



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

Transportation Research Part C

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

Maintenance optimization for transportation systems with demand responsiveness

Pablo L. Durango-Cohen^{a,*}, Pattharin Sarutipand^{b,1}

^a Department of Civil and Environmental Engineering, Northwestern University, 2145 Sheridan Road, A332, Evanston, IL 60208-3109, United States

^b 5, Bureau of Highway Maintenance Management, 2/486 Si Ayutthaya Rd., Thung Phaya Thai, Ratchathewi, Bangkok 10400, Thailand

ARTICLE INFO

Article history:

Received 7 May 2008

Received in revised form 31 December 2008

Accepted 5 January 2009

Keywords:

Transportation systems
Infrastructure management
Maintenance optimization
Demand responsiveness
Quadratic programming
ARMAX models

ABSTRACT

We present a quadratic programming framework to address the problem of finding optimal maintenance policies for multifacility transportation systems. The proposed model provides a computationally-appealing framework to support decision making, while accounting for functional interdependencies that link the facilities that comprise these systems. In particular, the formulation explicitly captures the bidirectional relationship between demand and deterioration. That is, the state of a facility, i.e., its condition or capacity, impacts the demand/traffic; while simultaneously, demand determines a facility's deterioration rate. The elements that comprise transportation systems are linked because the state of a facility can impact demand at other facilities. We provide a series of numerical examples to illustrate the advantages of the proposed framework. Specifically, we analyze simple network topologies and traffic patterns where it is optimal to coordinate (synchronize or alternate) interventions for clusters of facilities in transportation systems.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

Transportation infrastructure management refers to the process of making decisions concerning the allocation of resources for maintenance of the facilities that comprise transportation systems. The objective in making these decisions is to ensure that the systems are capable of performing the functions for which they were designed and built, while accounting for limited resource availability or level-of-service requirements. In this context, management decisions trade off user costs, which are associated with travel time, fuel consumption, vehicle depreciation and maintenance, and agency costs that are related to the type and intensity of maintenance activities, and include expenses for resource and personnel delivery, materials, etc. As facilities deteriorate, the rate at which user costs accrue increases. In turn, agency costs are incurred to improve condition, and thus, reverse the effects of deterioration.

The primary motivation for the work presented herein is that existing maintenance optimization models sacrifice important functional characteristics in favor of computational tractability/simplicity. In particular, models for multifacility transportation systems are usually formulated as constrained Markov Decision Processes (MDPs) using the notion of “randomized policies” and are solved as linear programs (Golabi et al., 1982; Golabi and Shepard, 1997; Murakami and Turnquist, 1997; Smilowitz and Madanat, 2000). In this framework, individual facilities are not identified and are assumed to be homogeneous. As a result, optimal policies specify the same set (probability distribution) of maintenance actions for all facilities that are in a given state, i.e., they are classified by their condition. Linear constraints can be included in these models to impose

* Corresponding author. Tel.: +847 491 4008; fax: +847 491 4011.

E-mail addresses: pdcc@northwestern.edu (P.L. Durango-Cohen), pattharinsarutipand2008@u.northwestern.edu (P. Sarutipand).

¹ Tel.: +66 2 354 6668x2906; fax: +66 2 644 5500.

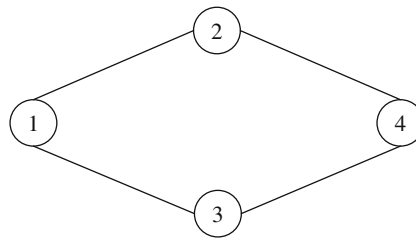


Fig. 1. Hypothetical transportation network.

restrictions that apply to the system, e.g., resource availability. These models are attractive because the computational effort to find optimal policies is independent of the number of facilities that comprise the system. However, omitting information that identifies the individual facilities makes it impossible to capture functional relationships that link them. This, in turn, constitutes an obstacle to providing effective decision support because significant benefits and costs in the management process can be directly attributed to interdependencies that link a system's facilities. To further emphasize this point consider the following example:

Fig. 1 represents a hypothetical transportation network with two paths between node 1 (origin) and node 4 (destination). We assume that the four links in the system are homogeneous with respect to their deterioration.

The fact that existing models classify facilities by their condition means that if links (1,2) and (1,3) are in the same condition, an optimal policy would specify the same action for both of them. Also, the homogeneous cost structure means that it costs exactly twice as much to apply an action to both links as it does to apply the action to either link individually. In practice, however, one expects maintenance policies to specify that (major) interventions on links (1,2) and (1,3) are performed at different times in order to minimize the disruption to the system. Such policies would account for a (user) cost structure in which disrupting the two links simultaneously costs much more than twice the cost of disrupting each link individually. Links (1,2) and (1,3) are said to exhibit a substitutable relationship. For analogous reasons, one also expects that interventions on links (1,2) and (2,4), or on links (1,3) and (3,4), would be synchronized. The links in each pair are said to exhibit a complementary relationship.

Recent advances in optimization theory, coupled with ever increasing availability of powerful computing platforms, make it possible to strike a more appealing and useful balance of capturing realism and ensuring tractability. In particular, as a step to address the limitations described in the preceding paragraphs, we formulate the problem of obtaining optimal maintenance policies for multifacility transportation systems as a quadratic program (QP). In the formulation, each facility's deterioration and demand/traffic are identified and represented as a linear system, i.e., an autoregressive moving average model with exogenous inputs (ARMAX) model. The model explicitly captures the bidirectional relationship between demand and deterioration. That is, the state of a facility, i.e., its condition or capacity, impacts the demand/traffic. The elements that comprise the system are linked because the state of a facility can impact the demand at other facilities. These relationships are captured by the specification of the cross-elasticities of demand. Simultaneously, demand/traffic determines a facility's deterioration rate. The quadratic objective can be used to capture nonlinearities in cost terms, which may, for example, reflect costs associated with congestion, vehicle wear and tear, and scale of maintenance activities. In addition to presenting the framework, we provide a series of numerical examples to illustrate its advantages. The numerical results illustrate how network topology and traffic patterns are key determinants of the structure of optimal maintenance policies.

The remainder of the paper is organized as follows: Section 2 provides an overview of maintenance optimization models for multifacility transportation systems. The proposed model is presented in Section 3. In Section 4, we provide numerical examples to illustrate the appealing features of the proposed model. Conclusions and directions for future work are discussed in Section 5.

2. Related work

The wide use and acceptance of finite (state and action) MDP formulations for periodic review maintenance optimization problems means that perhaps the most natural way to develop a framework for systems of interdependent facilities is to consider multidimensional MDP formulations for the problem. In the management science literature, these types of models are referred to as Parallel Machine/Equipment Replacement Problems (Vander Veen, 1985; Jones et al., 1991; McClurg and Chand, 2002; Chen, 1998; Childress and Durango-Cohen, 2005). These formulations provide an attractive framework to model interdependencies because the state and action space, as well as the deterioration process of each facility in the system is fully identified. In addition, there is a great deal of flexibility in terms of specifying M&R costs. Unfortunately, the fact that the state and action spaces are discrete, leads to significant computational problems in solving and analyzing such models. These difficulties are well-known and referred to as the "curse of dimensionality". The cause of these problems is that solution approaches for MDPs require enumerating all possible interventions for every system state and for every decision-making stage. Because the state and action spaces are discrete, the total numbers of possible system states and interventions

increase exponentially with the number facilities in the system. This consideration makes the approach impractical, even for transportation systems with small numbers of facilities.²

The computational difficulties associated with multidimensional MDP formulations have motivated the development of other approaches to address maintenance optimization problems for systems of interdependent facilities. Comprehensive reviews are presented in [Thomas \(1986\)](#), [Cho and Parlar \(1991\)](#) and [Dekker et al. \(1997\)](#). They classify the models based on the possible interdependencies between the facilities that comprise a system, which can be structural/functional, stochastic or economic. Functional dependence refers to situations where the successful operation of a system requires the successful operation of a minimum number of different types of components, i.e., the reliability of the system depends on the reliability of its components. In the context of maintenance optimization, system (or component) failures present opportunities to inspect, maintain, repair or replace both failed and functional components. Stochastic dependence refers to situations where the time-to-failure distributions of different components are dependent (either due to interactions or because they are influenced by similar factors, e.g., environment or loading). This means that knowledge about the condition (or failure) of a subset of the components provides an opportunity to update the time-to-failure distributions and the maintenance policy for other components. Finally, economic dependence refers to situations where the components of a system are linked by resource constraints or level-of-service requirements, or by the system-wide cost-benefit structure, i.e., the total cost of applying actions on a set of components can be different than the sum of the individual component costs. As an example of economically dependent components, consider a system in which the maintenance of each component requires preparatory or set-up work that can be shared when several components are maintained simultaneously.

The facilities that comprise transportation systems, simultaneously exhibit the three types of interdependencies described above. These interdependencies derive from the spatial distribution of facilities within the system, i.e., the network's topology, and from the trip distribution and the ensuing traffic flow. These relationships, in turn, have economic implications, and thus, our objective is to develop system-level maintenance policies for multifacility transportation systems. In addition to the review papers cited above, [Durango-Cohen and Sarutipand \(2007\)](#) and [Sarutipand \(2008\)](#) provide comprehensive reviews of applications in the management of transportation systems. Generally, models in the literature have focused on capturing economic dependencies, such as the ones described in the previous paragraph or such as resource-availability constraints or level-of-service requirements.

In this paper, we propose a maintenance optimization model to capture functional relationships, such as the complementary and substitutable relationships described in the example presented in Section 1. In the context of transportation systems, functional dependencies refer to situations where the state of a facility, and therefore its ability to perform the functions for which it was designed and built, impacts (either positively or negatively) the ensuing demand at another facility. Two important observations about functional dependencies between facilities in transportation systems are: (i) That the relationships tend to be weaker than the reliability example presented above, e.g., generally a facility in a poor/failed state does not eliminate the demand on other facilities (or on itself); and (ii) that traffic/demand, in turn, causes the state of a facility to deteriorate, i.e., the relationship between demand and facility state is bidirectional.

Two approaches have appeared in the literature to capture functional interdependencies in the context of maintenance of transportation systems. These approaches exemplify the modeling tradeoffs of realism vs. tractability. The first, a bottom-up or operational approach, where the impact of maintenance interventions on demand is evaluated through traffic simulation. [Uchida and Kagaya \(2006\)](#) and [Ouyang \(2007\)](#) are recent, and perhaps, the most representative studies employing this approach. [Uchida and Kagaya \(2006\)](#) develop a probit-based stochastic (and static) user equilibrium assignment framework. The effect of maintenance is captured in a generalized cost function that includes costs associated with delays caused by capacity reduction during interventions. In part, as a result of the details included in the traffic simulation, e.g., the duration of a given intervention, the framework does not lend itself to analytical (or direct numerical) solutions of the underlying maintenance optimization problem, and in turn, motivated the authors to develop a simulation–optimization procedure to select maintenance policies. [Ouyang \(2007\)](#), on the other hand, provides a different balance of the aforementioned trade-offs. In particular, he proposes a model where a facility's state (condition and traffic) determines link travel times and demand, which is estimated with a deterministic and static user equilibrium assignment model. In this framework, the traffic simulation model is less detailed, but can be integrated into a multidimensional dynamic programming formulation for the maintenance optimization problem. To overcome the “curse of dimensionality” the author presents an efficient method that relies on constructing an approximation of the value function using a finite number of basis functions.

In this paper, we consider a top–down or strategic approach that builds on [Friesz and Fernandez \(1979\)](#), who present a multidimensional control model that captures the relationship between demand and facility state. In particular, the effect of facility state on demand is captured by specifying the corresponding elasticity. The model presented in Section 3 can be interpreted as a multifacility generalization of [Friesz and Fernandez \(1979\)](#), where the effect of a facility's state on the demand at another facility is captured by the corresponding cross-elasticities of demand. To deal with the computational difficulties that arise in solving the ensuing maintenance optimization problem, rather than considering a generic control framework such as the one proposed by [Friesz and Fernandez \(1979\)](#), we consider a discrete-time problem, and assume that

² [Childress and Durango-Cohen \(2005\)](#), for example, show how the solution to a modest, single-stage PMRP with 15 facilities, 5 condition-states and 2 actions for each facility, requires solving a linear program with over 15 thousand variables and over 77 million constraints. This consideration is important given the size of transportation systems. [Golabi and Pereira \(2003\)](#), for example, recently developed a pavement management system for a road network with over 40 thousand sections.

the dynamics of the process are linear, and that the cost functions can be represented (or approximated) by second order polynomials. This approach is similar in spirit to the approximation method used by Ouyang (2007). Moreover, we have successfully used the framework in the analysis of closely-related maintenance optimization problems. Specifically, Durango-Cohen and Tadeipalli (2006) consider maintenance optimization for single facilities where multidimensional arrays of condition data are generated periodically. Durango-Cohen and Sarutipand (2007) consider a multifacility problem where maintenance decisions for the components are linked by a cost structure that captures either benefits associated with resource and personnel delivery when adjacent facilities are maintained simultaneously or costs associated with simultaneous disruption of substitutable links.

3. Model formulation

We consider the problem of managing, i.e., finding an optimal maintenance policy for a transportation system that consists of N facilities over a planning horizon of T periods. The variables in the model are as follows:

- y_t^n : Decision variable representing the (intensity of the) intervention that facility n receives during period t . The set of interventions for the system during t is collected in the vector Y_t , where $Y_t \equiv [y_t^1, y_t^2, \dots, y_t^N]'$.³
- x_t^n : State variable representing facility n 's condition at the start of period t . Similarly, $X_t \equiv [x_t^1, x_t^2, \dots, x_t^N]'$.
- q_t^n : State variable representing the demand/traffic on facility n during t ($Q_t \equiv [q_t^1, q_t^2, \dots, q_t^N]'$).
- u_t^n : State variable representing facility n 's effective capacity during t ($U_t \equiv [u_t^1, u_t^2, \dots, u_t^N]'$).
- ρ_t^n, η_t^n : Variables used to track facility n 's effective capacity deficit/surplus. Following a similar convention, $\bar{\rho}_t \equiv [\rho_t^1, \rho_t^2, \dots, \rho_t^N]'$ and $\bar{\eta}_t \equiv [\eta_t^1, \eta_t^2, \dots, \eta_t^N]'$.

To simplify the presentation, we define the parameters/coefficients used in the formulation in the context of describing the equations in the model. Thus, the problem of obtaining an optimal maintenance policy can be written as follows:

Minimize :

$$\sum_{t=1}^T \delta^{t-1} \Pi(X_t, Q_t, \bar{\rho}_t, Y_t) + \delta^T \Omega(X_{T+1}) \quad (1)$$

Subject to :

$$x_{t+1}^n = g_x^n x_t^n - h^n y_t^n + \lambda_q^n q_t^n + k_x^n, \quad n = 1, \dots, N, \quad t = 1, \dots, T \quad (2)$$

$$q_{t+1}^n = g_q^n q_t^n + \sum_{i=1}^N \lambda_x^{ni} x_{t+1}^i + \sum_{i=1}^N \lambda_u^{ni} u_{t+1}^i + k_q^n, \quad n = 1, \dots, N, \quad t = 1, \dots, T-1 \quad (3)$$

$$u_t^n = \kappa^n - \psi^n y_t^n, \quad n = 1, \dots, N, \quad t = 1, \dots, T \quad (4)$$

$$q_t^n - u_t^n = \rho_t^n - \eta_t^n, \quad n = 1, \dots, N, \quad t = 1, \dots, T \quad (5)$$

$$x_1^n = \hat{x}_1^n, \quad n = 1, \dots, N \quad (6)$$

$$q_1^n = \hat{q}_1^n, \quad n = 1, \dots, N \quad (7)$$

$$x_t^n, q_t^n, u_t^n, \rho_t^n, \eta_t^n \geq 0 \quad (8)$$

The objective function presented in Eq. (1) corresponds to the total discounted social cost, i.e., user plus agency costs, of managing the transportation system. We assume that the discount factor, δ , is defined over the open interval $(0, 1)$. $\Pi(\cdot)$ is the period cost function and $\Omega(\cdot)$ is the salvage value function. We assume that both $\Pi(\cdot)$ and $\Omega(\cdot)$ can be represented or approximated by second-order polynomials, and in particular consider the following specifications:

$$\Pi(X_t, Q_t, \bar{\rho}_t, Y_t) = Y_t' A Y_t + X_t' B Q_t + Q_t' C Y_t + \bar{\rho}_t' D \bar{\rho}_t \quad (9)$$

$$\Omega(X_{T+1}) = X_{T+1}' L X_{T+1} \quad (10)$$

where A, B, C, D , and L are $N \times N$ diagonal matrices. The elements of these cost functions are discussed in Section 3.1.

The constraints in the model are used to represent the dynamics of the managerial process. In particular:

- Equation set (2) describes the deterioration of the facilities that comprise the system. The coefficients g_x^n, h^n, λ_q^n , respectively capture the effect of previous condition, maintenance and traffic on the subsequent condition of facility n . k_x^n captures the effect of exogenous factors, e.g., weather, on facility n 's deterioration rate.
- Equation set (3) describes the demand/traffic on the facilities in the system. The parameter g_q^n captures the historical effect. k_q^n captures the rate of change that is based on factors other than the variables in the model. Significantly, we note that the traffic on a given link is sensitive both to its own condition and effective capacity, as well as to the condition and effective capacity of the other facilities in the system. This dependency is captured through the coefficients λ_x^{ni} and λ_u^{ni} ,

³ The notation V' is used to represent the transpose of a vector/matrix V .

which correspond to the elasticities of demand with respect to condition and effective capacity. The actual signs of these coefficients depend on the network's structure and trip distribution. In particular, for the coefficients λ_u^{ni} , positive signs apply to complementary facilities and negative signs to substitutable facilities. The signs are opposite for the coefficients λ_x^{ni} .

- Equation set (4) is used to set facility n 's effective capacity (in period t) to a nominal level, κ^n , minus an amount that is proportional to the magnitude/intensity of the intervention it receives during t . ψ^n is the coefficient that describes this relationship.
- The next set of equations, (5), are used to track the utilization of the facilities in the system. In particular, ρ_t^n and η_t^n are set to account for deficits/surpluses in effective capacity. A positive ρ_t^n indicates a capacity deficit in facility n during t , and is associated with a congestion cost captured in the period cost function $\Pi(\cdot)$.
- The equations in (6) and (7) are used to specify the system's initial condition and traffic. Facility n 's initial condition and traffic are denoted \hat{x}_1^n and \hat{q}_1^n , respectively.
- Finally, equation set (8) imposes logical restrictions on the model's variables.

Prior to continuing with the description, we discuss some of the assumptions/limitations embedded in the model. In particular, we note that:

- Continuous decision variables can be interpreted as maintenance rate or as an investment level. This interpretation is consistent with earlier models in the field (cf. (Friesz and Fernandez, 1979)) and is appropriate for tactical and strategic models, which might, for example, be used for budgeting.
- The quadratic cost structure is not overly restrictive because, for example, it is possible to obtain optimal maintenance policies for general classes of continuous cost functions by solving a sequence of problems. In each problem the cost function is approximated by a second-order Taylor Series expanded about a different point. The procedure is analogous to the Newton–Raphson method for solving systems of equations/optimization problems and is discussed further in Dreyfus (1977).
- The deterministic linear systems used to represent deterioration and demand in Eqs. (2) and (3), respectively, correspond to simple forms of Autoregressive Moving Average with exogenous inputs (ARMAX) models.⁴ These models provide a convenient, flexible and rigorous framework to formulate and estimate dynamic models. Indeed, preliminary statistical analysis in the context of formulating and estimating pavement deterioration models are encouraging (Chu and Durango-Cohen, 2007; Chu and Durango-Cohen, 2008). Traffic simulation/assignment provides a possible framework to generate the data needed to estimate the parameters, i.e., the elasticities, to specify the demand models. It is important to note, however, that the assumptions embedded in ARMAX models may not be universally valid, e.g., that the effect of interventions is linear and additive in the context of deterioration modeling, or that the relationship between traffic and capacity is linear in the context of demand modeling. Even though there is flexibility to (partially) address some of the aforementioned limitations, it is important to realize that the ARMAX framework may be inadequate in certain situations. Thus, the use of ARMAX models should be interpreted as analogous to the use of linear programming, even though most realistic problems probably do not exhibit the linear structure that is assumed.
- Our primary objective in this paper is to develop a framework that captures the effect of economic and demand interdependencies and of heterogeneity on optimal intervention policies. To simplify the presentation, we assume that facility deterioration and demand process are deterministic, i.e., k_x^n and k_q^n are deterministic. This assumption reduces the generality of the framework, but leads to an optimization problem that is both flexible enough to capture interdependencies and heterogeneity, and that is practical for systems with large numbers of facilities. To justify the latter, we note that (i) recent advances in optimization theory and in computational power allow for the solution of large-scale quadratic programs, and that (ii) the number of parameters in the model increases as a polynomial function of the number of facilities in the system (i.e.: the number of parameters is $\mathcal{O}(N^2)$). In contrast, the constrained MDP is practical but inflexible, i.e., the number of parameters and computational effort to find optimal policies are both small, but it is not possible to capture interdependencies or heterogeneity. Finally, the multidimensional MDP is flexible but impractical. While it offers the most flexibility, the number of parameters that require estimation, as well as the computational effort increase exponentially with the number of facilities that comprise the system. We also think it is appropriate to raise the following points:
 - The assumption of deterministic deterioration has been used in previous models in the literature;
 - Recent statistical studies such as Archilla and Madanat (2000); Prozzi and Madanat (2002) have shown that a large fraction of the variability in infrastructure deterioration should be attributed to heterogeneity among facilities, as opposed to aleatory uncertainty. Constrained MDPs, unnecessarily, attribute all of the variability to aleatory uncertainty. The proposed model (unnecessarily) attributes all of the variability to heterogeneity. Clearly, both are extremes and the ideal model lies somewhere in between; and

⁴ Simple extensions of these models would involve adding lagged dependent variables to capture the effect of historical trends on condition or demand. Considering non-stationary models (with time-dependent coefficients), or including additional exogenous variables are examples of other possible generalizations.

- The solution to optimization of problems whose dynamics are governed by stochastic linear systems and whose objective is quadratic, i.e., stochastic linear-quadratic regulators, is discussed in references such as Bertsekas (1995). Using this general framework for maintenance optimization would seem to be an important direction for research.

In the following subsection, we discuss the specification of the cost components in the model.

3.1. Cost function specification

We specify a period cost function, $\Pi(\cdot)$ that consists of three elements that capture costs incurred by both agencies or users. These elements are:

Maintenance costs: These costs correspond to the sum of costs associated with maintenance of the facilities in the system. In period t , these costs can be written as⁵:

$$\sum_{n=1}^N (a_{nn}(y_t^n)^2 + c_{nn}q_t^n y_t^n) \quad (11)$$

The first term corresponds to the direct maintenance costs. The second term corresponds to the traffic disruption costs associated with the set of interventions across the system. We assume $a_{nn} > 0, \forall n$, which means that the costs of an intervention increase (at an increasing rate) with the intensity of the maintenance rate, y_t^n . We also assume $c_{nn} > 0, \forall n$, meaning that the disruption costs incurred by the users is increasing with the maintenance rate and the traffic.

Vehicle operating costs: These are the costs incurred by the users that are associated with facility condition, e.g., vehicle wear-and-tear. Assuming $b_{nn} > 0, \forall n$ means that the costs increase as facility condition worsens and are proportional to traffic. These costs can be written as:

$$\sum_{n=1}^N b_{nn}x_t^n q_t^n \quad (12)$$

Congestion costs: These terms capture costs associated with travel time and other consequences associated with deficits in the effective capacity of the facilities that comprise the system. These costs can be written as:

$$\sum_{n=1}^N d_{nn}(\rho_t^n)^2 \quad (13)$$

Assuming $d_{nn} > 0, \forall n$ means that the congestion costs increase (at an increasing rate) with the magnitude of the capacity deficit.

In addition, we consider a salvage value function that only depends on the terminal condition of the system. The salvage value function can be written as follows:

$$\sum_{n=1}^N l_{nn}(x_t^n)^2 \quad (14)$$

4. Numerical examples

In this section, we present examples to illustrate maintenance policies obtained with the proposed model. Our objectives are to show how demand interdependencies influence optimal policies, and to show that the proposed model has appealing features that elude existing maintenance optimization models for transportation systems. We begin this section by considering simple transportation systems comprised of two links and one origin-destination pair.⁶ The systems: substitutable and complementary are presented in Figs. 2a and b, respectively. Later in the section, we consider more general, examples to show how the framework can be extended, as well as to discuss difficulties.

We solved the QPs to obtain the numerical results with a robust nonlinear optimization solver, KNITRO. This solver is publicly available online through the NEOS servers located at Northwestern University and the Argonne National Laboratory. Further information is available through <http://www-neos.mcs.anl.gov/neos/solvers/nc:KNITRO/AMPL.html>.

To highlight the impact of demand interdependencies on the structure of optimal maintenance policies, we consider facilities that are homogeneous in their capacity, deterioration and demand processes. We also examine how interdependencies impact the other state variables: demand, condition and capacity, when the optimal maintenance policies are implemented.

⁵ The notation a_{ij} is used to represent the $i - j$ element of matrix A .

⁶ The reader interested in computational issues associated with solving (large-scale) quadratic programs is referred to Gil et al. (1981).

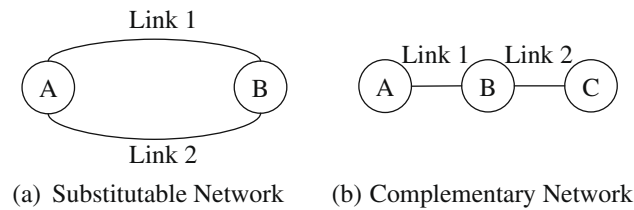


Fig. 2. Basic network types.

Intuitively, one expects different optimal maintenance policies for different network topologies and trip distributions. In the following section, we first compare optimal policies for the networks presented in Fig. 2. We also conduct a sensitivity analysis to develop qualitative insights about how parametric and topological changes affect the optimal policies, system condition, and traffic over the networks. The parameters used in the study are not representative of any particular facility, although the situation considered is “inspired” by a pavement management situation. To simplify the interpretation of the numerical results we consider a problem over a 25-period planning horizon, which in conjunction with other suitable choices leads to optimal policies that converge to a *steady-state*, i.e., the policy prescribes periodic interventions for the links in the system. Such policies provide a convenient way to contrast the results for the different cases. The discount factor δ , is set to $1/1.05$, which corresponds to a discount rate of 5%. The parameters used to specify instances of the model are presented in Table 1.

From the above specification we note that:

- Together (4) and (8) imply an upper bound on the decision variables, y_t^n , of 100.
- We set the initial condition of every facility in the system to new condition. The demand in each facility is set to 50% of the physical capacity, 100.
- Temporal changes in condition, $x_t^n - x_{t-1}^n$, are assumed to be independent of the history of the deterioration process. Also, the constant deterioration rate of 10 is such that, absent loading and maintenance, it takes 10 periods to deteriorate from the best condition state, 0, to the worst, 100. This is intended to mimic the PCI scale, which has a range of 100.
- For simplicity, in the numerical examples presented below, we set $\lambda_u^{ni} \neq 0$ and $\lambda_x^{ni} = 0, n, i = 1, 2$. Thus, we are assuming demand is sensitive to effective capacity and insensitive to condition.

4.1. Basic networks

In this section we use the model to obtain maintenance policies for the two networks shown in Fig. 2. In particular, we set the elasticities in the model to capture the two types of functional relationships that are illustrated. We begin by considering the case of complementary facilities. The results for the complementary network are presented in Fig. 3. We observe that the optimal steady-state policy specifies identical interventions for both links, i.e., an optimal maintenance rate of 46 units/period for each of the facilities. As a result, the state variables in the model (traffic, condition and capacity) converge to identical levels for the two facilities. We observe that the structure of the optimal maintenance policy, i.e., the fact that both facilities receive maintenance simultaneously, applies for $c_{nn} \geq 0, n = 1, 2$. However, the magnitude of the maintenance and of the other variables changes. The results in Fig. 3 are for $c_{nn} = 5, n = 1, 2$.

To illustrate how a network's topology influences the structure of optimal maintenance policies, we apply the model to the substitutable network presented in Fig. 2b. We set the elasticities as follows: $\lambda_u^{nn} = 0.2, n = 1, 2$ and $\lambda_u^{ni} = -0.2, n \neq i = 1, 2$.

Table 1
Parameter specification for numerical examples.

Cost parameters		Constraint parameters	
$a_{nn}, \forall n$	1	$g_x^n, \forall n$	1
$b_{nn}, \forall n$	1	$h^n, \forall n$	1
$c_{nn}, \forall n$	Variable	$\lambda_q^n, \forall n$	0.5
$d_{nn}, \forall n$	1	$k_x^n, \forall n$	10
$l_{nn}, \forall n$	0	$g_q^n, \forall n$	0
		$\lambda_x^{ni}, \forall n, i$	Variable
		$\lambda_u^{ni}, \forall n, i$	Variable
		$k_q^n, \forall n$	50
		$\kappa^n, \forall n$	100
		$\psi^n, \forall n$	1
		$\hat{x}_1^n, \forall n$	0
		$\hat{q}_1^n, \forall n$	50

Negative cross-elasticities are used to capture the substitutable relationship between the links. In these cases, the results are such that when the traffic disruption impact is “low”, $c_{nn} \leq 4$, $n = 1, 2$, the optimal policy is similar in structure to the policy for the complementary network, i.e., the facilities are maintained simultaneously. On the other hand, when the impact is “high”, $c_{nn} > 4$, the optimal policy specifies alternating interventions between the two facilities. As an example and to con-

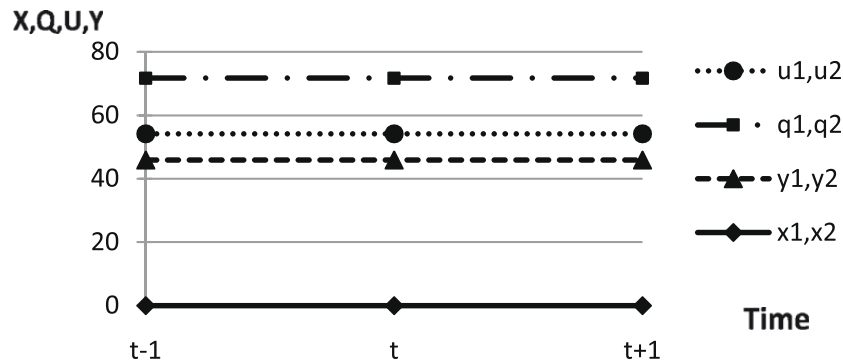


Fig. 3. Optimal solution of the complementary network.

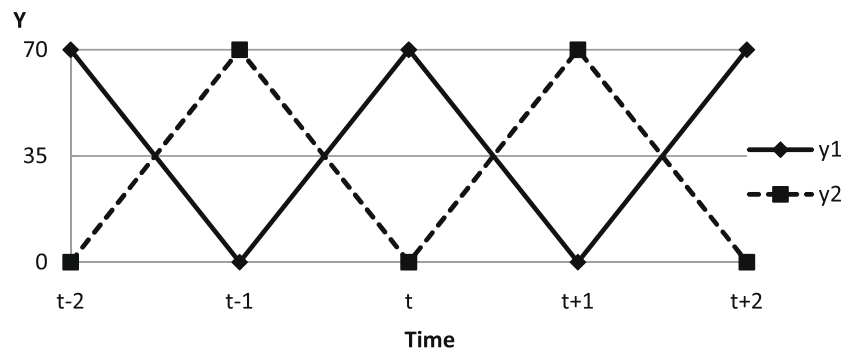


Fig. 4. Optimal policies for substitutable network with high disruption impact.

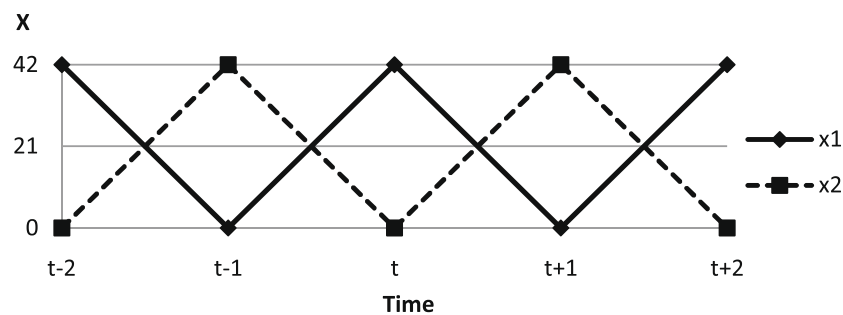


Fig. 5. Optimal condition for substitutable network with high disruption impact.

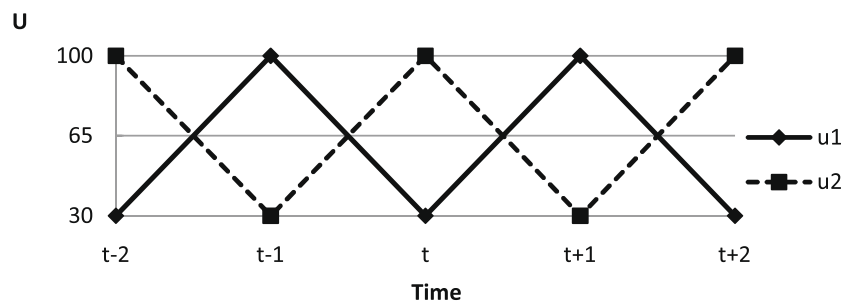


Fig. 6. Optimal effective capacities for substitutable network with high disruption impact.

trast with the above results for the complementary network, we set $c_{nn} = 5$, $n = 1, 2$ and present the results in Fig. 4. We observe that when the disruption impact is high, applying simultaneous interventions causes significant traffic delay due to disruption/loss of throughput and may also lead to congestion when the amount of traffic during the maintenance period is over the effective capacity of the facilities. As a result, the optimal policy specifies coordinated interventions where the facilities are maintained in alternating periods. The ensuing intervention schedule results in cycles in the levels of the state variables in the model, as shown in Figs. 5–7.

Another appealing feature of the optimal maintenance policies is that they lead to coordination between the variables in the model. In particular, the optimal policy leads to situations where high traffic is observed on facilities in good condition where there are no interventions. This, in turn, simultaneously reduces user costs (associated with vehicle wear-and-tear), as well as traffic disruption costs.

Next, we consider the effect of changing the demand's sensitivity to effective capacity on the results. In particular, we set $c_{nn} = 5$ and vary the elasticities, λ_u^{nn} , $\forall n$ and $-\lambda_u^{ni}$, $\forall n, i$, from 0 to 0.5 in increments of 0.05. The results are presented in Figs. 8 and 9.

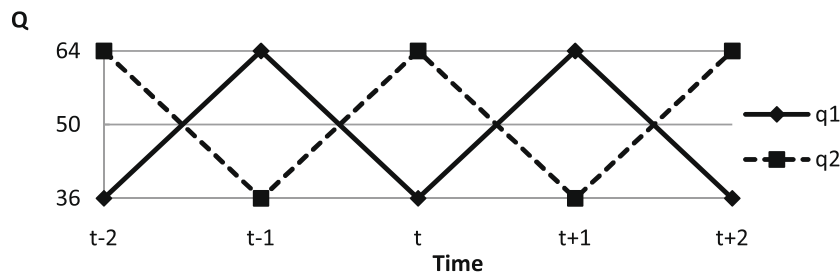


Fig. 7. Optimal amount of traffic for substitutable network with high disruption impact.

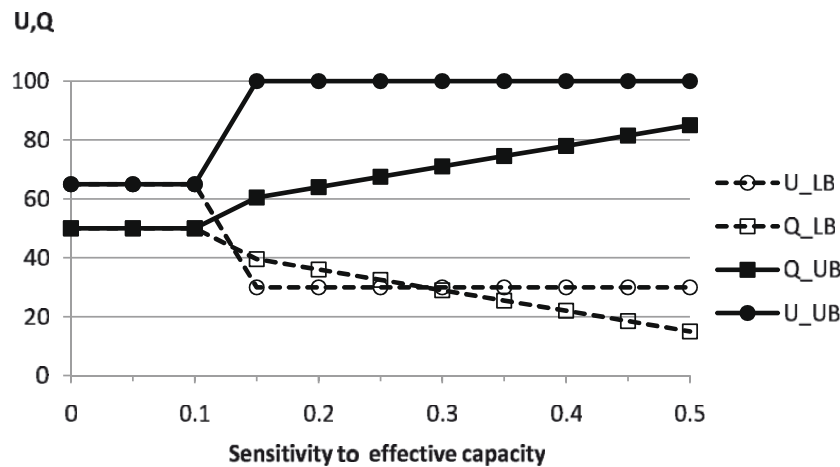


Fig. 8. System traffic condition with variation of sensitivity level to congestion at $c_{nn} = 5$.

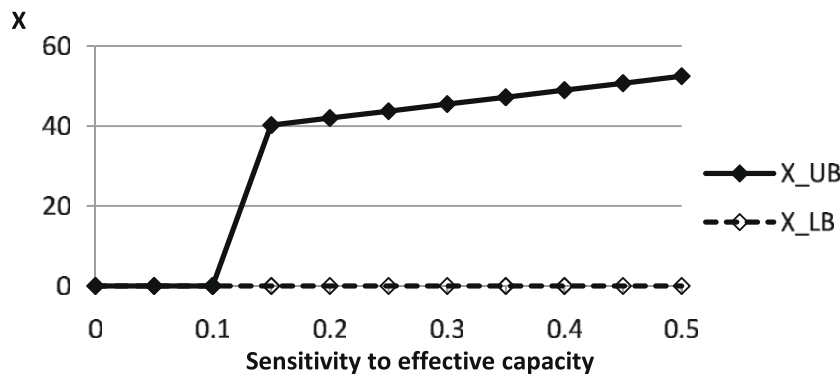


Fig. 9. The condition with variation of sensitivity level to congestion at $c_{nn} = 5$.

Both figures display two regimes: For elasticities with magnitudes below 0.15, optimal policies specify a maintenance rate of 35 units/period for both facilities. Such interventions restore the condition of each of the facilities in each period, and result in associated state variables that are maintained at the same level for both facilities. For elasticities above 0.15, the structure of optimal maintenance policies specifies cyclical interventions for the two facilities. In this regime, facility condition is restored in alternating periods. The magnitude of the intervention is 70. Figs. 8 and 9 show the magnitudes of the associated state variables. In consecutive periods and for a given elasticity, the facilities alternate between the upper and lower bounds that are presented in the figures. We observe that as the elasticity increases, meaning that additional traffic oscillates between the two facilities, the worse condition level, X_{UB} , increases. Because the average per-period demand that each facility serves is 50, explains why the magnitude of the interventions is 70. From Fig. 8 we observe that for the given for the nominal capacities and optimal interventions, congestion is induced (and an associated cost is incurred) in cases where the elasticity is low, reflecting users' lack of interest to consider an alternative route.

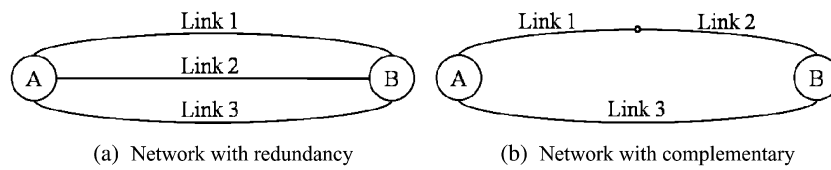


Fig. 10. Network with redundant links.

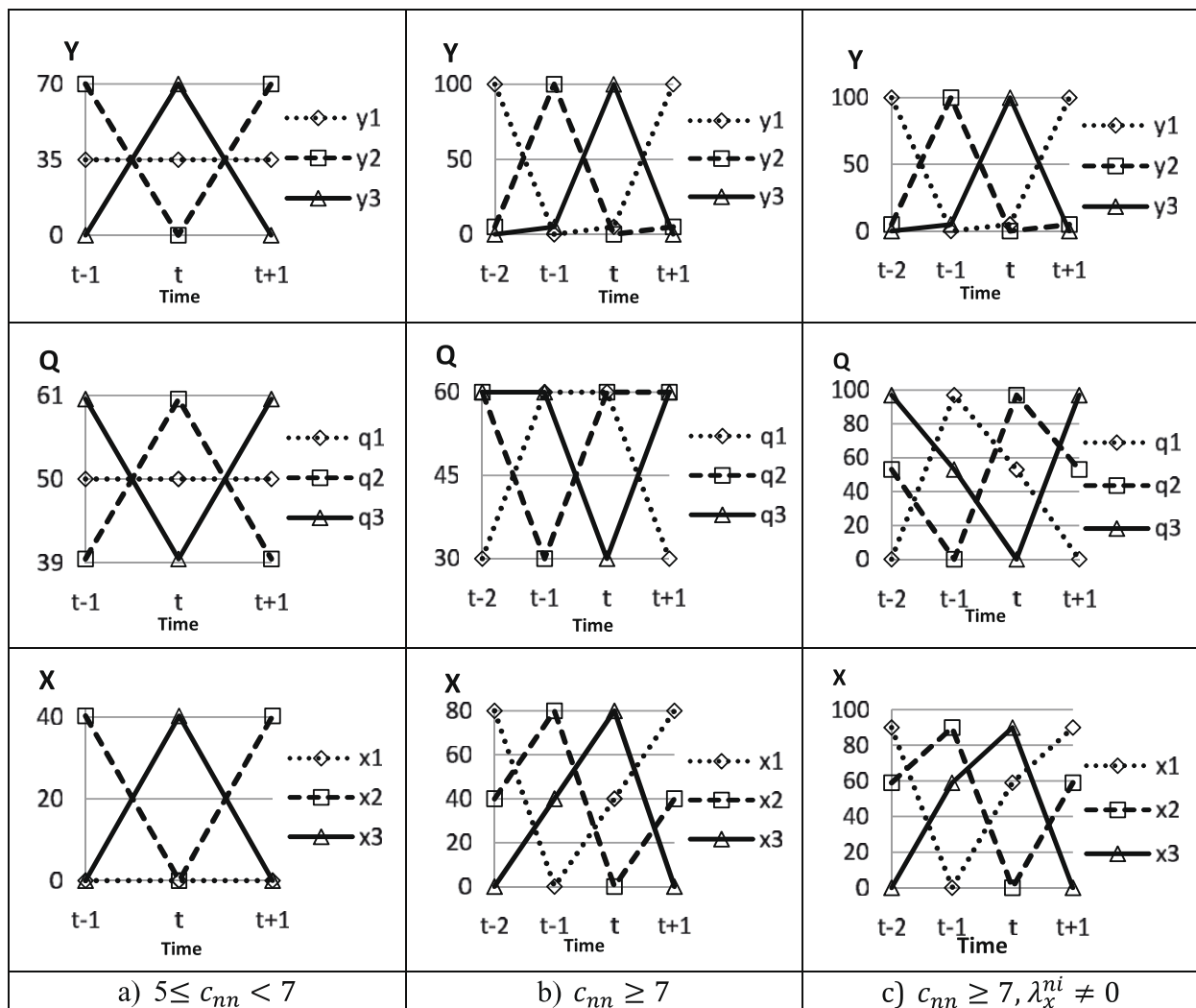


Fig. 11. Redundancy (a) Medium disruption impact (b) High disruption impact (c) High disruption impact when traffic are more sensitive to condition than effective capacity.

Table 2
Elasticities of demand.

Capacity		Condition	
$\lambda_u^{nn}, n = 1, 2, 3$	0.2	$\lambda_x^{nn}, n = 1, 2, 3$	–0.5
$\lambda_u^{1,2} = \lambda_u^{2,1}$	0.1	$\lambda_x^{1,2} = \lambda_x^{2,1}$	–0.25
$\lambda_u^{1,3} = \lambda_u^{3,1}$	–0.1	$\lambda_x^{1,3} = \lambda_x^{3,1}$	0.25
$\lambda_u^{2,3} = \lambda_u^{3,2}$	–0.1	$\lambda_x^{2,3} = \lambda_x^{3,2}$	0.25

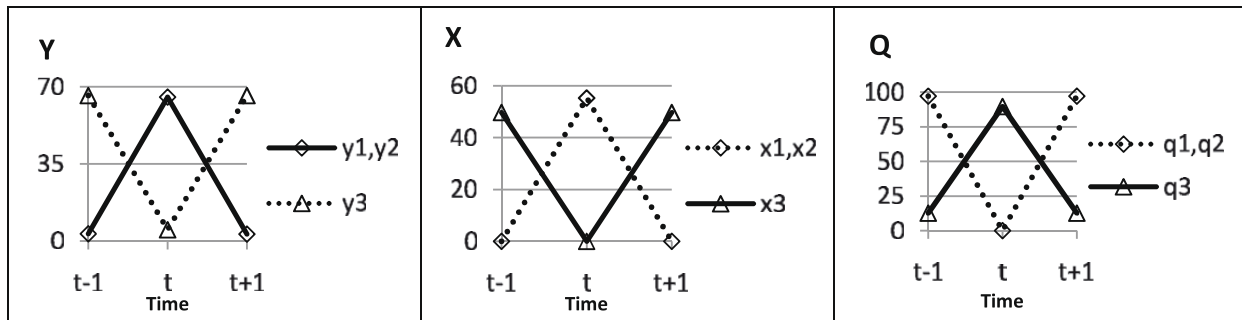


Fig. 12. Optimal solution for three-link network with complementary.

4.2. Additional numerical examples

In this section, we consider additional numerical examples that are intended to illustrate how the proposed framework can be generalized to more complex networks. We also discuss some difficulties. In particular, we consider the systems presented in Fig. 10 consisting of three facilities.

In the analysis of the system presented in Fig. 10a we consider three situations. The results are presented in Fig. 11. In the first two situations (Fig. 11a and b), we ignore demand sensitivity to condition, and set the demand elasticities with respect to capacity as follows: $\lambda_u^{nn} = 0.2, n = 1, 2, 3$ and $\lambda_u^{ni} = -0.1, n \neq i = 1, 2, 3$. These elasticities represent a situation where traffic on a facility is less sensitive to capacity in alternative facilities (than to its own capacity). Fig. 11a is for disruption costs where $5 \leq c_{nn} \leq 7, n = 1, 2, 3$. Fig. 11b is for $c_{nn} > 7, n = 1, 2, 3$. The main observation is that higher disruption costs lead to an optimal maintenance policy where the interventions between the three facilities are coordinated. The upper bound on the magnitude of an intervention means that, when the steady-state is reached, facilities are maintained in 2 out of every three periods.

In Fig. 11b we observe that when an intervention is applied on a facility, the swing level traffic that it serves is diverted equally onto the other two facilities without regard to their condition. Fig. 11c presents the results for a case where demand is also sensitive to condition. In particular, we set $\lambda_x^{nn} = -0.5, n = 1, 2, 3$ and $\lambda_x^{ni} = 0.25, n \neq i = 1, 2, 3$. In this case the state variables exhibit cycles that are symmetric for the three facilities that comprise the system.

In the analysis of the system presented in Fig. 10b, we set the elasticities as shown in Table 2.

The results for this case are presented in Fig. 12, where we observe cycles where the complementary facilities 1 and 2 are maintained simultaneously in periods where the substitutable facility, 3 is not being maintained.

5. Conclusions and directions for future work

We present a quadratic programming formulation for maintenance optimization of multifacility transportation systems. The proposed model provides a computationally-tractable framework that can be used to capture functional interdependencies that link facilities in transportation systems. In the formulation, each facility's deterioration and demand/traffic are identified and represented as a linear system, i.e., an autoregressive moving average model with exogenous inputs (ARMAX) model. The model explicitly captures the bidirectional relationship between demand and deterioration. That is, the state of a facility, impacts the demand/traffic, which in turn, determines the rate at which facilities deteriorate. The elements that comprise the system are linked because the state of a facility can impact the demand at other facilities. These relationships are captured by the specification of the cross-elasticities of demand. The quadratic objective can be used to capture nonlinearities in cost terms, which may, for example, reflect costs associated with congestion, vehicle wear and tear, and scale of maintenance activities. Although not used in the current paper, second order cost terms can also be used to capture pairwise economic dependencies, such as reductions in resource and personnel delivery costs when adjacent facilities are maintained simultaneously (cf. Durango-Cohen and Sarutipand, 2007).

In addition to presenting the framework, we provide a series of numerical examples to illustrate its advantages. The numerical results illustrate how network topology, traffic patterns, and significantly elasticity and cross-elasticity of demand

are key determinants of the structure of optimal maintenance policies. Specifically, the results provide insight about situations where it is optimal to coordinate (synchronize or alternate) maintenance interventions for clusters of facilities. Additional parametric analysis where we consider the impact of road-durability/structural-strength, deterioration rate, different network topologies, and larger network sizes are presented in Sarutipand (2008).

In terms of directions for future work, perhaps the most significant step would be to investigate the validity of the assumptions that deterioration and demand are adequately represented as ARMAX models (as shown in Equation sets (2) and (3) in the formulation). Preliminary statistical analysis and results in the context of formulating and estimating pavement deterioration models are encouraging (Chu and Durango-Cohen, 2007; Chu and Durango-Cohen, 2008). The appropriateness of using ARMAX models to represent demand/traffic, however, is still an open question. Traffic simulation/assignment may provide an interesting framework to address this issue, although embedded assumptions, e.g., (effect of road capacity or condition) on travel time estimates, criteria to select routes, etc., might impact the conclusions. Our conjecture is that the appropriateness of ARMAX models to represent demand depends critically on network structure/topology, traffic volume, and trip distribution, e.g., the linear approximation is probably valid for the single-origin, single-destination, two-link substitutable network presented in Fig. 2.

Acknowledgements

This work was partially supported by the National Science Foundation through Grant 0547471 awarded to the first author, and by a Royal Thai Government Scholarship awarded to the second author. The work was conducted while the second author was a doctoral student at Northwestern University.

References

- Archilla, A.R., Madanat, S., 2000. Development of pavement rutting model from experimental data. *ASCE Journal of Transportation Engineering* 126 (4), 291–299.
- Bertsekas, D., 1995. *Dynamic Programming and Optimal Control*. Athena Scientific.
- Chen, Z., 1998. Solution algorithms for the parallel replacement problem under economy of scale. *Naval Research Logistics* 45 (3), 279–295.
- Childress, S., Durango-Cohen, P., 2005. On parallel machine replacement problems with general replacement cost functions and stochastic deterioration. *Naval Research Logistics* 52 (5), 409–419.
- Cho, D., Parlar, M., 1991. Invited review: a survey of maintenance models for multi-unit systems. *European Journal of Operational Research* 51 (1), 1–23.
- Chu, C., Durango-Cohen, P., 2007. Estimation of infrastructure performance models using state-space specifications of time series models. *Transportation Research Part C* 15 (1), 17–32.
- Chu, C., Durango-Cohen, P., 2008. Empirical comparison of statistical pavement performance models. *ASCE Journal of Infrastructure Systems* 14 (2), 138–149.
- Dekker, R., Wildeman, R., Van Der Duyn Schouten, F., 1997. A review of multi-component maintenance models with economic dependence. *Mathematical Methods of Operations Research* 45 (3), 411–435.
- Dreyfus, S., 1977. *The Art and Theory of Dynamic Programming*. Academic Press, San Diego, CA.
- Durango-Cohen, P., Sarutipand, P., 2007. Capturing interdependencies and heterogeneity in the management of multi-facility transportation infrastructure systems. *ASCE Journal of Infrastructure Systems* 13 (2), 115–123.
- Durango-Cohen, P., Tadepalli, N., 2006. Using advanced inspection technologies to support investments in maintenance and repair of transportation infrastructure facilities. *ASCE Journal of Transportation Engineering* 132 (1), 60–68.
- Friesz, T., Fernandez, J.E., 1979. A model of optimal transport maintenance with demand responsiveness. *Transportation Research Part B* 13 (4), 317–339.
- Gil, P., Murray, W., Wright, M., 1981. *Practical Optimization*. Academic Press, Berkeley, CA.
- Golabi, K., Kulkarni, R., Way, G., 1982. A statewide pavement management system. *Interfaces* 12 (6), 5–21.
- Golabi, K., Pereira, P., 2003. An innovative pavement management and planning system for the road network of Portugal. *ASCE Journal of Infrastructure Systems* 9 (2), 75–80.
- Golabi, K., Shepard, R., 1997. Pontis: a system for maintenance optimization and improvement of US bridge networks. *Interfaces* 27 (1), 71–88.
- Jones, P., Zydiak, J., Hopp, W., 1991. Parallel machine replacement. *Naval Research Logistics* 38 (3), 351–365.
- McClurg, T., Chand, S., 2002. A parallel machine replacement model. *Naval Research Logistics* 49 (3), 275–287.
- Murakami, K., Turnquist, M., 1997. A dynamic model for scheduling maintenance of transportation facilities. *Transportation Research Record* 1524, 76–85.
- Ouyang, Y., 2007. Pavement resurfacing planning for highway networks: parametric policy iteration approach. *ASCE Journal of Infrastructure Systems* 13 (1), 65–71.
- Prozj, J., Madanat, S., 2002. A nonlinear model for predicting pavement serviceability. In: *Proceeding of ASCE Conference on Applications of Advanced Technology in Transportation*, Boston, MA, pp. 481–488.
- Sarutipand, P., 2008. An optimization framework for management of transportation infrastructure systems with interdependent facilities. Ph.D. thesis, Northwestern University, Evanston, IL.
- Smilowitz, K., Madanat, S., 2000. Optimal inspection, maintenance and rehabilitation policies for networks of infrastructure facilities under measurement and forecasting uncertainty. *Computer-Aided Civil and Infrastructure Engineering* 15 (1), 5–13.
- Thomas, L., 1986. A survey of maintenance and replacement models for maintainability and reliability of multi-item systems. *Reliability Engineering* 16 (4), 297–309.
- Uchida, K., Kagaya, S., 2006. Development of life-cycle cost evaluation model for pavements considering drivers route choices. *Transportation Research Record* 1985 2006, 115–124.
- Vander Veen, D., 1985. Parallel replacement under nonstationary deterministic demand. Ph.D. thesis, University of Michigan, Ann Arbor, MI.