This is a Portion of Integrated Crash Prediction and Resource Allocation Model

Crash Prediction Model (CPM)

Crash frequencies on a highway section are discrete and non-negative integer values, the Poisson regression technique is a natural first choice for modelling such data. However, past research has indicated that accident frequency data are likely to be over dispersed, making negative binomial regression a more appropriate choice (Washington, Karlaftis, and Mannering 2011). Using a negative binomial regression model, the probability of *t* crashes occurring at intersection *i* is given by

$$P(t_i) = \frac{{\lambda_i}^{t_i} \exp(-t_i)}{t_i!} \tag{1}$$

where $P(n_i)$ is the probability of *n* crashes occurring on an intersection *i* over a one year time period, and λ_i is the expected accident frequency for intersection *i*, that is, $\lambda_i = E(n_i)$. When applying the Poisson model, the expected accident frequency is assumed to be a function of explanatory variables such that

$$\lambda_i = \exp(\beta X_i + \varepsilon_i) \tag{2}$$

where X_i is a vector of explanatory variables that could include geometry, traffic characteristics, and weather conditions of highway section *i* that determine accident frequency, and β is a vector of estimable coefficients. Assuming that $\exp(\varepsilon_i)$ is a gamma-distributed disturbance term with mean of 1 and variance of α, we have

$$Var[t_i] = E[t_i][1 + \alpha E[t_i]] \tag{3}$$

The resulting probability distribution for the negative binomial distribution is

$$P(t_i) = \frac{\Gamma(t_i + \frac{1}{\alpha})}{t_i!\,\Gamma\left(\frac{1}{\alpha}\right)} \cdot \left(\frac{\frac{1}{\alpha}}{\frac{1}{\alpha} + \lambda_i}\right)^{\frac{1}{\alpha}} \left(\frac{\lambda_i}{\frac{1}{\alpha} + \lambda_i}\right)^{t_i} \tag{4}$$

This negative binomial model is used to predict crashes at intersection level with given highway geometry, and traffic conditions.

Resource Allocation Model

In the proposed Highway Safety-Resource Allocation Model (HS-RAM), the objective is to maximize the total benefits (*Z*) derived from crashes prevented at set of locations upon implementation of alternatives for the proposed planning period of *N* years. The integer optimization model is based on three binary variables, indexed by the intersection *i*, safety improvement choice *j*, and year of implementation *n*. Each improvement j has an effective duration of l_j years. The binary variable $x_{i,j}^{n} = 1$ if alternative j is implemented at location *i* during year *n*, and $y_{i,j}^{n,n'} = 1$ if alternative *j* is implemented at location *i* during year *n*, and is still active during year *n'*. (i.e. $y_{i,j}^{n,n'} = 1$ if $x_{i,j}^{n} = 1$ and $0 \le n' - n \le l_j$). Here, *x* indicates the year of construction, while *y* indicates the years of effectiveness. The objective function is based on maximization of benefits, with several constraints: a budget constraint, constraints based on the feasible alternatives for each intersection, and definitional constraints relating *x* and *y*.

3.2.1 Objective Function

Let f_i^n, m_i^n, and p_i^n denote the expected number of fatal crashes, injury or non-fatal crashes, and property damage only (PDO) collisions at location *i* during year *n*. Similarly, assuming $r_{i,j}^{f}$, $r_{i,j}^{f}$, and $r_{i,j}^{f}$ denote the crash reduction factors for these three types of crashes if treatment *j* is applied at intersection *i*, and let c^f, c^m, and c^p denote the economical costs of each type of crash obtained from NSC, 2013. National Safety Council estimates the average costs of fatal and nonfatal unintentional injuries to illustrate their impact on the nation's economy. According to NSC 2013, the costs are a measure of the dollars spent and income not received due to accidents, injuries, and fatalities that is another way to measure the importance of prevention work. Hence, the objective function can then be written as:

Maximize

$$Z = \sum_{n=1}^{N} \sum_{j=1}^{J} \sum_{i=1}^{I} \sum_{n'=1}^{N} \left[f_i^n r_{i,j}^f c^f + m_i^n r_{i,j}^m c^m + p_i^n r_{i,j}^p c^p \right] y_{i,j}^{n,n'} \tag{5}$$

3.2.2 Constraints

Equation (6) is a budget constraint, that ensures the sum total of capital investment and operation and maintenance (O&M) cost should not exceed the total budget in the planning period. However, there is a flexibility of expenditure between the years in the planning period. Such flexibility in expenditure between years within a planning period can be incorporated into the procedure through a planning based budget model applied in transit resource allocation (Mishra et al. 2013). In these models a planning period budget is based on the assumption that the agency has the flexibility of borrowing monies from subsequent years' allocation or past year surplus. Let $\pi_{i,j}^{n}$ represent the capital cost of constructing improvement j at intersection i in year n, and $o_{i,j}^{n,n'}$ the operating costs in year *n'*. Also, let b_n be the available budget available for year *n*. Then we require

$$\sum_{n=1}^{N} \sum_{j=1}^{J} \sum_{i=1}^{I} \left[\pi_{i,j}^{n} x_{i,j}^{n} + \sum_{n'=1}^{N} o_{i,j}^{n,n'} y_{i,j}^{n,n'} \right] \le \sum_{n=1}^{N} b_n \tag{6}$$

For a variety of reasons, not all alternatives can be implemented at all locations. Accordingly, constraint (7) ensures that the alternatives implemented at a location, using pre-specified parameters $\hat{x}_{i,j}^{n}$:

$$x_{i,j}^{n} \le \hat{x}_{i,j}^{n} \text{ for all } i, j, n \tag{7}$$

Expression (8) denotes that each location *i* can have a limited number of active alternatives (γ_i^n) during the analysis year *n*, pre-specified by the planning agency.

$$\sum_{j=1}^{J}\sum_{n=1}^{N} y_{i,j}^{n,n'} \le \gamma_i^n \quad \text{for all } i, n' \tag{8}$$

When the alternatives are mutually exclusive, as in the base case, γ_i^n is uniformly equal to one. This provides the following features:

- *Feature 1*: A location can receive only one alternative in a given year.
- *Feature 2:* A location, that has the carry-over effect from an alternative implemented in previous years, may not receive any funds during the service life of the alternative. (Note: This constraint can be modified as desired).

Furthermore, the definitions of *x* and *y* require

$$y_{i,j}^{n,n'} \le \begin{cases} x_{i,j}^{n} & 0 \le n' - n \le l_j \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

$$x_{i,j}^{n} \le y_{i,j}^{n',n} \text{ for all } i, j, n, n' \tag{10}$$

$$x_{i,j}^{n}, y_{i,j}^{n,n'} \in \{0,1\}, \forall\, i, j, k_j \tag{11}$$

Equation (9) requires that the *y* values are consistent with the *x* values such that an improvement cannot be active at a given year unless it was implemented in a year within its duration of effectiveness. Equation (10) prohibits an already-active improvement from being selected again during its duration of effectiveness. Finally, Equation (11) reflects the binary nature of the decision variables.

3.3 Integration of CPM and HS-RAM Model

Both CPM and HS-RAM are integrated for simultaneous prediction of crashes and performing optimal resource allocation. The outcome of CPM serves as input to the resource allocation model (HS-RAM) for benefit maximization in the planning period. The crash prediction model forecasts probability of number of crashes for each location ($P(t_i)$ - see equation (4)) but is dependent upon HS-RAM for inferring best preventative improvements locations for maximum crash reduction. Advantages of crash prediction model lies in the fact that it eliminates the unrealistic assumption of deterministic growth factors for crash prediction.

4. MODEL APPLICATION

Table 1 lists a vector of estimable coefficients (β) for crash prediction model (equation (2)). The resulting coefficients (Table 2) are used in integrated crash prediction model to determine the probable frequency of crashes at each location over the planning period. We developed these coefficients from the comprehensive crash and roadway dataset made available by MDOT. The derived coefficients in Table 3 appear intuitive from the viewpoint of magnitude and sign. For example ADT of the major road has positive sign suggesting that higher ADT causes the probability of increasing crashes at the intersection.

Table 1: Coefficients for Crash Prediction Model (CPM)

Independent variables	Fatal Coefficient (t-stat) (β)	Injury Coefficient (t-stat) (β)	PDO Coefficient (t-stat) (β)
Constant	-13.379** (*-2.842*)	-1.859*** (*-3.658*)	-1.612*** (*-4.336*)
ln (ADT of major road)	0.978* (*2.019*)	0.264*** (*4.781*)	0.269*** (*6.841*)
Five or more number of lanes on major street		0.220* (*2.547*)	0.168** (*2.933*)
Presence of exclusive left turn phase on minor street		-0.730*** (*-3.889*)	
Five or more number of driveways on minor street		0.175* (*1.801*)	0.138** (*2.118*)
Speed limit more than 30 mph		0.150* (*1.657*)	0.218*** (*3.369*)
Parking lane on minor street		-0.196* (*-1.762*)	-0.216** (*-2.934*)
Presence of divider with barrier		0.018** (*2.928*)	0.141** (*2.194*)
Intersection in close proximity to freeway		0.188*** (*9.961*)	0.527*** (*6.513*)
Hazard rating more than 3		-0.200*** (*-7.609*)	0.265*** (*4.490*)
Cycle length over 60 seconds		-1.443*** (*-7.151*)	0.167** (*2.565*)
Rho-square	0.063	0.070	0.289

Note: t-statistics are in parenthesis
*** Significant at 99%; ** Significant at 95%; * Significant at 90%