Chapter 5

Defining Traffic Conflict in Non-Lane-Based Traffic Condition

5.1 Preface

In a heterogeneous traffic stream, vehicles do not necessarily travel at the center of the lane. In such a traffic scenario, vehicles interact in longitudinal as well as lateral directions [2, 200]. Therefore, lateral interactions should also be incorporated while defining conflict in these traffic conditions. This chapter fulfills objective 1 of this dissertation and presents a novel methodology for defining conflicts in non-lane-based 2-D interactions. A bivariate peak-over threshold model is utilized to define conflict and crash risk (Section 3.2.2). Further, the estimated crash risk is compared with the observed crash data.

The chapter is organized as follows: Section 5.1 illustrates the review of relevant literature on defining conflict in non-lane-based traffic. Section 5.2 presents the extraction of conflict indicators from the trajectory data. Section 5.3 describes the application of the bivariate EVT model for defining conflict in non-lane-based traffic. Section 5.4 presents the model estimates and a discussion of the results. Finally, Section 5.5 presents the summary of significant findings.

5.2 Suitability of Conflict Indicators in Non-Lane-Based Traffic

In developing countries, traffic stream is composed of multiple vehicle types sharing same space. Vehicles in such traffic conditions try to travel ahead by maintaining lateral gaps without following lane discipline, generating non-lane-based traffic. Vehicles in these traffic streams interact in longitudinal as well as lateral directions, leading to a two-dimensional interaction [2, 33, 200]. To quantify traffic conflict and, hence safety in non-lane-based traffic, Nguyen et al. [135] proposed the idea of safety space. Based on spacing around a vehicle, authors defined safety space as an elliptical boundary, which they maintain in traffic stream. Further, based on position heat maps of vehicles in traffic streams, Chen et al. [86] estimated the safety space for motorized two-wheelers and cars in mixed traffic. Even though researchers attempted to define traffic conflict in non-lane-based traffic, these studies were limited to motorized two-wheelers and cars.

SSM-based safety assessment in non-lane-based traffic conditions is mostly based on previously defined temporal conflict indicators. These indicators are defined considering one-dimensional vehicle interactions based on longitudinal proximity alone. For example, researchers have utilized TTC, PET, and MTTC to quantify traffic conflict in non-lane-based traffic conditions [124, 130, 150, 152]. Further, few researchers have investigated traffic conflict during lane changing and passing maneuvers where vehicle maintain a lateral gap along with longitudinal gap [96, 97, 189, 201–203]. Although these studies attempted to define conflict in non-lane-based interaction, these were only limited to one-dimensional vehicle interactions. These studies did not account for lateral vehicle interactions. So far, research on traffic conflict-based road safety assessment has mainly focused on one-dimensional longitudinal vehicle interactions, and conflict has been defined using different longitudinal proximity indicators alone. TTC, the most used conflict indicator, is suitable for defining conflict when vehicles are on a collision course and interactions are only in longitudinal directions [204]. Das and Maurya [43] investigated the variation of TTC with lead vehicle type and lateral spacing. They reported that in non-lane-based traffic, TTC is dependent on lead vehicle type and lateral spacing. They further suggested including lateral vehicular interaction along with TTC while defining conflict in these scenarios. In a recent study, Venthuruthiyil and Chunchu [31] defined a novel conflict indicator called as Anticipated Collision Time (ACT) that could define all types of conflict. While ACT could be useful, extensive validation to use it in non-lane-based traffic is required.

From the past literature, it has been found that proximity-based conflict indicators are widely used for safety assessment. TTC, a longitudinal proximity indicator, has been used as a valid conflict indicator for defining conflict and crash risks in different traffic scenarios. Further, the literature suggests that TTC is only appropriate for defining conflict in one-dimensional longitudinal vehicular interactions. Vehicular interaction in non-lane-based traffic scenarios is 2-dimensional; therefore, using longitudinal indicators alone to define conflict is inappropriate. Although a plethora of literature is available on defining conflict in lane-based traffic interactions, only limited research has been conducted concerning the applicability of these safety indicators and to define conflict in a non-lane-based vehicle following.

To fill the aforementioned gap, crash risk associated with rear-end and side-swipe conflicts is estimated for non-lane-based traffic scenarios. This study develops a bivariate model to utilize lateral and longitudinal conflict indicators together (2-dimensional interactions) to define traffic conflict.

5.3 Data

This chapter utilized trajectory data of around 8000 vehicles as described in Chapter 4. In this Section, the estimation of proximity indicators from the extracted vehicle trajectories and their summary statistics are presented. The crash data used in this study is presented in Table 4.1.

5.3.1 Extracting conflict indicators

Conflict indicators were defined so as to quantify rear-end and side-swipe conflicts. In this regard, vehicular interactions on the major road were filtered out and considered further in this study. TTC was originally defined in a homogeneous traffic scenario where vehicles share a common path (Hayward 1971), and the subject vehicle interacts with a leader vehicle primarily in the longitudinal direction. TTC, a conflict indicator, sufficiently captures vehicular interactions in one-dimension [204]. However, a conflict indicator in non-lane-based traffic would be a function of both the longitudinal and the lateral interactions. Therefore, a 2-dimensional conflict indicator must be defined. In non-lane-based traffic scenarios TTC and lateral gap (Gap_{lat}) can be used to quantify longitudinal and lateral proximity, respectively. For non-lane-based traffic streams, TTC is computed as the ratio of longitudinal gap between the interacting pair (leader and rear follower) of vehicles and the relative speed between them as shown in Eqn. 5.1.

$$\text{TTC} = \begin{cases} \frac{X_{i-1}(t)-X_i(t)-L_{i-1}}{V_i(t)-V_{i-1}(t)} & \text{if } V_i(t) > V_{i-1}(t) \\ \text{not defined} & \text{if } V_i(t) \leq V_{i-1}(t) \end{cases} \tag{5.1}$$

Where, $i-1$ and i imply the leader and follower vehicles respectively. X, L and v represent longitudinal position, vehicle length and the speed, respectively. Gap_{lat}, the second conflict indicator, was computed as shown in Eqn. 5.2.

$$\text{Lateral gap } (\text{Gap}_{\text{lat}}) = y_{i-1} - y_i - \left(\frac{w_{i-1}+w_i}{2}\right) \tag{5.2}$$

$$\text{Longitudinal gap } (\text{Gap}_{\text{lon}}) = x_{i-1} - x_i - w_{i-1} + l_i \tag{5.3}$$

Where, $i-1$ and i imply the leader and follower vehicles respectively, while x, y, l and w represent longitudinal position, lateral position, length and vehicle width, respectively.

TTC and minimum Gap_{lat} were both positive and negative based on the relative vehicle positions. Since this study aims to model conflicts, TTC values between -10 to 10 s and Gap_{lat} less than 5 m which contains all possible conflicts were used for modelling. Descriptive statistics of both the indicators are presented in Table 5.1.

Table 5.1 Descriptive statistics of conflict indicators (TTC and Gap_{lat})

Site	**Sample Size**	**TTC**			**Gap_{lat}**		
		Median	**Mean**	**SD**	**Median**	**Mean**	**SD**
Site1	1000	1.60	2.43	3.27	1.20	1.05	0.88
Site2	1173	3.10	3.20	3.41	0.60	0.48	0.97
Site3	562	0.85	2.10	2.82	1.00	0.91	0.81
Site4	363	0.80	2.13	3.38	1.30	1.10	0.93

5.4 Defining traffic conflicts in non-lane-based traffic

Depending on the relative position of vehicles, vehicular interactions can be grouped into three categories, as depicted in Chapter 1, Fig. 1.1. Inline interactions involve interaction of vehicles moving with negative lateral gap and positive longitudinal gap. Interactions with positive lateral and longitudinal gaps as oblique interactions, and interactions with positive lateral gap and negative longitudinal gaps as parallel interactions. TTC is generally defined for inline interactions and appropriate to define conflict and crash risk for these interactions only [204], where TTC less than or equal to zero represents a crash. In non-lane traffic, as vehicular interaction is in the longitudinal and lateral direction, defining conflict using TTC alone will lead to bias since it cannot capture the conflict in oblique and parallel interactions.

In the present study, TTC and Gap_{lat} were used together for defining conflicts with the help of bivariate POT approach. Bivariate POT approach is used for modeling the probability of joint exceedances of two variables. For modeling the minimum of two variables, negative values of TTC and Gap_{lat} were used. Using bivariate GPD model, a conflict can be defined as extreme situations where both the variables exceed their pre-specified thresholds. Further, when a crash will occur when both the variables become zero together. Classification of vehicular interactions into normal, conflicts and crashes based on threshold values of both indicators is depicted in Fig. 5.1.

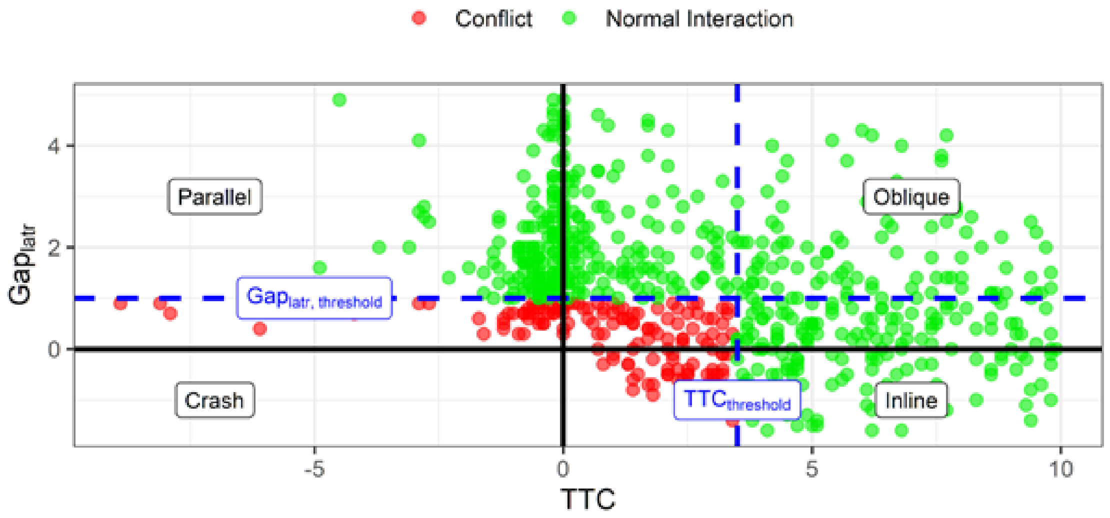

Fig. 5.1 Conflict and crash in non-lane-based traffic for all site data.

Using both the indicators, identified conflict observations as extreme were fitted using a bivariate GPD model as in previous studies [187, 205]. The probability of crash (R), can be calculated using estimated model parameters as shown in Eqn. 5.4.

$$R = \Pr((-TTC \geq 0) \vee (-Gap_{\text{lat}} \geq 0)) = 1 - G(0,0) = 1 - \exp\left[-\left(\tilde{x}^{-1/\alpha} + \tilde{y}^{-1/\alpha}\right)^{\alpha}\right] \tag{5.4}$$

Further, annual crash frequency (N) can be computed as (Zheng et al. 2018):

$$N = \frac{365 \times 24 \times R}{T} \tag{5.5}$$

Where, T is the conflict observation period in hours.

5.5 Results and Discussion

5.5.1 Dependence Structure between TTC and Gap$_{\text{lat}}$

Bivariate EVT model, using conflict indicators that are independent, leads to a better crash estimate [206]. Correlation analysis was performed to check the relation between these two variables. TTC was found to be negatively dependent on Gap$_{\text{lat}}$ and positively dependent upon longitudinal gap based on correlation test mentioned in Table 5.2. Chi-plot [207] was used to explore the nature of dependence among the conflict indicators (Fig. 5.2). Tests show that TTC is negatively dependent on lateral and positively dependent on the longitudinal gaps. This indicates that driver behavior is two dimensional and vehicles interact laterally as well longitudinally. Therefore, using longitudinal proximity alone is not sufficient to model crash risk. In this study, Gap$_{\text{lat}}$ along with TTC was used for bivariate extreme value modelling of conflicts in heterogeneous traffic scenarios. In bivariate extreme value modelling, variables are assumed to be either asymptotically dependent or perfectly independent. When the dependence between variables disappears at extreme levels, standard Bivariate EV modelling approach is not recommended [208]. TTC and Gap$_{\text{lat}}$ were found to have weak negative dependence (Fig. 5.2 and Table 5.2).

Table 5.2 Correlation between lateral and longitudinal parameters

Interaction Type	**Parameters**	**Pearson's rho (p-value)**	**Kendall's tau (p-value)**	**Spearman's rho (p-value)**
Staggered car-following	TTC and Gap_{long}	0.59 ($p <$ 2.2e-16)	0.50 ($p <$ 2.2e-16)	0.67 ($p <$ 2.2e-16)
	TTC and Gap_{lat}	-0.36 ($p <$ 2.2e-16)	-0.27 ($p <$ 2.2e-16)	-0.39 ($p <$ 2.2e-16)

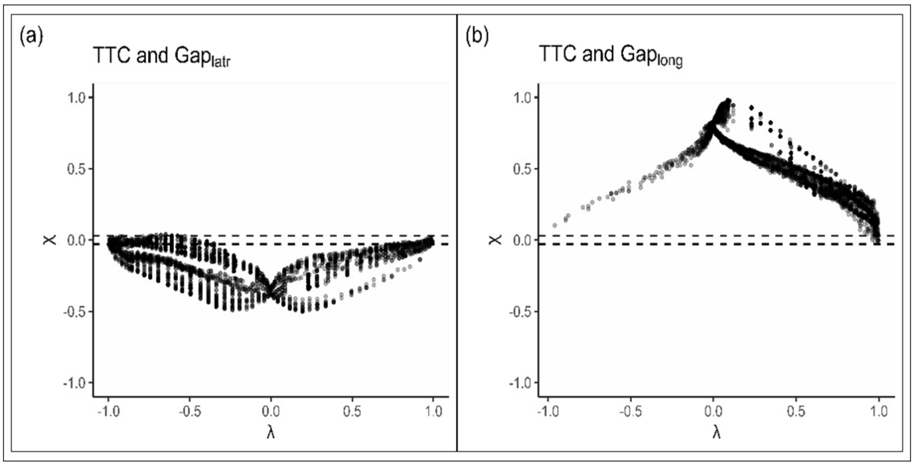

Fig. 5.2 Fisher's chi-plot for test of dependence.

In order to apply standard bivariate models, further tests were conducted to examine extremal dependency. From the extremal index plot (Fig. 5.3), dependence structure between variables can be obtained at extreme levels. The value of Chi (χ) is close to zero at higher quantiles, indicating asymptotic independence. Furthermore, the negative value of Chi-bar ($\bar{\chi}$) indicates weak negative association between variables at the extreme level [209].

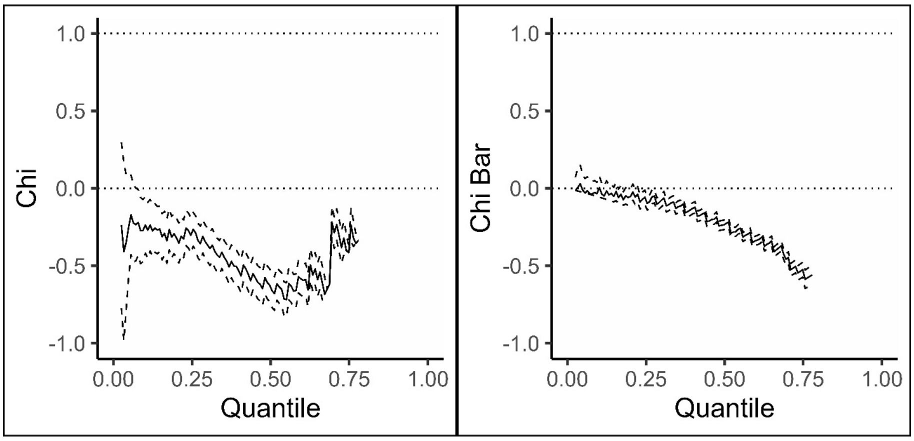

Fig. 5.3 Asymptotic dependence between negated TTC and negated Gap_{lat}.

5.5.2 Determining the threshold

For fitting threshold excess models, selection of suitable thresholds is a critical step. There are a few EVT-based techniques that are used to select a suitable threshold. For a marginal distribution, threshold could be selected using Mean residual life plots and threshold stability (TS) plots. Spectral measure plot is used to select the threshold for bivariate cases [193]. Previous literature has utilized a combination of these plots to select a proper threshold in case of univariate and bivariate EVT models. In the present study, TS plot and spectral measure plot were used for threshold selection. TS plot is based on the argument that if threshold excesses for a random variable follow GPD, then for any higher threshold, observations above the threshold would also follow a GPD. The parameters (shape and scale parameter) of the new distribution will be almost same, with some sampling variability. This indicates that parameters will not be exactly constant but should be stable with some sampling error [172]. Using TS plot, a threshold is selected from the curve, where scale and shape parameters are almost constant.

5.5.3 EVT modeling results

Selection of threshold for joint distribution was done using Spectral measure plot. This plot is based on the argument that, theoretically, at a large threshold value, empirical and theoretical spectral measure should be the same. For the bivariate case, the value of the theoretical spectral measure is 2. Therefore, the plot directs toward the location (k_0) of the threshold corresponding to spectral measure value 2. The highest ($k - k_0$) sorted observations would be used as exceedances for joint distribution. More details about these threshold selection methods can be found in Coles [172] and Beirlant et al. [193]. Fig. 5.4 and Fig. 5.5 show both the threshold plots for the joint-site model used in present study. Further, as suggested in previous studies, final thresholds were selected from the region based on both plots [185, 190]. The selected thresholds for all the models are listed in Table 5.3.

Logistic distribution has been mostly used to model bivariate joint distribution of extreme values of conflict indicators [190, 205]. The parameters of bivariate GPD models were estimated using the censored maximum likelihood estimation method. Values above the thresholds were used to construct the likelihood function. Model estimation results for the separate and joint-site models, along with predicted crashes, are presented in Table 5.3. Crash probability was estimated using parameters of the bivariate GPD model (Eqn. 5.3). Also, the total number of crashes were estimated using Eqn. 5.4. Further, 95% confidence interval of the estimated crashes were simulated assuming normal distribution of bivariate GPD model parameters as suggested by Zheng, Ismail and Meng [182].

Table 5.3 Estimation results of bivariate threshold excess models

Models	**Thresholds**		**Estimates (standard error)**					**Joint exceedances**	**Log-likelihood value (LL)**	**Estimated crash probability**	**Crash rate (95% C.I.)**
	$-$TTC (u_x)	**$-$Gap$_{lat}$ (u_y)**	**σ_x (SE)**	**ξ_x (SE)**	**σ_y (SE)**	**ξ_y (SE)**	**α (SE)**	**(Proportion)**			
Site1	-2.5	-0.5	2.56 (0.09)	-0.25 (0.01)	0.74 (0.04)	-0.30 (0.02)	0.99 (2E-06)	59 (0.06)	-2435	0.14	276 [244, 312]
Site2	-2.5	-0.8	2.99 (0.12)	-0.23 (0.02)	1.60 (2E-06)	-0.57 (2E-06)	0.75 (0.02)	157 (0.13)	-3461	0.20	396 [371, 419]
Site3	-2.0	-1.0	2.16 (0.13)	-0.19 (0.05)	1.06 (0.09)	-0.32 (0.09)	0.99 (2E-06)	110 (0.12)	-1473	0.23	445 [360, 532]
Site4	-3.0	-1.0	3.67 (0.23)	-0.39 (0.03)	1.08 (0.11)	-0.27 (0.07)	0.99 (2E-06)	46 (0.12)	-1046	0.19	369 [291, 445]
Joint-site Model	-3.0	-1.0	3.43 (0.07)	-0.28 (0.01)	1.30 (2E-06)	-0.40 (2E-06)	0.75 (0.01)	648 (0.20)	-11436	0.24	115 [95, 128]

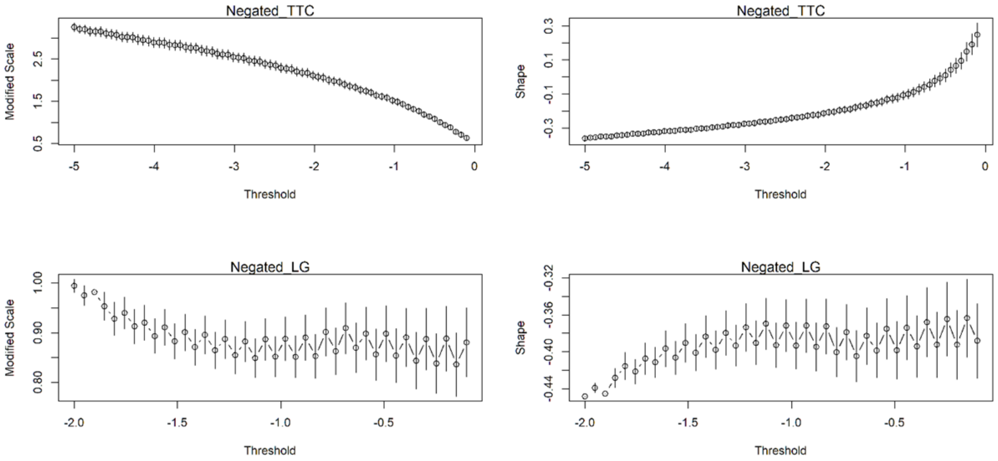

Fig. 5.4 Threshold stability plot for joint-site model.

Diagnostic plots for the joint-site model are presented in Fig. 5.6. The dependence parameter (α) was found to be in the range of 0.75 to 0.99. This implies that variables have a very weak dependence at extremes. Further, based on the correlation test as mentioned in Section 5.1, a weak negative dependence was obtained between the variables TTC and Gap_{lat}. In the field, lower TTC values are compensated by road users by maintaining a higher Gap_{lat}. This means in any interacting vehicle, TTC and Gap_{lat} simultaneously will not be approaching the extreme. The quantile curves are convex and have similar shapes, indicating a good fit for the bivariate data [189]. Further Pickand dependence function plot for the bivariate vector is symmetric, indicating a well-fitted model.

5.5.4 Overall model performance

The EVT approach can be used for modeling crash risk based on conflict data [181]. Moreover, during the modeling, crash data are not required and are only used for validation. For assessing crash risk, the present study uses the EVT approach which has been validated by several studies [183, 187, 190, 210–215]. The estimated crash risk increases with the risk of conflict (correlation;r=0.98; p-value=0.001). This supports the suitability of

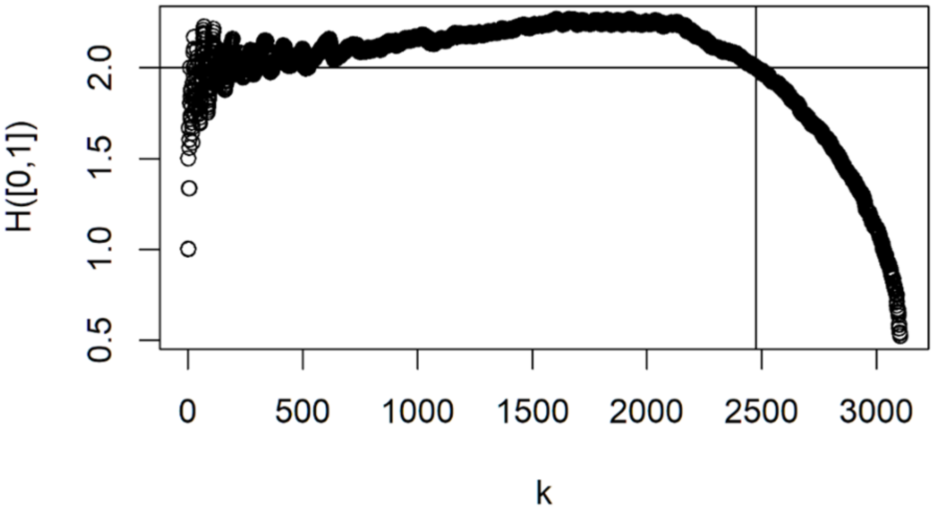

Fig. 5.5 Spectral measure plot for joint-site model.

using the bivariate model (TTC and $\mathrm{Gap_{lat}}$) to define conflict and hence, crash risk in non-lane-based traffic conditions.

In the present study, fatal crash data for a period of only 1 year were available for all 4 sites located on state highways (Table 4.1). Most available crash data include only fatal crashes, which is often erroneous. This is a fundamental issue in the historical crash database available across all developing countries [45].

Based on MORTH (Ministry of Road Transport and Highways) report on accidents on Indian roads [216], fatal crashes on Indian state highways were reported to be around 25% of total crashes. In this study, for comparing the estimated crash results, total crashes were obtained, assuming fatal crashes to be 25% of the overall crashes. The interval estimates for observed crashes were obtained using $\lambda : \frac{1}{2n}\chi^2_{2y_0,(1-\alpha/2)} \leq \lambda \leq \frac{1}{2n}\chi^2_{2(y_0+1),(\alpha/2)}$, as suggested by Songchitruksa and Tarko [181], where α is the level of significance, n is the duration of crash data, and y_0 is the number of observed crashes. The confidence intervals for crashes at four sites are calculated to be [9.2, 26], [18.6, 40.4], [9.2, 26], and [28.6, 54.4], respectively. The confidence interval for the joint-site was found to be [81.4, 122.7].

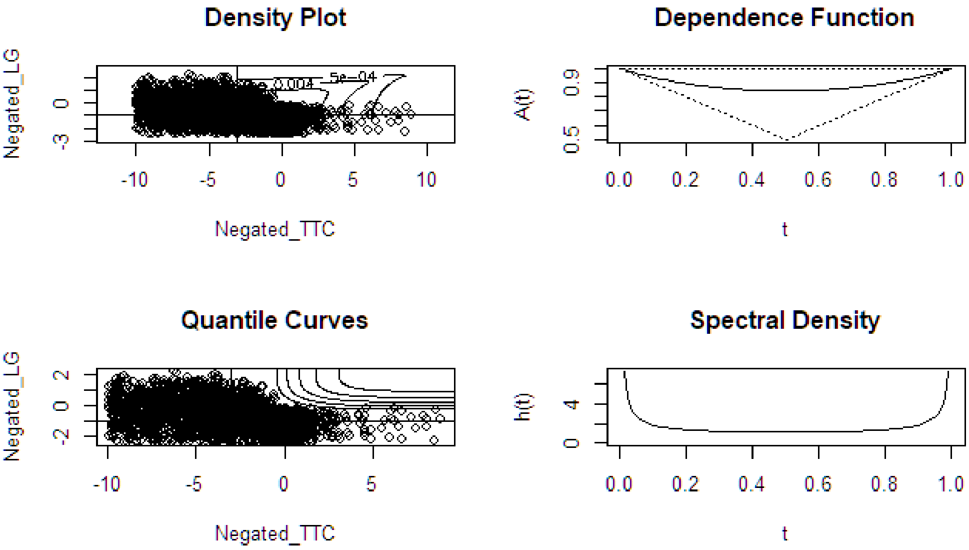

Fig. 5.6 Diagnostic plots for the joint-site model.

For the joint-site model, crash estimates (Table 5.3) were close to the observed crashes, and the majority of estimated crash confidence intervals overlapped with the Poisson confidence interval of observed crashes. This supports the usability of the bivariate GPD model for defining conflict and crash risk in non-lane-based traffic conditions. Further, the site-based estimates overpredict the crashes. The possible reason for this deviation could be the short duration of the conflict observation period. The uncertainty in crash estimates gets amplified by the large difference in conflict and crash observation periods. For instance, in the estimate based on 4.5 hours of conflict observation period, an inaccuracy of $\pm 1\%$ leads to an inaccuracy of around ± 20 crashes $(0.01 \times \frac{24}{4.5} \times 365 = 19.5)$, and an inaccuracy of $\pm 5\%$ in estimate leads to an inaccuracy of around ± 97 crashes $(0.05 \times \frac{24}{4.5} \times 365 = 97.33)$. This issue has also been highlighted in recent studies [190, 206].

5.5.5 Comparison of separate and joint-site models

The joint-site model-based predicted crashes (total 115 crashes per year) (Table 5.3) are much closer to the observed crashes at 4 sites (total 100 crashes). Also, a more important finding is that the joint-site model tends to generate crash estimates (95% CI for total crashes [95, 128]) that overlap with the Poisson confidence interval of observed crashes [81, 123]. Separate site-based model tends to overpredict the total number of crashes each year, ranging between 244 to 532 crashes at each site. The 95% confidence interval for the joint-site model is much smaller than the separate site-based model, suggesting a more precise estimate. This signifies that with an increase in sample size due to complete pooling, prediction capability of the fitted model improves significantly. The accuracy of joint-site model is better than either of its marginal models, and it is expected to be improved further with data from a longer observation period. This finding is similar to the finding of Arun et al. [190].

5.6 Chapter Summary

The present study proposed a bivariate extreme value modeling approach for defining conflict in non-lane-based traffic. Most of the previously defined conflict indicators are only suitable for one-dimensional vehicle interactions as in lane-based traffic. For example, rear-end conflicts can be quantified using TTC in lane-based traffic where vehicles are moving in strict lane discipline with 1-D interactions. However, it is not appropriate to use such conflict indicators alone to define conflict when vehicles move in non-lane-based traffic where 2-dimensional interaction is more common.

This study considers lateral along with longitudinal interactions to define conflicts in such traffic streams. Bivariate EVT model was proposed for crash risk assessment using TTC and Gap_{lat}. The results show that incorporating lateral and longitudinal conflict indicators together into the bivariate models can significantly improve the conflict-based risk assessment in non-lane-based traffic. Another important finding is that the joint-site model tends to generate crash estimates closer to observed crashes than separate site-based models.