MAINTENANCE AND MARKOV DECISION MODELS

(EQR085)

Rommert Dekker
Research Professor, Econometric Institute, Erasmus University Rotterdam, PO Box 1738, NL-3000 DR Rotterdam, The Netherlands
Telephone: +31 10 4081274
Fax: +31 10 4089162
E-mail: rdekker@few.eur.nl

Robin P. Nicolai
PhD student, Econometric Institute and Tinbergen Institute, Erasmus University Rotterdam, PO Box 1738, NL-3000 DR Rotterdam, The Netherlands
Telephone: +31 10 4082524
Fax: +31 10 4089162
E-mail: rnicolai@few.eur.nl

Lodewijk C.M. Kallenberg
Professor, Mathematical Institute, Leiden University,
Leiden, P.O. Box 9512, NL-2300 RA Leiden, The Netherlands
Telephone: +31 71 5277101
Fax: +31 71 5277005
E-mail: kallenberg@math.leidenuniv.nl

Corresponding Contributor: **R. Dekker**

Draft 3/12/2007

Keywords: Markov decision theory. Maintenance.

Abstract: In this contribution we first give an introduction to Markov decision theory. We state the main optimality criteria and solution approaches. Next we sketch how it can be applied in Maintenance theory. In particular we deal with the civil infrastructure sector and show what kind of results it brings. Finally we also indicate which problems arise in applications.

1 INTRODUCTION

Maintenance modeling and Markov decision chains have a large joint history. For both areas the main research started in the fifties. They formed a fruitful partnership, with theory allowing applications and applications guiding the need for theoretical developments. In this contribution we sketch the main ideas behind Markov decision chains and we indicate how they are applied in the maintenance area. In particular we treat the area of maintenance of civil structures. Apart from providing ample references, we state the pros and cons of the Markov decision modeling.

2 HISTORICAL PERSPECTIVE

Intense scientific interest in problems related to maintenance management originated only a few decades ago. The basis for the scientific support of maintenance is found in reliability engineering. The book *Mathematical Theory of Reliability* (Barlow and Proschan, 1965) indicates the start of scientific interest in maintenance problems. In the early stages, maintenance was almost exclusively interpreted as preventive replacement of components (*see* eqr 106). The development of some simple but insightful models describing aging phenomena of technical components, as well as the choice of appropriate actions to cope with these phenomena, showed that a scientific approach can be useful. One such model is the Markov decision chain. In the seventies and early eighties, the basic models were extended in several directions. In the eighties, scientists made a step forward in bridging the gap between the analytical models and practice by a systematic analysis of systems. Moreover, similar to many other areas of applied (stochastic) optimization, the major advances in information technology brought the use of quantitative analytic models within computational reach.

During the last fifteen years substantial progress has been made in developing quantitative decision support systems for maintenance management of complex systems. Sophisticated decision support systems are in operation nowadays in the oil industry (both for maintenance of refineries as well as offshore installations), the civil infrastructure sector (road, bridges and railway) and electric power generation. The number of organizations that make use of some kind of quantitative tools to support inspection and replacement of equipment is increasing rapidly.

3 MAINTENANCE

Maintenance can be defined as the combination of all technical and associated administrative actions intended to retain an item or system in, or restore it to a state in which it can perform its required function. The maintenance objectives can be summarized under four headings – ensuring system function (availability, efficiency and product quality); ensuring system life (asset management), ensuring safety and ensuring human well-being.

For production equipment, ensuring the system function is often the prime objective. Here, maintenance has to provide the right (but not the maximum) reliability, availability, efficiency and capability (i.e. producing at the right quality) of production systems in accordance with the need for these characteristics. In principle it is possible to give an economic value to the maintenance results, and a cost-balance can be done. Ensuring system life and asset management refers to keeping systems as such in proper conditions, whilst there are only indirect links to a possible production of goods or services. This objective is appropriate for civil structures, like buildings, dams, offshore platforms and roads, as their function is complex and not easy to measure. Often norms have to be set to define failure and the benefits of maintenance are therefore more difficult to quantify. In this case one has to minimize maintenance costs in order to meet the norms or conditions on states. Safety plays a role in case failures can have dramatic consequences, e.g., in the case of airplanes, nuclear and chemical plants. In this case testing and inspection activities constitute an important part of the maintenance work. Here, costs of maintenance have to be minimized while keeping the risks within strict limits and meeting statutory requirements. Finally, we refer to human well-being or shine as an objective, if there is no direct economic or technical necessity but primarily a psychological one (which indirectly may be economical). An example is painting which is not for protective reasons.

Many models for maintenance consider replacing of parts or systems. Typical models are the age and block replacement models (*see* eqr111; eqr131). These models consider only one state of the system, viz. working or failed. Next to that there are condition based models that consider multiple states and this can be modeled by Markov decision chains. For mechanical equipment, deterioration can be quickly, while also the economic importance of failures can be clear. In case of civil infrastructures deterioration can be very slow, and failure does not always have direct economic consequences. As a result it is much more important to decide when to do maintenance for such systems and to allocate the budget to those assets with the highest needs. Markov decision models can be helpful in this respect.

4 MARKOV DECISION MODELS

A Markov model is a special type of dynamic model with which the probabilistic evolution of a system can be modeled in time. The main assumption underlying such a model is that all information about the future behavior is captured in the state description. In other words, the present state provides all relevant information about the future behavior and knowledge about previous states is not necessary. More formally, a Markov chain is a discrete time process governed by a discrete state space E (observed at discrete time points) and transition matrix P, for which the Markov property holds, i.e.

$$P_{ij} = P(X_{t+1} = j \mid X_0 = i_0, X_1 = i_1, \dots, X_t = i) = P(X_{t+1} = j \mid X_t = i)$$

If the transition probabilities do not depend on t, then the Markov chain is said to be stationary. A Markov decision chain (MDC) is an extension of a Markov chain, where the Markov chain can be steered by actions and with which optimal actions can be determined.

An alternative name is Markov decision process as it basically describes a stochastic process. The term chain is used to indicate that it gives rise to a chain of states in time. It is defined by a four-tuple E, A, P and r, where E denotes the state space, $A(i)$ the action set in state $i \in E$, $P_{ij}(a)$, $i, j \in E$ the transition probabilities and $r_i(a)$, $i \in E$, $a \in A(i)$, the immediate rewards in state i when action a is chosen. The control is defined through policies and decision rules. A policy π is a sequence of decision rules $(\pi^1, \pi^2, \ldots)$ where π^t is function that assigns to each action a the probability of that action being taken at time t. For a memoryless or Markov policy the decision rule at time t is independent of the states and actions before time t. A policy is said to be stationary and deterministic if all decision rules are identical and nonrandomized. For any decision rule π we denote by $P(\pi)$, $r(\pi)$ the matrix, vector with $P_{ij}(\pi(i))$ and $r_i(\pi(i))$ as i-th element, respectively. Let $P^k(R) = P(\pi^k) \cdots P(\pi^1)$ and $P^0(R) = I$, the identity matrix. The evolution of the Markov decision chain under a given stationary and deterministic policy f is a Markov chain with transition probability matrix $P(f)$. Let C be the class of all policies and C_M be the class of Markov policies. In policy optimization the issue is to find the best action a in each state such that some criterion is optimized. The expected total α–discounted rewards are defined by

$$v^{\alpha}(R) = \sum_{k=0}^{\infty} \alpha^k P^k(R) r(\pi^{k+1}).$$

Here α is the discount factor which is taken from [0,1). For a stationary policy f one can also obtain the discounted rewards $v^{\alpha}(f)$ by solving a set of equations, i.e. (in vector notation)

$$v^{\alpha}(f) = r(f) + \alpha P(f) v^{\alpha}(f). \tag{1}$$

Again if $\alpha < 1$, this set has a unique solution. Note that the size of this set of equations is equal to the size of the state space. Normally solving such a set takes a number of operations which increases with the cube of the size of the state space. A policy R_{α} is called α-discounted optimal if for all $i \in E$, $R \in C$ we have

$$v_i^\alpha(R_\alpha) \geq v_i^\alpha(R).$$

A second criterion is the long-run expected average rewards, defined by

$$g_i(R) = \liminf_{N\to\infty} \frac{1}{N+1}\sum_{k=0}^{N}\sum_{j\in E} P_{ij}{}^k(R) r_j(\pi^{k+1}),$$

Let $g_i^* = \sup_{R\in C} g_i(R)$. In the finite-state and finite-action case there exists an average optimal policy and the supremum can be replaced by a maximum. In the case of a denumerable state space, an average optimal policy may exist only under certain conditions (like unichain and a certain recurrence). A policy R is long run average optimal if $g_i(R) = g_i^*$, for all $i \in E$. For both criteria it can be shown that one can restrict oneself to the class of Markov policies. The first step in optimization is the derivation of the so-called optimality equations. For the α-discounted rewards they are defined as follows.

$$v_i^\alpha = \max_{a\in A(i)}\{r_i(a) + \sum_{j\in E}\alpha P_{ij}(a) v_j^\alpha\}, \quad i \in E.$$

It appears that there exists a unique solution to them and that any policy which takes maximizing actions in each state is α-discounted optimal. The optimality equations for the long-run average costs consist of two equations, viz.

$$g_i = \max_{a\in A(i)}\{\sum_{j\in E} P_{ij}(a) g_j\}, \; i \in E,$$
$$g_i + v_i = \max_{a\in A^0(i)}\{r_i(a) + \sum_{j\in E} P_{ij}(a) v_j\}, \; i \in E,$$
$$\textit{where } A^0(i) = \{a \in A(i) \mid g_i = \sum_{j\in E} P_{ij}(a) g_j\}.$$

Again a solution *g*,*v* to this set of equations is unique up to a constant vector for v and the solution g equals g^*. Any policy which takes maximizing actions in both equations in all states is average optimal. These equations can be simplified in case the Markov decision chain is *unichain* (which means that under all policies the Markov chain has one minimal closed set) or even weaker, if it is *communicating*, i.e. for every two pairs of states, *i* and *j* there is a policy f such that $P_{ij}^k(f) > 0$ for some $k > 0$. In this case the optimal average reward vector is a constant vector g and the first equation in the optimality equation is automatically satisfied. The next problem is to solve the equations.

4.1 Solution methods

In principle there are three solution methods, i.e. policy improvement, value iteration and linear programming. Policy improvement (PI) works with a starting policy f_0 and determines in each step for each state an improving action. It appears that the stationary policy belonging to these actions also improves the objective function. As there are only a finite number of policies, the algorithm is finite. An algorithm for the α-discounted rewards is the following.

Step 0: let k=0 and choose a starting policy f_0

Step 1: policy evaluation: determine $v^{\alpha}(f_k)$ from equation (1)

Step 2: policy improvement: determine for each state i: $\max_{a \in A(i)} r_i(a) + \alpha P_{ij}(a) v_j^{\alpha}(f_k)$.

Let $f_{k+1}(i)$ be the policy consisting of the maximizing action for each state i. Choose $f_{k+1}(i) = f_k(i)$ where possible.

Step 3: If f_{k+1}=f_k, then stop, else increase k by one and return to step 1.

The successive approximation algorithm works with value iteration. It tries to find a solution to the optimality equations by repeatedly evaluating it. An algorithm is as follows

Step 1: choose a starting vector v_i^0, $i \in E$, e.g. by taking $v_i^0 = \max_{a \in A(i)} r_i(a)$ and set k=0.

Step 2: let $v_i^{k+1} = \max_{a \in A(i)} r_i(a) + \alpha \sum_{j \in E} P_{ij}(a) v_j^k, i \in E$.

Step 3: check if $| v_i^{k+1} - v_i^k | < \varepsilon$ for all $i \in E$, if not go back to step 2.

Step 4: determine the policy which takes maximizing actions in the last step.

The resulting policy has α-discounted rewards which are less than $2\alpha(1-\alpha)^{-1}\varepsilon$ different from the maximal α-discounted rewards. The algorithm may be speeded up by eliminating non-optimal actions. It is faster than PI if the transition matrix is sparse and only few transitions are possible. Finally the linear programming (LP) approach consists of formulating a LP for the optimality criterion. One can prove the following formulation

$$\begin{aligned} &\min \sum_{j\in E} \beta_j v_j, \\ &s.t. \quad v_i \geq r_i(a) + \alpha \sum_{j\in E} P_{ij}(a) v_j,\, a \in A(i),\, i \in E \end{aligned}$$

where β_j, $j \in E$, is just a vector of positive components. The solution equals the optimal discounted rewards and the policy taking maximizing actions is α-discount optimal. The advantage is that standard LP solvers can be used. From the LP formulation it follows that a Markov decision chain can be solved in polynomial time. More complexity results are given in Papadimitriou and Tsitsiklis (1987).

4.2 Issues in applying MDCs

In general there are two problems with applying the Markov decision model. The first is identifying states and establishing the Markov property and the second, which is related to the first, is the issue that state-spaces can be very large, with the consequence that computation times can be prohibitive.

New techniques concentrate on finding a functional form for the value function. Remember that if one has an approximation for the value function one can apply the PI algorithm to obtain the optimal policy. Several approaches have been suggested to this end. We like to mention neuro-dynamic programming from Bertsekas and Tsitsiklis (2005), reinforcement learning from Das et al. (1999) and simulation based approaches by Marbach and Tsitsiklis (2001).

4.3 Extensions of the Markov decision chain model

In a *semi-Markov decision chain* the time between observations of the system is not fixed, but also a random variable, with a distribution which depends on the state and action chosen. Most parts of the analysis can be extended to semi-Markov chains. It is also possible to define a corresponding discrete-time Markov chain, by discretizing the transition time distribution, but this does increase the state space considerably.

In a *continuous-time Markov process* the time between consecutive transitions is exponentially distributed. These Markov processes can be converted into discrete-time Markov chains. We refer to *eqr* 057 for an extensive treatment of Markov processes.

In a *partially observable Markov decision chain* the problem is that not all states are observable and hence no specific actions can be taken once the system enters these states. A separate branch of theory is devoted to these types of Markov chains. A problem is that it is no longer optimal to restrict oneself to Markov polices, but information on the past history plays a role. This makes both analysis and solution methods inherently more difficult. Overviews on this type of models have been given by Monahan (1982) and Lovejoy (1992).

5 MDC APPLICATIONS TO MAINTENANCE PROBLEMS

Maintenance problems have been among the first applications of Markov decision chains, as the model both allows a modeling of the deterioration as well as the determination of the structure of optimal policies. Early examples were given by Sasieni (1956) and Howard (1960) in his Car replacement problem. Let us describe the latter.

Consider a car, which age is expressed in periods of 3 months, so an age of j indicates an age of $3j$ months. Every period the car could either age with another period or have a catastrophic failure / accident after which we replace it with another car. Instead of letting the car age, we can also decide to replace it preventively with another car. In both cases one needs to choose what kind of age the replacement car should have. The system can be modeled by taking as state the age of the present car. Transitions to other states occur because of either ageing of the car or of a catastrophic failure in which case we replace the car. From a lifetime distribution the transition probabilities can be derived. Next we need cost figures for operating (e.g due to small maintenance), for failures and for acquiring a car of a certain age j, for all j. We would like to remark that in the Markov chain modeling with expected rewards criteria, we do not need to bother whether costs were actually incurred before or after a transition. One just takes an expectation over all possible transitions in a state to define the expected immediate rewards. One of the results of MDC theory is that the optimal policy is of the control-limit type, i.e. replace the car if its age is beyond a critical age threshold. This can be proven by showing that the α-discounted costs increase in the states. Other approaches would not give such results.

The Markov decision approach has particularly been followed in condition-based maintenance, where different conditions of equipment are considered and maintenance actions are based on these. Next there are a lot of studies in maintenance that make use of Markov models (*see* eqr057), but they use other methods to optimize rather than Markov decision theory.

Real applications have been very much in the sector of civil infrastructure, viz. roads, bridges and buildings. Nice overviews of applications have been given by Hontelez et al (1996), Frangopol et al (2004). Applications in other maintenance areas are less common, and as example we like to mention Amari et al (2006) for inspection planning of condition based maintenance and Stengos and Thomas (1980), who consider the replacement of blast furnaces in a steel works. It appears that preventively replacing both furnaces at the same time is not beneficial and that a specific cycle should be followed to reduce the probability that both furnaces fail together. In general there are two different ways to model states. In the first one takes the age of a structure central and takes that as basis to derive transition probabilities. An alternative is that one takes as state the remaining life span. In the other modeling one classifies the condition of the object according to some scale. Below we show a classification given by Scherer et al. (1994) for bridge components.

<< Insert Table 1 about here >>

Condition state	Description
9	New condition
8	Good condition: no repairs needed
7	Generally good condition: potential exists for minor maintenance
6	Fair condition: potential exists for major maintenance
5	Fair condition: potential exists for minor rehabilitation
4	Marginal condition: potential exists for major rehabilitation
3	Poor condition: repair or rehabilitation required immediately

Table 1: Condition state descriptions (taken from Scherer et al. (1994)).

Van Winden and Dekker (1998) also use such a condition scale, albeit with 5 levels. These conditions are typically identified by inspection by classified personnel. One problem however with such a scale is that the time to transition is not constant and that a semi-Markov modeling should be used. The latter can be converted by taking the residence time as extra state variable, but it would make the state space 2-dimensional.

Another practical issue is that often the state of multiple components or different deterioration mechanisms have to be taken into account. Van Winden and Dekker (1998) consider various elements in a building, like window frames, masonry, pointing and painting. Dekker et al. (1998) consider for a road segment longitudinal, transversal unevenness, cracking and raveling. All these aspects can be modeled but they make the state space multi-dimensional. A decomposition is not always possible because there can be actions which address multiple deterioration aspects or multiple components and which are cheaper than addressing these components separately.

The final issue is that often multiple objects need to be considered because of either economies of scale in maintenance or budget restrictions which allow only some maintenance to be carried out. We now would like to discuss approaches for roads, bridges and buildings.

A seminal paper on road maintenance was published by Golabi et al. (1992) and it discusses a pavement maintenance system for Arizona's highways. In the system the whole state highway system was modeled. LP was used as technique; it has the advantage that restrictions on the fraction of road segments in a certain state can be taken into account. A problem is namely that no costs are involved with bad states, so one has to find a driver to stay out of these states. A special modeling advantage of roads is that they are more or less similar objects and that large sums of money are involved in their maintenance. The disadvantage is that one needs to consider road segments of 100 meter and build up a quite large state space. The MDC approach is typically split up into several phases. First for a single segment a Markov decision model is formulated and an optimal policy is derived. From this policy it follows what kind of budget is needed to keep the road in a good condition. Next one considers a whole network of road segments and determines priorities to work on the segments in case the budget is restricted. The advantage of this approach is that the state of all the highways can be taken into account which gives a much farer budget allocation. Many papers extended this approach, e.g. Chen et al. (1996), Abaza and Ashur. (1999), Bako et al. (1995) and Dekker et al. (1998). Later on pavement management systems, using Markov decision models came into use in many US states and other countries.

For bridges the first paper was written by Scherer and Glagola (1994), who investigate the applicability and discuss the problems to keep the state-space limited. The first real application, PONTIS, was again made by a team by Golabi (1997). Later on more applications followed (see Hawk and Small (1998)). A problem of bridges compared to roads is that bridges are more or less unique, which makes the modeling quite time consuming.

For building maintenance similar things as for bridges can be done. One trick with buildings is that one can work with generic elements and determine policies for these. Next an upscaling is made. For example, pointing is measured in square meters and one can determine the total costs by considering the total number of square meters per building multiplied by the cost per square meter. Sometimes a subdivision is necessary, but otherwise most of the same parts have a similar ageing. This limits somewhat the number of components in a building, but considering a whole inventory of building does give a large state space. One of the interesting results of MDC in this respect is that a trade-off between quality and costs can be obtained which facilitates decision making. Van Winden and Dekker (1998) derived Figure 1. Again, this is difficult to achieve with other techniques.

<< Insert Figure 1 about here >> (eps figure will also be supplied)

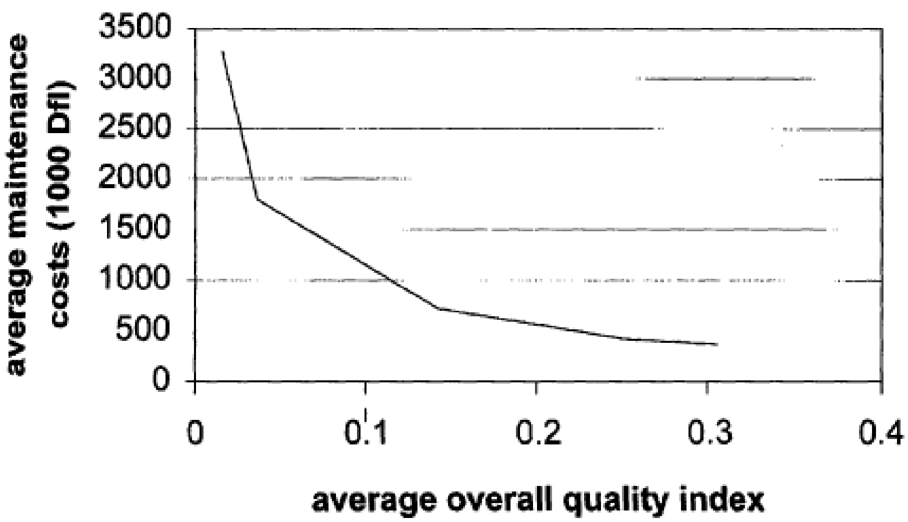

Figure 1 Trade off between quality and costs.

Finally we would like to mention quite some papers on partially observable MDCs and maintenance. Again this is a natural combination as quite often deterioration is not visible, but it can be modeled. Frangopol et al. (2004) give a number of examples for bridge maintenance.

A natural extension of these is the application of MDCs in health maintenance. Here again the modeling is on the evolution of a disease within a human with several stages. The MDCs are used to determine the effectiveness of early screening for diseases. As example we like to mention Singer (2001).

6 CONCLUSIONS

Markov Decision Chains have been very successful tool in modelling maintenance. As such they are the main model in civil infrastructure maintenance. A main problem lies however in the size of the state space. But new techniques allow for much larger state spaces. So far no real alternative models have emerged which allows to determine optimal policies. Hence, the main future issue is to develop new techniques to tackle the large multi-dimensional state spaces.

7 RELATED ARTICLES

eqr057, eqr 106, eqr 111, eqr 131

8 REFERENCES

- Abaza KA, Ashur SA. Optimum decision policy for management of pavement maintenance and rehabilitation. Journal of Transportation Research Board 1999. Transportation Research Record 1655: 8-15.
- Amari, S.V, L. McLaughlin and H.Pham, Cost-effective Condition-Based Maintenance using Markov Decision Processes, Proceedings Annual Reliability and Maintainability Symposium, RAMS 2006, IEEE 2006; 464-469.
- Bako A, Klafszky E, Szantai T, Gaspar L. Optimization techniques for planning highway pavement improvements. Annals of Operations Research 1995 58(1): 55-66.
- Barlow RE, Proschan F. Mathematical theory of reliability; John Wiley: New York, 1965.
- Bertsekas W, Tsitsiklis JN. Neuro-Dynamic Programming; Athena Scientific: Nashua 1996.

- Chen X, Hudson S, Pajoh M, Dickinson M. Development of new network optimization model for Oklahoma department of transportation. Annual Meeting TRB, 1996, no 1524, 103-108.
- Das, TK, Gosavi A, Mahadevan S, Marchalleck N. Solving semi-Markov decision problems using average reward reinforcement learning. Management Science 1999 45(4): 560-574.
- Dekker R, Plasmeijer R, Swart J. Evaluation of a new maintenance concept for the preservation of highways. IMA Journal of Mathematics applied in Business and Industry 1998 9: 109-156.
- Frangopol, DM, Kallen, M-J, Van Noortwijk, JM Probabilistic models for life-cycle performance of deteriorating structures: review and future directions. Prog. Struct Engng. Maer. 2004 6:197-212.
- Golabi K. A statewide pavement management system. Interfaces 1982 6: 5-21.
- Golabi K, Shepard R. Pontis: A System for Maintenance Optimization and Improvement of US Bridge Networks. Interfaces 1997 27: 71-88.
- Hontelez, JAM, Burger HH, and Wijnmalen DJD. Optimum condition-based maintenance policies for deteriorating systems with partial information, Reliability Engineering and System Safety, 1996 98(1) 52-63.
- Howard RA. Dynamic programming and Markov Processes; Technology Press of MIT. 1960.
- Lovejoy WS. A survey of algorithmic methods for partially observed Markov decision processes. Annals of Operations Research 1991 28: 47-66.
- Marbach P, Tsitsiklis JN. Simulation-Based Optimization of Markov Reward Processes. IEEE Transactions on Automatic Control 2001 46(2): 191-209.

- Monahan, GE. A survey of partially observable Markov decision processes: theory, models and algorithms. Management Science 1982 28(1): 1-16.
- Papadimitriou, C. and Tsitsiklis JN. The complexity of Markov decision processes, Mathematics of Operations Research, 1987 12: 441-450
- Sasieni M. A Markov chain process in industrial replacement. Operational Research Quarterly 1956 7: 148-155.
- Scherer WT, Glagola DM. Markovian models for bridge maintenance management. Journal of Transportation Engineering 1994 120(1): 37-51.
- Singer, M. Cost effectiveness of screening for hepatitis C virus in asymptotic, average risk adults. The American Journal of Medicine 2001 111(8): 614-621.
- Stengos D, Thomas L. The blast furnaces problem. European Journal of Operational Research 1980 4: 330-336.
- Winden C. van, R. Dekker, Markov decision models for building maintenance. Journal of the Operational Research Society 1998 49(9): 928-935.