## **Problem Formulation**

### **Decision Problem**

Determine the optimal locations for temporary EMT stations/field offices across the Pennsylvania highway network for each quarter, minimizing the distance between EMT stations and predicted crash locations while considering resource constraints and temporal variations in crash patterns.

### **Decision Variables**

* **x\_ij^t** ∈ {0,1}: Binary variable, 1 if EMT station i is assigned to highway section j in quarter t, 0 otherwise
* **y\_i^t** ∈ {0,1}: Binary variable, 1 if EMT station i is activated in quarter t, 0 otherwise
* **z\_j^t** ∈ {0,1}: Binary variable, 1 if highway section j is covered by at least one EMT station in quarter t, 0 otherwise

### **Parameters**

* **d\_ij**: Distance from potential EMT location i to highway section j
* **p\_j^t**: Predicted number of crashes in section j during quarter t
* **K**: Maximum number of EMT stations available
* **R\_max**: Maximum response distance threshold
* **C\_i**: Cost of operating EMT station at location i
* **B**: Total budget constraint

### **Objective Function**

**Minimize Z = Σ\_t Σ\_j Σ\_i p\_j^t × d\_ij × x\_ij^t**

This minimizes the weighted distance between EMT stations and crashes, where weights are the predicted crash frequencies.

### **Constraints**

1. **Coverage constraint**: Each section must be assigned to exactly one EMT station (if covered)  
   * Σ\_i x\_ij^t = z\_j^t, ∀j,t
2. **Station activation**: Can only assign from activated stations  
   * x\_ij^t ≤ y\_i^t, ∀i,j,t
3. **Capacity constraint**: Limited number of stations  
   * Σ\_i y\_i^t ≤ K, ∀t
4. **Budget constraint**:  
   * Σ\_i Σ\_t C\_i × y\_i^t ≤ B
5. **Maximum response distance**:  
   * x\_ij^t × d\_ij ≤ R\_max, ∀i,j,t
6. **Coverage requirement** (optional): Minimum percentage of high-risk sections covered  
   * Σ\_j (z\_j^t × I(p\_j^t > threshold)) ≥ α × |high-risk sections|, ∀t

## **Method 1: Predictive Analytics + Mixed Integer Linear Programming**

### **Step 1: Crash Prediction Model**

* Use historical crash data with features:
  + Weather patterns by quarter
  + Traffic volume variations
  + Road conditions
  + Historical crash frequencies
  + Special events/holidays
* Apply time series models (ARIMA, Prophet) or ML models (Random Forest, XGBoost) to predict p\_j^t

### **Step 2: MILP Optimization**

* Solve the formulated MILP using:
  + Commercial solvers: CPLEX, Gurobi
  + Open-source: CBC, GLPK
* Implement quarterly re-optimization with updated predictions

### **Step 3: Rolling Horizon Approach**

* Solve for all 4 quarters simultaneously but with decreasing weight for future quarters
* Modified objective: **Z = Σ\_t w\_t × Σ\_j Σ\_i p\_j^t × d\_ij × x\_ij^t** where w\_t = 1/(1+0.1×(t-1)) gives more weight to near-term quarters

## **Method 2: Robust Optimization with Scenario-Based Stochastic Programming**

This method accounts for uncertainty in crash predictions rather than using point estimates.

### **Step 1: Scenario Generation**

* Generate S scenarios for crash occurrences: p\_j^t(s) for scenario s
* Use Monte Carlo simulation based on:
  + Historical crash distribution patterns
  + Confidence intervals from predictive models
  + Extreme event scenarios (severe weather, major accidents)

### **Step 2: Two-Stage Stochastic Program**

**First Stage Variables** (here-and-now decisions):

* **y\_i**: Binary, whether to establish EMT station at location i (fixed for all quarters)

**Second Stage Variables** (wait-and-see decisions):

* **x\_ij^t(s)**: Assignment decisions for each scenario and quarter

### **Modified Objective:**

**Minimize Z = Σ\_i C\_setup × y\_i + E\_s[Q(y,ξ(s))]**

Where Q(y,ξ(s)) is the recourse function: **Q(y,ξ(s)) = min Σ\_t Σ\_j Σ\_i p\_j^t(s) × d\_ij × x\_ij^t(s)**

### **Step 3: Solution Approach**

1. **Sample Average Approximation (SAA)**:  
   * Solve with finite scenarios: **Min Z = Σ\_i C\_setup × y\_i + (1/S)Σ\_s Q(y,ξ(s))**
2. **Progressive Hedging Algorithm**:  
   * Decompose by scenarios
   * Iteratively converge to implementable solution
3. **Chance Constraints** (alternative):  
   * Add reliability requirements: **P(response\_time ≤ τ) ≥ 1-ε**
   * Convert to deterministic equivalents using CVaR

### **Advantages of Each Method:**

**Method 1 (Prediction + MILP)**:

* Computationally tractable
* Clear interpretation
* Easy to implement with standard solvers
* Good for stable, predictable patterns

**Method 2 (Robust Stochastic)**:

* Handles uncertainty explicitly
* More resilient to prediction errors
* Better for high-variance situations
* Provides confidence bounds on performance

### **Implementation Considerations:**

1. **Data Requirements**: 3-5 years of historical crash data with temporal, spatial, and contextual features
2. **Computational Resources**: Method 2 requires more computational power (potentially parallel processing for scenarios)
3. **Update Frequency**: Quarterly re-optimization with monthly prediction updates
4. **Validation**: Use cross-validation on historical data, comparing actual response times vs. optimized allocations

Both methods can incorporate additional real-world constraints like EMT station setup/teardown costs, staff scheduling constraints, and coordination with existing permanent facilities.