[*CONCEPT*, Vol. XXXVIII (2015)]

Application of a Markov Chain Traffic Model to the Greater Philadelphia Region

Joseph Reiter
Mathematics and Statistics

Introduction

Traffic congestion in urban areas has been an issue since the beginning of the 20th century, and it continues to be one of the most persistent problems facing urban planners. The phenomenon has been extensively studied and several models have been used to explain where, when, and how vehicles move through a network of roads. Limited access highways changed how vehicles flowed in and around cities and across the country. They allow vehicles to maintain constant speeds between exits and are partially responsible for the expansion of cities (Rephann & Isserman, 1994).

The challenge in studying a road system populated with vehicles is that the number of parts interacting at the same time is very large. Each individual vehicle in the system must be considered on its own and also as a part of the system as a whole. In some ways, this system acts like a complex economy; each vehicle acts to maximize its utility by minimizing travel time, avoiding hazards, and seeking the least troublesome route to its destination. Each constituent part of the system has some effect on all the other parts in the system.

This paper looks at the issue of traffic congestion around cities and demonstrates a model that predicts areas of traffic congestion based on census data along with statewide traffic data. The model utilizes a Markov chain process in order to analyze the origin-destination paths of vehicles in a network system. This method is applied to the Greater Philadelphia Region and evaluated for its predictive value. The implications of this model and possible extensions are also considered.

Review of Traffic Models

Microscopic Models

Microscopic models focus on the individual interactions between vehicles and the roadway. Vehicles follow specific rules, outlined in the model, that govern their

movement in the system (Velasco & Saayedra, 2008). These rules relate values such as velocity, distance to nearby vehicles, and time. One common microscopic traffic model is "follow-the-leader", which utilizes a dynamical equation or system of differential equations relating the motion of the (n+1)*th* vehicle following the n*th* vehicle in a single lane (Gazis, Herman, & Rothery, 1961):

$$\ddot{x}_{n+1}(t+T) = \lambda[\dot{x}_n(t) - \dot{x}_{n+1}(t)],$$

where x_n is the position of the nth vehicle, T is the time lag of response to the stimulus, λ is the sensitivity, and the dots denote differentiation with respect to time t.

Nagel and Schreckenberg developed a cellular automata model that dictates rules for traffic movements. These rules account for how vehicles accelerate with relation to each other and move along the roadway. The product of the model is a time vs. space (road) plot which tracks how dense traffic is at each point in the road during a time interval. This plot shows how congestion patterns can travel in both time and space (Nagal & Schreckenberg, 1992).

A model put forth by Indrei utilizes Markov chains to construct a theoretical traffic system in a real space $\mathfrak{R}$ (Indrei, 2006). An object located in a particular element (i, j) of the configuration matrix transitions to a new location in discrete time. Each element of this matrix represents a position on the roadway. Vehicle movements are defined as a transition where rows represent highway lanes and columns represent length of the highway segments:

$$(i,j) \bowtie \big(f(i), j + \sigma(object)\big),$$

The symbol $\bowtie$ represents a natural join between elements of two matrices, $\mathbf{A_0}$ and $\mathbf{A_1}$. This relation describes how an object in the (i, j)*th* slot in a matrix $\mathbf{A_0} \in \mathfrak{R}$ will end up in the $(f(i), j + \sigma(object))$*th* slot of another matrix $\mathbf{A_1} \in \mathfrak{R}$ during one time step, where

$$\sigma(object) = \begin{cases} 1, & \text{with probability } p \\ 2, & \text{with probability } q \\ 3, & \text{with probability } r \\ & p + q + r = 1 \end{cases}$$

and

$$f: \{i\} \mapsto \{i, i+1, i-1\}$$

Mesoscopic Models

Another category of traffic models, called mesoscopic or kinetic traffic models, examines vehicular movements as parts of a larger scale mechanism. Headway distribution models focus on how much time passes between two successive vehicles. The time between vehicles (headway) is described by a random variable with either a single distribution or mix of different distributions. Some of these distributions are normal, gamma, and exponential (Zhang, Wang, Wei, & Chen, 2007).

Cluster models are characterized by groups of vehicles that all exhibit the same property, such as velocity. This clustering effect may be due to how a roadway narrows to fewer lanes or other factors such as weather. The size of each cluster, which determines how vehicles will flow on a highway, is dynamic and can grow or diminish over time (Hoogendoorn & Bovy, 2001).

The gas-kinetic model of traffic flow comes from statistical mechanics and models vehicles as particles in a traffic flow. Prigogine and Andrews developed this model based on the Boltzmann-Maxwell equation that models the velocity of ideal gas particles. The distribution of velocities $f(x,v,t)$, where x is position on road, v is velocity, and t is time (Prigogine & Andrews, 1960). This distribution function is described as:

$$\frac{\partial f}{\partial t} + v\frac{\partial f}{\partial x} = \frac{\partial f}{\partial t}_{rel} + \frac{\partial f}{\partial t}_{coll}$$

where df/dt_{rel} is the relaxation term (reflected by a desire to return to the ideal distribution) and df/dt_{coll} is the 'collision' term, which in this case refers to the interaction between vehicles ahead slowing down the ones behind.

Macroscopic Models

Macroscopic models describe averaged quantities in traffic such as density, average velocity, and velocity variance. Individual vehicles are not considered, but rather their aggregate behavior is examined in order to understand the overall traffic flow on a roadway. Lighthill and Whitham developed a theory that at any point of the road the flow q (vehicles per hour) is a function of the concentration k (vehicles per mile), and that the quantity in a small element of length (dk) changes at a rate equal to the difference between inflow and outflow (dq) (Lighthill & Whitham, 1955):

$$\frac{\partial k}{\partial t} + \frac{\partial q}{\partial x} = 0$$

This partial differential equation describes a phenomenon from physics called a kinematic wave. The dynamics of these waves describe how traffic jams can occur at one point on a road and travel backwards through the traffic flow. A model built upon the Lighthill/Whitham model that uses a system of partial differential equations was proposed by Harold J. Payne and includes a convection, relaxation, and anticipation term in the system (Hoogendoorn & Bovy, 2001):

$$V(x(t+T), t+T) = V^e(r(x+D, t)),$$

where $x(t)$ is the location of the vehicle at time t, $V(x,t)$ is velocity at x and t, V^e is the equilibrium velocity, $r(x,t)$ is the density, T is reaction time and D is gross-distance headway with respect to the preceding vehicle. The equation describes how drivers will adjust their velocity to an equilibrium velocity, which is affected by the traffic density. The left and right sides of this equation can be expanded and combined to form a single equation:

$$\partial_t V + V\partial_x V = \frac{V^e(r) - V}{T} - \left(\frac{c_0^2}{r}\right)\partial_x r,$$

where $c_0^2 = \xi/T > 0$ is the anticipation constant, and $\xi = -\mathrm{d}V^e/\mathrm{dr}$ is the decrease in the equilibrium velocity with increasing density. Other single and multi-equation models exist that extend the Lighthill/Whitham and Payne models to particular situations (Hoogendoorn & Bovy, 2001).

Two methods for describing traffic flow outlined by Sasaki and Myojin involve the use of Markov chains (Sasaki & Myojin, 1968). The first is the branching probability matrix method, in which the highway is divided into m sections. A volume of traffic over each section of highway is denoted x_i , and a row vector of volumes can be constructed $X = (x_1, x_2, \ldots , x_m)$. Inflows and outflows are also denoted within a row vector; U_i = inflow through ramp i $(i = 1,2,3,\ldots k)$, V_j = outflow through ramp j $(j = 1,2,3,\ldots r)$, and k = the number of on-ramps and r = the number of off-ramps. The branching probability matrix is written in canonical form:

$$P = \begin{pmatrix} I & 0 & 0 \\ R_1 & 0 & Q_1 \\ R_2 & 0 & Q_2 \end{pmatrix}$$

where

R_1= the probability of an inflow from on-ramp i to off-ramp

R_2= the branching probability of traffic flow from section *i* to off-ramp *j*
Q_1= the probability of an inflow from on-ramp *i* to section *j*
Q_2= the branching probability of traffic flow from section *i* to section *j*

Traffic flows over sections are given by:

$$X = uQ_1(I - Q_2)^{-1}$$

and outflows through ramps are given by:

$$v = u[R_1 + Q_1(I - Q_2)^{-1}R_2]$$

The second method described by Sasaki and Myojin is the route matrix method. If we are given the distribution of trips between on-ramps and off-ramps, we can form a route matrix, which describes how a vehicle entering at on-ramp *i* will travel to off-ramp *j*. This will be an (r × m) matrix in which the entry r_{im} is equal to 1 if the route from on-ramp *i* to off-ramp *j* includes section *m*, and equal to 0 otherwise. Each on-ramp will have its own matrix R_i. The trip distributions can be denoted by row vector $p_i = (p_{i1}, p_{i2}, \ldots, p_{ir})$, where p_{ij} = transition probability of a trip from on-ramp *i* to off-ramp *j*. The overall expected traffic flow over the traffic network is:

$$Q = \begin{pmatrix} p_1R_1 \\ p_2R_2 \\ \vdots \\ p_kR_k \end{pmatrix} = (Q_{ij}), (k \times m)$$

and gross actual traffic flow:

$$X = uQ$$

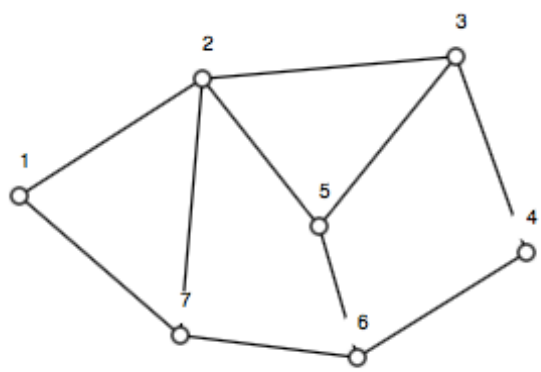
Another article describes this process in detail and applies it to a hypothetical traffic network (Crisostomi, Kirkland, Schlote, & Shorten, 2010). The literature refers to this approach as the origin-destination (OD) matrix method. The estimation of these matrices by various means is the subject of several articles (Van Zuylen & Willumsen, 1980) (Perrakis, Karlis, Cools, Janssens, & Wets, 2012) (Youngblom, 2013).

III. Model

The purpose of this model is to predict when and where heavy traffic is most likely to occur in a highway network. Since the focus of this model is on particular segments of highway rather than individual vehicles or traffic flow, it would need to be macroscopic in its level of detail. Sasaki and Myojin's model from the previous section provides a good foundation on which to build this model since their model's level of focus is on highway segments. But we should first reexamine the situation we are trying to model.

Representing a Highway System as a Matrix

A road network can be looked at as a system of exits connected by segments of highway that can be represented by a directed graph. The vertices of the graph represent exits and edges represent highway segments:



For this example, there are 7 exits connected by highway segments. An adjacency matrix **A** can be formed to describe this system:

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

A new matrix **D** representing the distance between exits in miles can be defined by inserting the distances between exits instead of 1s, replacing the 0s that represent unconnected exits with a very large number to simulate an infinite distance, and leaving the 0s in place for elements where $i = j$:

$$\mathbf{D} = \begin{pmatrix} 0 & 3 & \infty & \infty & \infty & \infty & 4 \\ 3 & 0 & 3 & \infty & 2 & \infty & 5 \\ \infty & 3 & 0 & 2 & 3 & \infty & \infty \\ \infty & \infty & 2 & 0 & \infty & 2 & \infty \\ \infty & 2 & 3 & \infty & 0 & 1 & \infty \\ \infty & \infty & \infty & 2 & 1 & 0 & 1 \\ 4 & 5 & \infty & \infty & \infty & 1 & 0 \end{pmatrix}$$

The matrix **D** describes the relevant information about how the exits are connected to each other, but we also need to know how many lanes of traffic there are on each segment of highway. A matrix to describe the number of lanes between exits **L** is defined, where elements representing segments that are not connected are given a value of -1 as a placeholder (avoiding possible division by zero later in the model):

$$\mathbf{L} = \begin{pmatrix} -1 & 2 & -1 & -1 & -1 & -1 & 2 \\ 2 & -1 & 2 & -1 & 3 & -1 & 3 \\ -1 & 2 & -1 & 2 & 3 & -1 & -1 \\ -1 & -1 & 2 & -1 & -1 & 2 & -1 \\ -1 & 3 & 3 & -1 & -1 & 1 & -1 \\ -1 & -1 & -1 & 2 & 1 & -1 & 3 \\ 2 & 3 & -1 & -1 & -1 & 3 & -1 \end{pmatrix}$$

The two matrices **D** and **L** together describe all the important information about the geometry of the highway system.

Determining Populations

Now that the road network is established, we can consider how vehicles travel on the system. Assume that at any time t there are a particular number of vehicles in the area of each exit. These populations can be represented by a row vector $\mathbf{q} = (q_1, q_2, \ldots, q_m)$, where q_i represents the population at exit i. So $\mathbf{q}_t = (q_{1t}, q_{2t}, \ldots, q_{mt})$ represents the populations at each exit at time t.

For this model, time takes on discrete values of one-hour increments, so for each hour a new population must be determined. This new population consists of vehicles that came from another exit as well as vehicles that did not travel away from their exit. First, consider the vehicles that come from another exit. Let the exit traveled from be e_i and the exit traveled toward be e_j. Then there is a probability that a vehicle leaving e_i goes to e_j which can be expressed as p_{ij}. A matrix **P** can be formed from these probabilities:

$$\mathbf{P} = \begin{pmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,n} \\ p_{2,1} & \ddots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ p_{m,1} & \cdots & \cdots & p_{m,n} \end{pmatrix}$$

This matrix represents a transition matrix that states the probabilities associated with traveling from exit *i* to exit *j*. Each row of this matrix will be stochastic (sum to 1).

Another probability representing the relative volume of traffic on the highway network during a particular hour must also be defined. Let $v(t)$ be a function representing the relative volume of traffic using the network during time interval t, then the number of vehicles leaving any particular exit during time interval t is given by $v(t)q_i$, and the total number of vehicles traveling to exit j is given by:

$$v(t)q_1p_{1,j} + v(t)q_2p_{2,j} + v(t)q_3p_{3,j} + \cdots + v(t)q_mp_{m,j} = v(t)\sum_{x=1}^{m} q_xp_{x,j}$$

The other portion of the new population consists of vehicles that did not leave their position. Since the relative proportion of vehicles that did not travel during a time increment is the complement of the proportion that did travel, the number of vehicles remaining is the complementary probability times the number of vehicles at a particular destination exit, given by:

$$[1 - v(t)]q_j$$

Combining these two expressions gives the new population for each exit:

$$q_{(\boldsymbol{t+1})\boldsymbol{i}} = v(t)\sum_{x=1}^{m} q_xp_{x,j} + [1 - v(t)]q_j$$

This equation can be expressed using matrices and vectors that have already been defined:

$$\mathbf{q}_{\boldsymbol{t+1}} = v(t)\mathbf{q}_{\boldsymbol{t}} \cdot \mathbf{P} + [1 - v(t)]\mathbf{q}_{\boldsymbol{t}}$$

This process of finding the next population is similar to using a Markov chain, except that there is an extra term for the proportion of a population not participating in the transition process.

Traffic Density

Since the goal of this model is to predict where and when heavy traffic occurs, it is necessary to determine how many vehicles are traveling on each segment of the highway system during each time interval. In order to determine how many vehicles are traveling on a segment, we must determine which routes pass through each highway segment. A complication occurs at this point in the process since there are several ways to go from one exit to another. To simplify this, an assumption is made that vehicles will follow the shortest path between two exits. It can be argued that this assumption is not completely valid and that other ways to determine a path may yield a more precise model. This paper does not consider other path-determining methods (Lim, Balakrishnan, Gifford, Madden, & Rus, 2011), yet nothing in the general framework of this model prohibits using other methods. An algorithm for finding the shortest paths on a graph between two vertices was developed by EW Dijkstra and will be used in this model.

The number of vehicles traveling on a particular route from exit i to exit j was determined above as $v(t)q_i p_{i,j}$. This value can be determined for each origin-destination pair and a matrix constructed from them:

$$\mathbf{R} = \begin{pmatrix} v(t)q_1p_{1,1} & v(t)q_1p_{1,2} & \cdots & v(t)q_1p_{1,n} \\ v(t)q_2p_{2,1} & \ddots & & \cdots \ \vdots \\ \vdots & \vdots & \ddots & \vdots \\ v(t)q_mp_{m,1} & \cdots & \cdots & v(t)q_mp_{m,n} \end{pmatrix} = \begin{pmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,n} \\ r_{2,1} & \ddots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ r_{m,1} & \cdots & \cdots & r_{m,n} \end{pmatrix}$$

We can construct an (m × m) square matrix **Q** where each row is **q**:

$$\mathbf{Q} = \begin{pmatrix} q_1 & q_2 & \cdots & q_m \\ q_1 & q_2 & \cdots & q_m \\ \vdots & \vdots & \ddots & \vdots \\ q_1 & \cdots & \cdots & q_m \end{pmatrix}$$

Then the route matrix **R** can be rewritten using matrices:

$$\mathbf{R} = v(t)\mathbf{P} \odot \mathbf{Q}^{\mathbf{T}}$$

where the symbol $\odot$ represents element by element multiplication of the matrices. The transpose of **Q** is needed since each row of **R** is concerned with a single origin exit.

This matrix **R** shows how many vehicles are traveling on a particular route, but we are interested in how many vehicles are traveling on a particular segment of highway. This can be determined by taking the sum of all vehicles on all routes that pass through this segment. A matrix **C** can be constructed to count the total number of vehicles passing through each segment. An element of **C** can be described:

$$c_{i,j} = \sum_{0<x\leq m} \sum_{0<y\leq n} r_{x,y} \text{ , for all } r_{x,y} \text{ which are shown to pass through segment } (i,j)$$

The routes $r_{x,y}$ that pass through segment (i, j) are determined by application of Dijkstra's Algorithm (Dijkstra, 1959) to the matrix **D**, which is the matrix that shows the distance between exits. From the example above:

$$c_{6,7} = r_{3,7} + r_{4,1} + r_{4,7} + r_{6,7}$$

Heavy traffic occurs in areas in which the traffic density passes above a certain limit, so it is necessary to determine the density for each section of highway. Since the number of vehicles per hour traveling on a particular segment is given by $c_{i,j}$, the density (vehicles per mile lane) can be determined from:

$$\rho = \frac{c_{i,j}}{s \cdot l_{i,j}},$$

where s is the average speed of vehicles in miles per hour and $l_{i,j}$ is the number of lanes from the matrix **L** on segment (i, j). This density can be compared to typical densities that occur during heavy traffic to yield a prediction about what the traffic flow should be like on this segment.

Hypothetical Example of the Model

Going back to the example from above, let us define a population vector **q** = (2000,3000,2500,4000,5000,3500,4500). The matrix **Q** is:

$$\mathbf{Q} = \begin{pmatrix} 20000 & 30000 & 25000 & 40000 & 50000 & 35000 & 45000 \\ 20000 & 30000 & 25000 & 40000 & 50000 & 35000 & 45000 \\ 20000 & 30000 & 25000 & 40000 & 50000 & 35000 & 45000 \\ 20000 & 30000 & 25000 & 40000 & 50000 & 35000 & 45000 \\ 20000 & 30000 & 25000 & 40000 & 50000 & 35000 & 45000 \\ 20000 & 30000 & 25000 & 40000 & 50000 & 35000 & 45000 \\ 20000 & 30000 & 25000 & 40000 & 50000 & 35000 & 45000 \end{pmatrix}$$

Let us also declare **P** with equal probabilities that a vehicle will come from any exit, which means that $p_{1,j} = p_{2,j} = p_{3,j} \ldots$:

$$\mathbf{P} = \begin{pmatrix} 0.1 & 0.15 & 0.15 & 0.2 & 0.25 & 0.05 & 0.1 \\ 0.1 & 0.15 & 0.15 & 0.2 & 0.25 & 0.05 & 0.1 \\ 0.1 & 0.15 & 0.15 & 0.2 & 0.25 & 0.05 & 0.1 \\ 0.1 & 0.15 & 0.15 & 0.2 & 0.25 & 0.05 & 0.1 \\ 0.1 & 0.15 & 0.15 & 0.2 & 0.25 & 0.05 & 0.1 \\ 0.1 & 0.15 & 0.15 & 0.2 & 0.25 & 0.05 & 0.1 \\ 0.1 & 0.15 & 0.15 & 0.2 & 0.25 & 0.05 & 0.1 \end{pmatrix}$$

and let $v(t) = 0.1$, then:

$$\mathbf{R} = \begin{pmatrix} 200 & 300 & 300 & 400 & 500 & 100 & 200 \\ 300 & 450 & 450 & 600 & 750 & 150 & 300 \\ 250 & 375 & 375 & 500 & 625 & 125 & 250 \\ 400 & 600 & 600 & 800 & 1000 & 200 & 400 \\ 500 & 750 & 750 & 1000 & 1250 & 250 & 500 \\ 350 & 525 & 525 & 700 & 875 & 175 & 350 \\ 450 & 675 & 675 & 900 & 1125 & 225 & 450 \end{pmatrix}$$

so the number of vehicles traveling on highway section (6,7) during this hour is:

$$c_{6,7} = 250 + 400 + 400 + 350 = 1400$$

If the average speed is 65 miles per hour, then the density $\rho(6,7)$ is:

$$\rho_{6,7} = \frac{c_{6,7}}{s \cdot l_{6,7}} = \frac{1400 \text{ vehicles/hr}}{65 \frac{\text{miles}}{\text{hr}} \cdot 3 \text{ lanes}} \approx 7.179 \text{ vehicles/(mile} \cdot \text{lane)}$$

Comparing this result to actual traffic densities would allow prediction of heavy traffic. Notice that the only information needed besides the geometry of the road

system was number of vehicles around an exit, probability of destination, average speed, and relative traffic volume for the entire road system during the time interval. It is worth noting that individual traffic volume measurements on each segment are not needed in this model.

If we are interested in what traffic density is like at the next time interval, the formula for $\mathbf{q}_{t+1}$ can be used:

$$\begin{aligned}\mathbf{q}_{t+1} &= v(t)\mathbf{q}_t \cdot \mathbf{P} + [1 - v(t)]\mathbf{q}_t \\ &= (2045, 3067, 2617.5, 4090, 5112.5, 3272.5, 4295)\end{aligned}$$

This vector can be used to find the densities during the next time interval.

IV. Application of Model to the Greater Philadelphia Region

The highway system around the Philadelphia region is commonly plagued by heavy traffic that can be difficult to anticipate. The model developed in this paper will now be applied to the Philadelphia highway system. In order for the model to be applied, certain assumptions about the region must be made. The level of validity of these assumptions will affect the predictive value of the model, and are not absolute.

Assumptions in the Model

Instead of looking at all of the hundreds of highway exits in the region, this analysis only looks at 46 strategically selected exits. The implicit assumption here is that only a certain number of exits need to be accounted for in order to get a reasonably accurate prediction, so long as they are chosen in such a way that the exits adequately cover the entire region being modeled.

To determine the initial values for the population vector **q**, an assumption is made that the number of households (US Census Bureau, 2013) in an area is proportional to the number of vehicles in this area during the early morning hours (approximately 1AM). The number of households within a 1 and 2 mile radius of each exit was obtained from census data. This data is used as the population at time = 1AM. The rest of the populations are a result of the iterative process in the model, which evolves dynamically.

The destination probabilities were derived from data about the number of workers in a given radius of each exit (US Census Bureau, 2013). During the hours of 5AM to 10AM, the proportion of workers in the area of an exit matches the probability that a vehicle will travel to this exit. This produces a transition matrix in which each column contains elements with the same probability. During the hours of 3PM to 4AM, the values of the transition matrix switches to being

proportional to the number of households around an exit, which is what we assumed the initial population was. This should return the number of vehicles around an exit to the initial population at the end of each day. During the hours of 11AM to 2PM, the transition matrix is assumed to be a combination of the other two transition matrices. So the average of these probabilities is used during the mid-day hours. The values of the $v(t)$ function representing the relative volume of traffic using the network during time interval t was determined from past hourly traffic volumes (PennDOT, 2011).

The average speed of vehicles on the highway is assumed to be 65 miles per hour. There are two problems with this assumption. First, every section of highway may not have an average speed of exactly 65 miles per hour. Second, during heavy traffic, the average speed of vehicles will decrease. The first consideration will not drastically affect the model, since dividing by a slightly different value, say 70 miles per hour, will only reduce the density by a small percentage (7.1%). The second consideration has a greater possibility of affecting the results of the density. Heavy traffic will decrease the average speed and density will rise. Since detecting heavy traffic is the goal of this model, densities above the heavy traffic threshold will already be accounted for and an increase in density value will not change the categorical classification as heavy traffic.

When applying the model to a real network, there will be highways that lead away from the network and do not connect to another exit. These lengths of highway should be considered when applying the model, since incoming traffic will affect densities. In order to account for these highways, exits were added at the end of these highway segments and given a starting population = 9999 and number of workers = 10000. These false exits provide a buffer to the road network to more realistically replicate traffic conditions at the edge of the network. Also, these false exits provide a way to represent traffic coming into or leaving the highway network at the edges.

Data Analysis

Four versions of this model were executed and analyzed. These versions are a result of manipulating the radius around exits in which population and number of workers were considered. To evaluate the models' abilities to predict heavy traffic, the results of each model were compared to historical data (Google Maps, 2013). The Highway Capacity Manual for 2010 describes levels of services, which relate the traffic densities to relative heaviness of traffic flow (Transportation Reseach Board, 2011). The number of lanes on each highway segment was determined in order to calculate vehicle density (ITO Map, 2013). For this paper, levels of service of C (18 pc/mi/ln) or greater were considered heavy for the models based on populations within 1 mile of exits and levels of service D (26 pc/mi/ln) or

greater were considered heavy for models based on populations within 2 miles of exits.

The positive predictive value (PPV) was determined for each hour between 7AM and 7PM for each weekday. Positive predictive value is obtained through application of Bayes' theorem. The goal is to determine the conditional probability that there will be heavy traffic, given that the model predicts heavy traffic. Let H be the event that there is heavy traffic on a highway segment, then:

$$\begin{aligned} \mathrm{PPV} &= \mathrm{P}(\mathrm{H} \mid \text{Predict H}) \\ &= \frac{\mathrm{P}(\text{Predict H} \mid \mathrm{H}) \cdot \mathrm{P}(\mathrm{H})}{\mathrm{P}(\text{Predict H} \mid \mathrm{H}) \cdot \mathrm{P}(\mathrm{H}) + \mathrm{P}(\text{Predict H} \mid \text{Not H}) \cdot \mathrm{P}(\text{Not H})} \end{aligned}$$

The positive predictive value is a measure of the true positive rate of heavy traffic compared to the total number of predicted heavy traffic areas. This value allows the models to be compared to each other and shows how likely their predictions are to be correct.

The models will be referred to by the following for the rest of this paper:

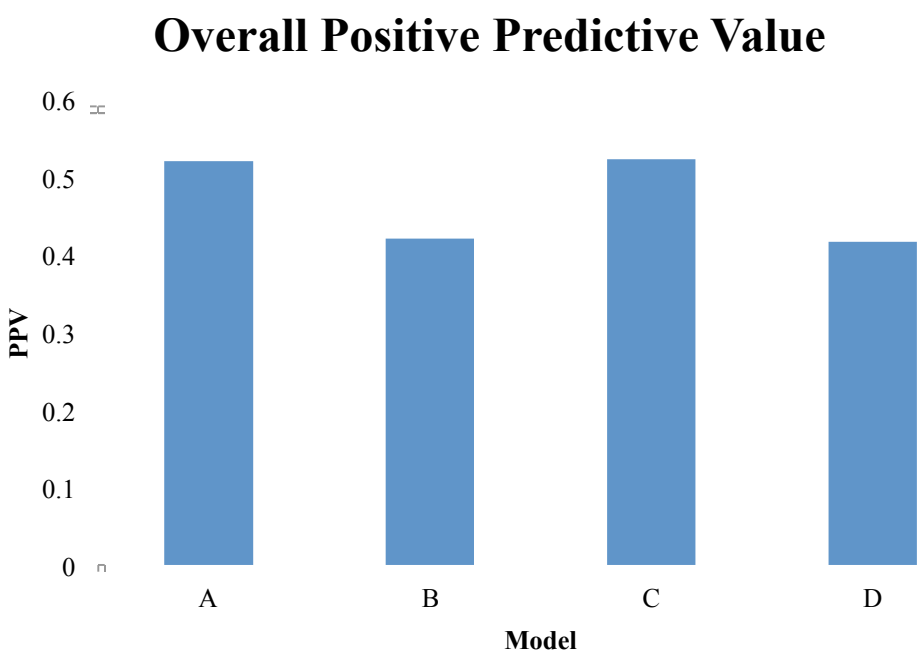
	1 miles population radius	2 mile population radius
1 mile workers radius	Model A	Model B
2 mile workers radius	Model C	Model D

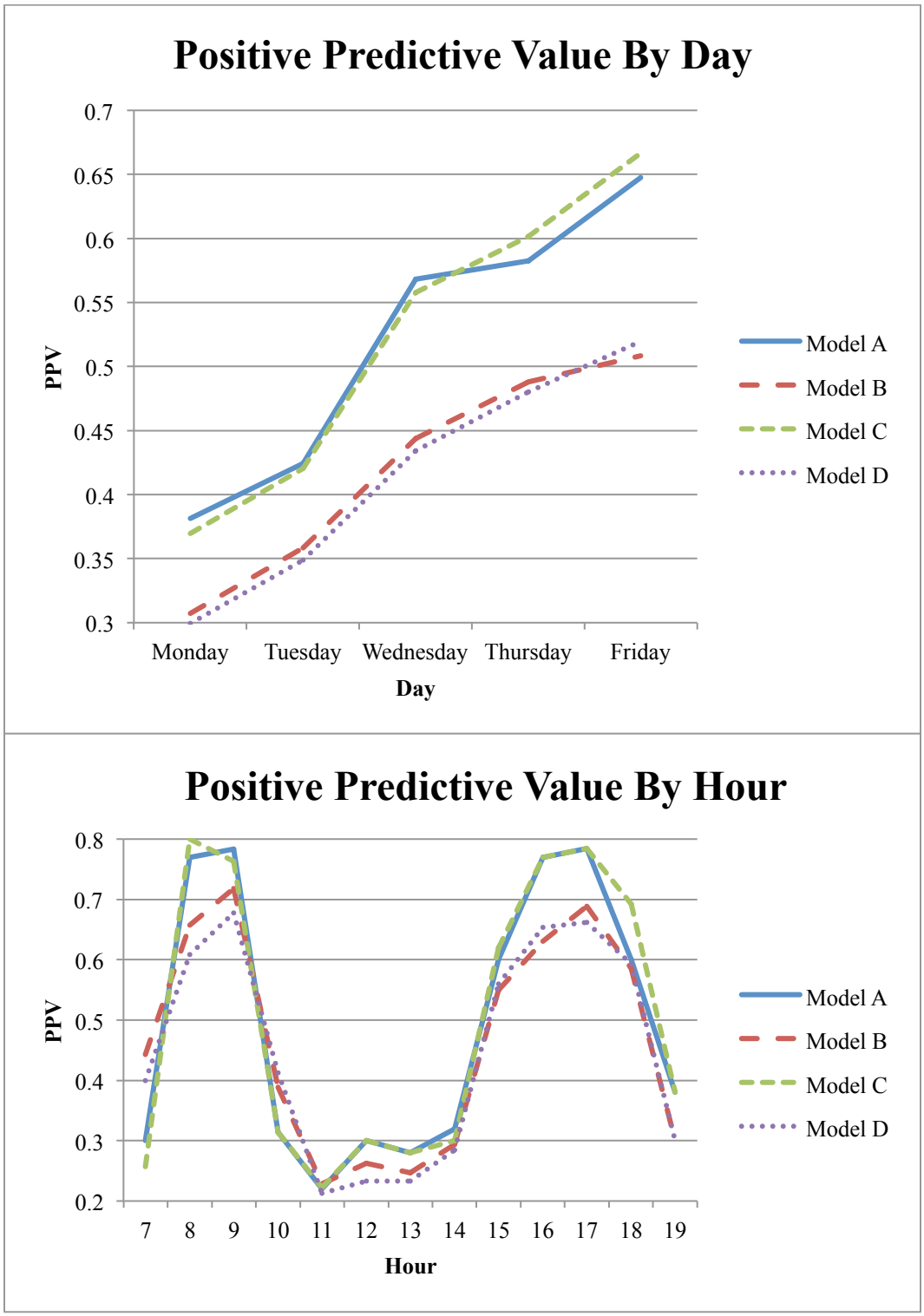
Positive Predictive Value of Models

(pop, workers)	Model A (1,1)	Model B (2,1)	Model C (1,2)	Model D (2,2)
Overall	0.52086	0.42116	0.52319	0.41645
Monday	0.38129	0.30717	0.36957	0.29934
Tuesday	0.42446	0.35836	0.42029	0.34868
Wednesday	0.56835	0.44369	0.55797	0.43421
Thursday	0.58273	0.48805	0.60145	0.64748
Friday	0.64748	0.50853	0.66667	0.51974
7AM	0.3	0.44286	0.25714	0.4
8AM	0.76923	0.65714	0.8	0.60869
9AM	0.78333	0.71765	0.76364	0.67777
10AM	0.31429	0.38824	0.31429	0.4125

11AM	0.22	0.22857	0.22222	0.2125
12PM	0.3	0.2625	0.3	0.23333
1PM	0.28	0.24706	0.28	0.23333
2PM	0.32	0.29412	0.3	0.28421
3PM	0.6	0.55	0.62	0.56
4PM	0.76923	0.63077	0.76923	0.65385
5PM	0.78462	0.68889	0.78462	0.66207
6PM	0.6	0.58519	0.69231	0.59286
7PM	0.52086	0.42116	0.52319	0.41645

Graphs of Positive Predictive Value





A comparison of the models indicates a greater sensitivity to the population radius than the workers radius. This is most clearly observed when evaluating the day-to-day change in PPV. In general, the predictive value was higher during times in which the relative traffic volumes were higher. Models A and C performed well,

with overall PPVs of 0.52086 and 0.52319 respectively. Both models considered populations within 1 mile. These models mostly performed better during times of heavier traffic, whereas all models performed about the same during non-peak hours.

Conclusion

The model presented here can be applied to any urban highway traffic network provided that sufficient demographic data is available for the region being studied. This is an economic advantage over other modeling methods that need to collect and update traffic data to construct an origin-destination matrix. An evaluation of these other models' predictive value may give insight into whether or not there is an advantage to using them.

Through the use of technology, vehicles are beginning to use an adaptive approach to route selection, which looks at current traffic conditions along with predicted conditions. Vehicles can now automatically be notified of changing traffic while *en route* and adjust accordingly (Dragoi & Dobre, 2011). As traffic control technologies continue to advance, our ability to avoid traffic congestion will improve and possibly change our disposition towards navigating on urban highway systems.

Bibliography

Crisostomi, E., Kirkland, S., Schlote, A., & Shorten, R. (2010). A Google-like Model of Road Network Dynamics and its Application to Regulation and Control. *International Journal of Control , 84* (3), 633-651.

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische mathematik , 1* (1), 269-271.

Dragoi, V.-A., & Dobre, C. (2011). A model for traffic control in urban environments. *Wireless Communications and Mobile Computing Conference (IWCMC), 2011 7th International*, (pp. 2139-2144).

Gazis, D. C., Herman, R., & Rothery, R. W. (1961). Nonlinear follow-the-leader models of traffic flow. *Operations Research , 9* (4), 545-567.

Google Maps. (2013, October 20). *Philadelphia, PA Historic Traffic Data Map for Monday.* Retrieved October 20, 2013, from Google Maps: https://maps.google.com/maps

Hoogendoorn, S. P., & Bovy, P. H. (2001). State-of-the-art of vehicular traffic flow modelling. *Proceedings of the Institution of Mechanical Engineers, Part I , 215* (4), 283-303.

Indrei, E. (2006). Markov Chains and Traffic Analysis. *Department of Mathematics, Georgia Institute of Technology* .

ITO Map. (2013, October 1). *Highway Lanes*. Retrieved October 1, 2013, from ITO Map: http://www.itoworld.com/map/179

Lighthill, M. J., & Whitham, G. B. (1955). On kinematic waves. II. A theory of traffic flow on long crowded roads. *Proceedings of the Royal Society of London, Series A. Mathematical and Physical Sciences , 229* (1178), 317-345.

Lim, S., Balakrishnan, H., Gifford, D., Madden, S., & Rus, D. (2011). Stochastic Motion Planning and Applications to Traffic. *International Journal of Robotic Research* .

Nagal, K., & Schreckenberg, M. (1992). A cellular automation model for freeway traffic. *Journal de Physique , 2* (12), 2221-2229.

PennDOT. (2011). *Factoring Process, Hourly Percent Total Vehicles.* Bureau of Planning and Research. PennDOT.

Perrakis, K., Karlis, D., Cools, M., Janssens, D., & Wets, G. (2012). A Bayesian approach for modeling origin-destination matrices. *Transportation Research Part A: Policy and Practice , 46* (1), 200-212.

Prigogine, I., & Andrews, F. C. (1960). A Boltzmann-like approach for traffic flow. *Operations Research , 8* (6), 789-797.

Rephann, T., & Isserman, A. (1994). New highways as economic development tools: An evaluation using quasi-experimental matching methods. *Regional Science and Urban Economics , 24* (6), 723-751.

Sasaki, T., & Myojin, S. (1968). Theory of inflow control on an urban expressway system. *Proceedings of the Japan Society of Civil Engineers , 160*.

Transportation Reseach Board. (2011, March-April). *Summary minifeature on the HCM2010.* Retrieved October 2013, from Transportation Research Board: onlinepubs.trb.org/onlinepubs/trnews/trnews273HCM2010.pdf

US Census Bureau. (2013, October 20). *Demographic Snapshot Summary, Total Households.* Retrieved October 20, 2013, from Free Demographics: http://www.freedemographics.com

US Census Bureau. (2013, Oct 20). *Longitudinal-Employer Household Dynamics Program.* Retrieved Oct 20, 2013, from OnTheMap Application: http://onthemap.ces.census.gov/

Van Zuylen, H. J., & Willumsen, L. G. (1980). The most likely trip matrix estimated from traffic counts. *Transportation Research Part B: Methodological , 14* (3), 281-293.

Velasco, R., & Saayedra, P. (2008). Macroscopic Models in Traffic Flow. *Qualitative Theory of Dyanamical Systerms , 7* (1), 237-252.

Youngblom, E. (2013). Travel Time in Macroscopic Traffic Models for Origin-Destination Estimation. *University of Wisconsin-Milwaukee* .

Zhang, G., Wang, Y., Wei, H., & Chen, Y. (2007). Examining headway distribution models with urban freeway loop event data. *Transportation Research Record: Journal of the Transportation Research Board , 1999* (1), 141-149.