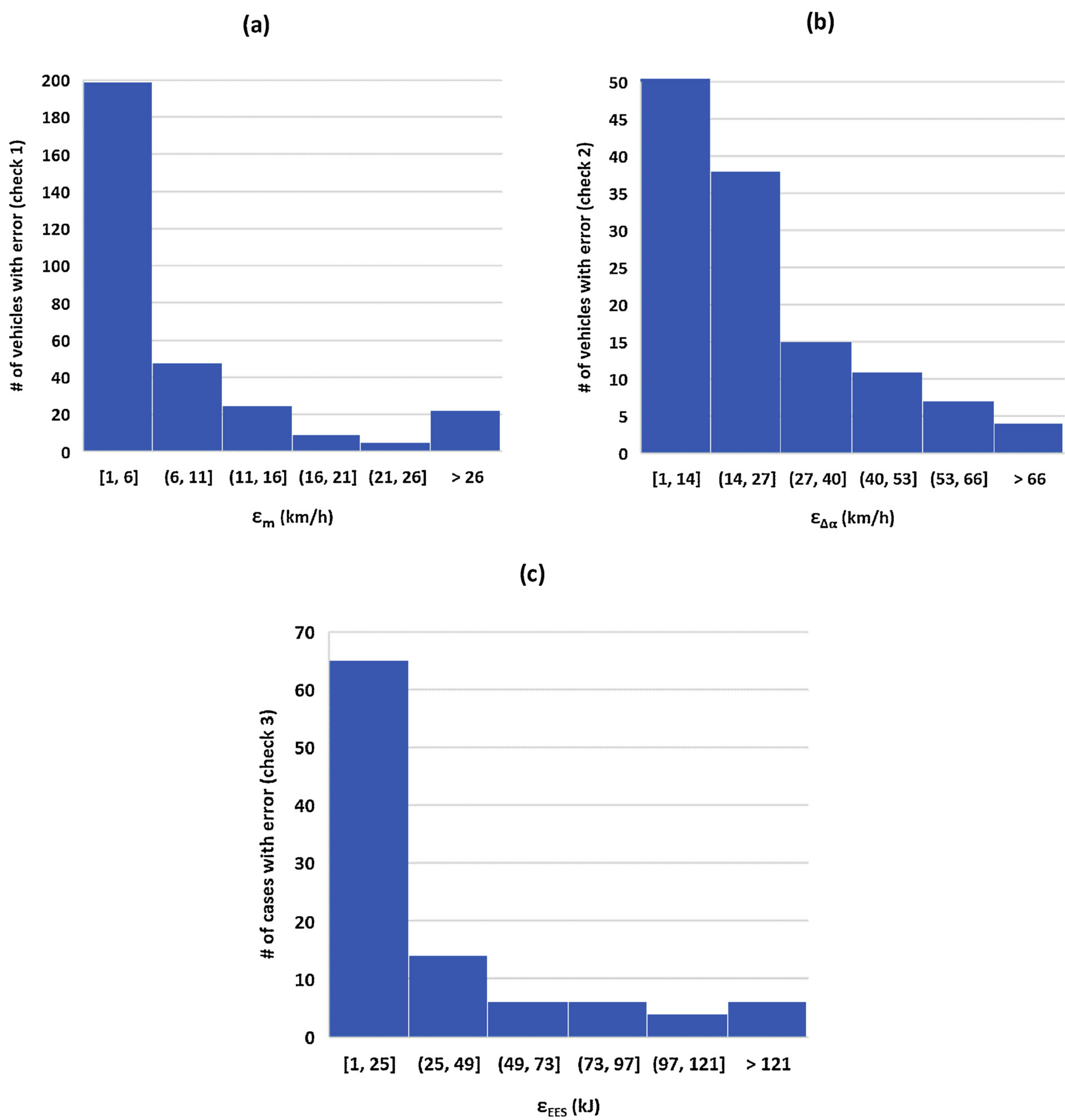

Fig. 6. Distribution of errors for a single vehicle highlighted by the application of (a) check 1, (b) check 2, and (c) check 3 to cases collected in the IGLAD database.

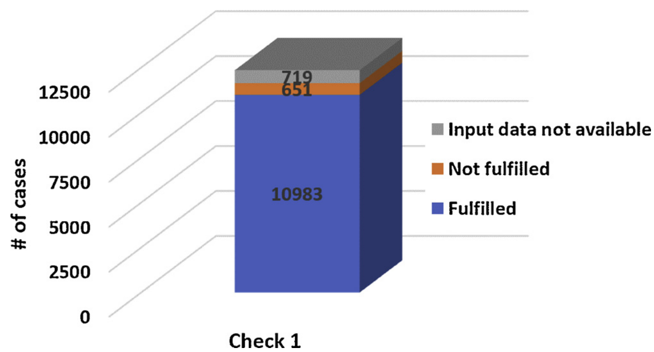

Fig. 7. Results of check 1 applied to data collected in the NASS/CDS database.

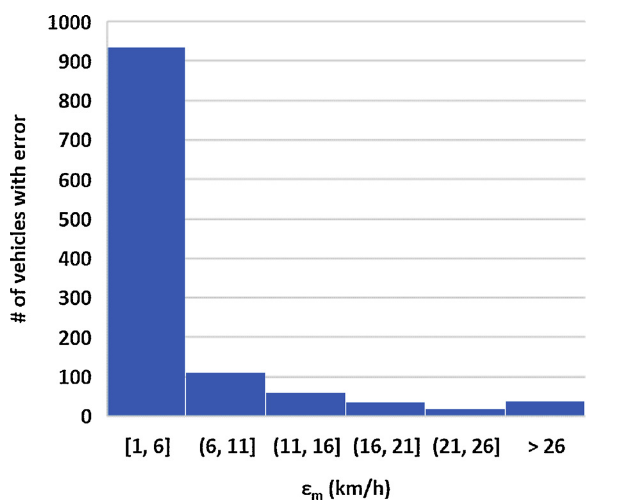

Fig. 8. Distribution of errors in a single vehicle, highlighted by the application of check 1 to cases in the NASS/CDS database.

vehicles in a total of 12 353 accidents (2004–2015).

The distribution of errors for vehicles that do not satisfy check 1 (Eq. (16)) is reported in Fig. 8.

4. Discussion

Although only car-to-car collisions have been considered in this study, the proposed checks can be applied to all vehicle-to-vehicle crashes, including motorcycles. With the methodology described above, any impact configuration can be analysed both at high and moderate velocities. The imposed conditions to verify data correctness allow the combination of cases from different databases to increase the number and enhance the quality of cases with substantiated data. This is particularly important because different countries generally have different collection methods and accuracies (Fildes et al., 2013). The tolerance values m_{min}, m_{max}, and t_e that were used in this work are only illustrative and can be modified based on the desired accuracy level for subsequent analyses.

The examples regarding cases in IGLAD and NASS/CDS databases show how coherence errors can be present in a remarkable number of accidents. It is possible to highlight a number of incongruities, which are undoubtedly lower in the NASS/CDS database compared to that of IGLAD. On the other hand, for the latter database, all three checks are applicable; thus, the remaining data can be assessed with high quality. This suggests that further proper investigations would be necessary for data in the NASS/CDS database. Figs. 5 and 6 identify an additional problem, which researchers frequently encounter during in-depth accident database analysis: the presence of several cases with incomplete data. This includes both the possibility of a parameter not codified and

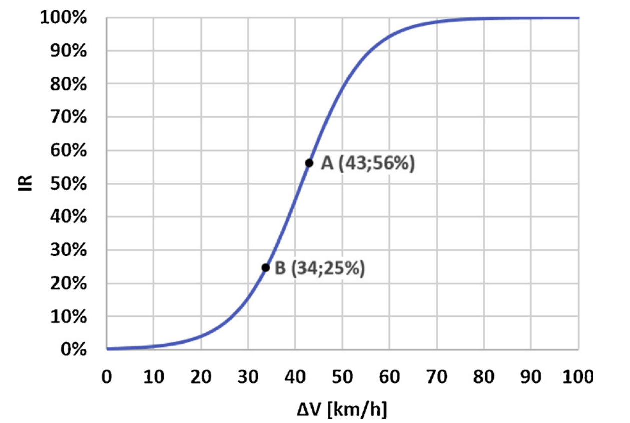

Fig. 9. Example of error propagation from ΔV to IR (IR curve referred to frontal impacts, MAIS 3+ (Jurewicz et al., 2016)).

a datum not collected in certain accidents. Under both circumstances, the contribution that the database offers in a specific analysis, where kinematic data are needed, can be limited. In this regard, technicians and groups involved in the data collection process can refer to the checks described in this work as an additional verification for the quality of uncertain data, which are often excluded from the database.

Assume that a search is made for information regarding the performance of a vehicle in terms of crashworthiness in a frontal impact. For a crash, this is typically obtained as the combination of IR and associated injury severity (Newstead et al., 2016). Without the loss of generality, for a fixed injury severity—which can be expressed by a MAIS index—it is possible to refer only to IR curves obtained through logistic regression. If the occupant suffers no severe injury in a crash characterised by an IR higher than 50%, it can be stated that the vehicle possesses a high crashworthiness in the first instance; literature-based IR curves are found in (Jurewicz et al., 2016), depicted in Fig. 9, and referred to MAIS higher than 2 (MAIS 3+). For instance, if a ΔV (impact severity index) from the database is equal to 43 km/h (point A), with an associated error of -9 km/h (a value encountered with a non-negligible frequency in the analysed database, point B), the error propagates on the IR value (shifting it from 56 to 25%) and makes the analysis inconsistent.

On the other hand, if IR curves are obtained by logistic regression models, the use of correct data certainly benefits the analysis. To enhance the overall quality, the errors on ΔV can be compensated by translating the IR curves derived using reconstructed ΔV to match the real associated ΔV values (Funk et al., 2008). In the case of incongruity, the kinematic data that are affected by errors can be excluded. For example, considering only occupants with a maximum MAIS inside a vehicle involved in a crash, Fig. 10 shows two IR curves (MAIS 3+) representing the IGLAD data, which were obtained by employing 1) all cases and 2) only cases with congruent values of ΔV (verified by check 1). The difference between the two curves reaches values that are higher than 7% (corresponding to ΔV = 65 km/h (IR $\cong$ 60%)); the extent of this difference is higher than that of the confidence interval typically adopted in a statistical analysis, which can be assumed equal to 10% of the IR value (Pride et al., 2013). Thus, the use of as-is data only introduces erroneous information. Hence, the application of the described checks is always advisable in database analysis. The benefits derived from the proposed procedure becomes more evident as the number of cases that are subject to disposal decreases (Ye and Lord, 2014). Whereas confounding factors, such as occupants' age, position, and seatbelt use, are not considered in determining IR curves as in (Funk et al., 2008), results depict how low-quality kinematic data can affect the analysis.

The checks can also be employed by reconstruction experts, who participate in the data collection process to enhance the quality of

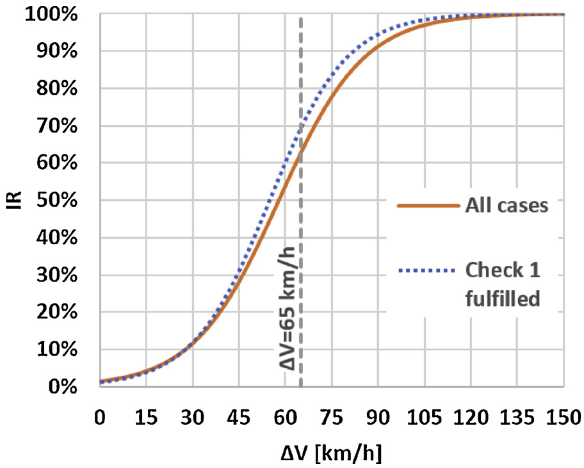

Fig. 10. Two IR curves obtained from IGLAD database analysis (MAIS 3+), considering all available data or only those that satisfy check 1.

accident reconstruction prior to data insertion. For this purpose, ΔV can be considered as the main kinematic datum to assess reconstruction plausibility, because all proposed checks are only applicable if its value is known. The values of the reconstructed ΔV approximating the extremities of the congruence interval expressed by Eq. (7) can imply an inadequacy in the reconstruction model (e.g., the stiffness of considered vehicles). In this case, shifting from categorical stiffness coefficients (based on the corresponding vehicle body type) to that of a specific stiffness can result in a more accurate estimate of ΔV by approximately 5% (Niehoff and Gabler, 2006). The quality can further be increased by considering the restitution phase of the crash; this results in extremely small differences between real ΔV values and those calculated (approximately 1%) (Niehoff and Gabler, 2006). The reconstruction software setup can thus be adjusted to make all parameters agree to the fullest with the conditions expressed by Eqs. (7)–(9) and (13).

While special attention has been paid to real accident cases in which only two vehicles collided, the proposed procedure identically applies if more than two vehicles are involved in the accident. In particular, the kinematic parameters of interest must be reported singularly for each event composing the accident (i.e., each crash between two vehicles): for what regards the analysed databases, this condition applies.

5. Limitations

Whereas the procedure represents an efficient tool to assess internal coherence among data for cases codified in in-depth accident databases, the absolute correctness of reconstructed data cannot be verified. In fact, although internal coherence is a necessary condition to assess the appropriateness of the reconstruction, it is not sufficient; if the assumptions for the reconstruction are not consistent with real accident kinematics, then the reconstructed data can still be internally coherent. To highlight eventual conflicts between real accident kinematics and associated data recorded in the database, an additional reconstruction of the event would be required. Moreover, because this operation cannot be automated, the benefits and potentials of collecting synthetic data in in-depth accident databases would be diminished.

Because the technique cannot be applied to the complete NASS/CDS database, the procedure cannot be entirely implemented as-is for the analysis of any in-depth accident database. This is because of the different types of data stored in the database and reconstruction methodologies used for their calculation. Hence, different formulations of the same physical laws (or others) should be implemented to comprehend the parameters that differ from those reported in this work.

6. Conclusions

The procedure described in the present work represents an efficient tool for checking the internal coherence among kinematic data collected in accident databases. The proposed data checks are based on 1) momentum conservation, 2) compatibility of velocity triangles, and 3) energy conservation.

Among all possible checks for the internal coherence of data, those proposed in this paper are applicable without the necessity to analyse each case individually. This can be achieved through the recalculation of parameters or by a sketch analysis; on the other hand, the checks can easily be automated in the database. In principle, although no conclusion can be derived regarding the correspondence between true kinematic data resulting from the crash and those obtained by reconstruction, the internal congruence check among the parameters guarantees that the physical models have been properly applied. Consequently, these checks are useful in the accident reconstruction phase, data collection, data codification into the database, and accident analysis phase.

Considering the uncertainties present in the measurable and calculable parameters obtained from accident reconstruction, the proposed checks define threshold values as a function of the verified parameters. The threshold values assumed in the present work are based on physical considerations that can be modified according to the accuracy required on a case-by-case basis.

The application of checks to IGLAD and NASS/CDS demonstrates that a relevant number of incongruent kinematic data are present in these two different databases. These incongruities inevitably affect the statistical analysis based on kinematic parameters (implementable by database analysis) as the correlation between impact severity and IR. Moreover, the individuation of an IR associated with a certain accident can be inconsistent if the real case is characterised by incongruities in the collected parameters.

Tolerances proposed in the present work for each check are illustrative and can be modified to account for uncertainty (if known) on data collected in the analysed accident database. In general terms, the choice of more strict tolerances is associated to a decrease in the sample size of the dataset, making the results of the associated analysis less reliable: for this reason, appropriate values for tolerances must be selected to cater both needs of data accuracy and adequate sample size.

The application of the procedure to check data congruence is essential to perform the analyses of real accidents based on kinematic data.

References

Antonetti, V.W., 1998. Estimating the coefficient of restitution of vehicle-to-vehicle bumper impacts. SAE Technical Papers, n. 980552. pp. 1–9.
Baker, S.P., O'Neill, B., Haddon, W., et al., 1974. The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care. J. Trauma 14 (3), 187–196.
BASt. [Online]. 2018 Available: https://www.bast.de/BASt_2017/EN/Automotive_Engineering/Subjetcs/gidas.html?cms_gtp=1832216_Dokumente%253D3 (Accessed 29 September 2018).
Brach, R.M., 1994. Uncertainty in accident reconstruction calculations. SAE Technical Papers Series, no. 940722. pp. 147–153.
Brach, R.M., Brach, M.R., 2005. Vehicle Accident Analysis and Reconstruction Methods. SAE International, Warrendale, PA, pp. 1–12.
Burnett, R., Carter, J., Roberts, V., et al., 2004. The influence of seatback characteristics on cervical injury risk in severe rear impacts. Accid. Anal. Prev. 36, 591–601.
40 CFR, 2018. 40 CFR §86.1803-01. US Government Publishing Office.
Cicchino, J.B., 2017. Effectiveness of forward collision warning and autonomous emergency braking systems in reducing front-to-rear crash rates. Accid. Anal. Prev. 99, 142–152.
Day, T.D., Hargens, R.L., 1987. An overview of the way EDCRASH computes delta-V. SAE Technical Papers, No. 870045. pp. 183–198.
Fildes, B., Keall, M., Thomas, P., et al., 2013. Evaluation of the benefits of vehicle safety technology: the MUNDS study. Accid. Anal. Prev. 55, 274–281.
Flannagan, C.A.C., Bálint, A., Klinich, K.D., et al., 2018. Comparing motor-vehicle crash risk of EU and US vehicles. Accid. Anal. Prev. 117, 392–18397.
Funk, J.R., Cormier, J.M., Gabler, H.C., 2008. Effect of delta-V errors in NASS on frontal

crash risk calculations. In: 52nd AAAM Annual Conference. San Diego, CA.
Gabler, H.C., Hampton, C.E., Hinch, J., 2004. Crash severity: a comparison of event data recorder measurements with accident reconstruction estimates. In: SAE World Congress. Detroit, MI.
Gennarelli, T.A., Wodzin, E., 2008. Abbreviated Injury Scale 2005: Update 2008. Association for the Advancement of Automotive Medicine, Barrington, IL.
IGLAD, 2018. Initiative for the Global Harmonization of Data, "IGLAD". [Online] Available:. (Accessed 29 September 2018). www.iglad.net.
Jurewicz, C., Sobhani, A., Woolley, J., et al., 2016. Exploration of vehicle impact speed—injury severity relationships for application in safer road design. 6th Transport Research Arena.
Kononen, D.W., Flannagan, C.A.C., Wang, S.C., 2011. Identification and validation of a logistic regression model for predicting serious injuries associated with motor vehicle crashes. Accid. Anal. Prev. 43, 112–122.
Kullgren, A., 2008. Dose-response models and EDR data for assessment of injury risk and effectiveness of safety systems. In: Proceedings of the 2008 International IRCOBI Conference. Bern, CH.
Mastandrea, M., Vangi, D., 2005. Influence of braking force in low-speed vehicle collisions. Proc. Inst. Mech. Eng. Part D: J. Automobile Eng. 219, 151–164.
McDonald, C.C., Curry, A.E., Kandadai, V., et al., 2014. Comparison of teen and adult driver crash scenarios in a nationally representative sample of serious crashes. Accid. Anal. Prev. 72, 302–308.
Mendez, Á.G., Izquierdo, F.A., Ramírez, B.A., 2010. Evolution of the crashworthiness and aggressivity of the Spanish car fleet. Accid. Anal. Prev. 42, 1621–1631.
National Highway Traffic Safety, 2018. National Automotive Sampling System. [Online]. Available:. (Accessed 29 September 2018). https://www.nhtsa.gov/research-data/national-automotive-sampling-system-nass.
Newstead, S., Watson, L., Cameron, M., 2016. Vehicle Safety Ratings Estimated From Police Reported Crash Data: 2016 Update. Monash University Accident Research Centre, Melbourne.
Niehoff, P., Gabler, H.C., 2006. The accuracy of WinSmash delta-V estimates: the influence of vehicle type, stiffness, and impact mode. In: 50th Annual Proceedings Association For The Advancement Of Automotive Medicine. Chicago, IL.
Pride, R., Giddings, D., Richens, D., et al., 2013. The sensitivity of the calculation of delta-V to vehicle and impact parameters. Accid. Anal. Prev. 55, 144–153.
Sharma, D., Stern, S., Brophy, J., et al., 2007. An overview of NHTSA's crash reconstruction software WinSMASH. In: 20th International Technical Conference on the Enhanced Safety of Vehicles (ESV). Lyon, FR.
Tolouei, R., Maher, M., Titheridge, H., 2013. Vehicle mass and injury risk in two-car crashes: a novel methodology. Accid. Anal. Prev. 50, 155–166.
Vangi, D., 2005. Collision stage reconstruction in traffic accidents: sensitivity analysis. Proc. Inst. Mech. Eng. Part D: J. Automobile Eng. 219, 285–294.
Vangi, D., 2009. Simplified method for evaluating energy loss in vehicle collisions. Accid. Anal. Prev. 41, 633–641.
Vangi, D., Begani, F., Gulino, M.S., et al., 2018a. A vehicle model for crash stage simulation. IFAC PapersOnLine, vol. 51–2. pp. 837–842.
Vangi, D., Cialdai, C., Gulino, M.S., et al., 2018b. Vehicle accident databases: correctness checks for accident kinematic data. Designs 2 (4), 1–11.
Weaver, A.A., Talton, J.W., Barnard, R.T., et al., 2015. Estimated injury risk for specific injuries and body regions in frontal motor vehicle crashes. Traffic Inj. Prev. 16 (Sup 1), 108–116.
Yan, X., Ma, M., Huang, H., et al., 2011. Motor vehicle–bicycle crashes in Beijing: irregular maneuvers, crash patterns, and injury severity. Accid. Anal. Prev. 43, 1751–1758.
Ye, F., Lord, D., 2014. Comparing three commonly used crash severity models on sample size requirements: multinomial logit, ordered probit and mixed logit models. Anal. Methods Accid. Res. 1, 72–85.