International Institute for
Applied Systems Analysis
Schlossplatz 1
A-2361 Laxenburg, Austria

Tel: +43 2236 807 342
Fax: +43 2236 71313
E-mail: publications@iiasa.ac.at
Web: www.iiasa.ac.at

Interim Report IR-02-071

Applying optimal control to minimize energy use due to road infrastructure expansion

Sander M. Lensink (S.M.Lensink@fwn.rug.nl)

Approved by

Arkadii Kryazhimskii, (kryazhim@iiasa.ac.at, kryazhim@aha.ru)
Project Coordinator, *Dynamic Systems*

November 2002

Contents

Abstract

It is assumed that traffic jams have negative environmental effects. This implies an overlap between transportation policy and environmental policy. This paper quantifies the relative energy effects of congested traffic. By comparing these effects to the energy costs of construction and maintenance of roads, it is possible to balance those effects. The paper determines how fast and to what level the road infrastructure should expand, under the condition that life cycle energy consumption of the transportation system is minimized. By using the Pontryagin maximum principle, it is shown that optimal control theory can provide the solution that minimizes energy use. The paper concludes that the level of final expansion depends on the highest construction effort possible.

Acknowledgments

I like to thank the whole Dynamic Systems Project at IIASA for the good working conditions and especially my supervisors, Sergei Aseev and Arkadii Kryazhimskii, for their motivating support during my three months of stay at IIASA.

About the Author

Sander Lensink graduated in physics from the University of Groningen in 1998, with a specialization in environmental studies. He started his PhD-project at the University of Groningen in 1999, where his work aims to analyze the energy requirements of transport infrastructure expansion.

Applying optimal control to minimize energy use of road infrastructure expansion

Sander M. Lensink (S.M.Lensink@fwn.rug.nl)

1. Introduction

1.1. Global energy issues

As global warming manifests itself, a global effort is launched to reduce the emissions of greenhouse gases, prominently among which the gas carbon dioxide CO_2. The main sources of CO_2-emissions are fossil fuel combustion and cement production. Two ways of reducing the fuel related carbon dioxide emissions are to reduce the amount of fossil fuel burnt and to shift energy sources from fossil fuel to renewable sources. The Netherlands has committed itself in the Kyoto-treaty to reduce CO_2 equivalent emissions in the period 2008-2012 to 6% below the 1990 level. Energy is being used throughout every sector of the economy, as shown in table 1. Also every sector is expected to contribute to some extent to the needed emission reductions. These necessary reductions of which half is to be implemented in the Netherlands, are estimated at 50 Mton CO_2-equivalents in 2010 compared to unchanged policy, and are for 50% to be achieved by inland measures. The national Dutch transportation sector has to see its greenhouse gas emissions levels drop by 3 Mton in 2010, thus relatively to unchanged policy.

Table 1 Emission reductions in the Netherlands as implementation of the Kyoto-treaty. Of the total reduction of 50 Mton, mechanisms of Joint Implementation and the Clean Development Mechanism will achieve 25 Mton. Therefore, only 25 Mton reduction has to be achieved inside the Netherlands. Source: The climate policy implementation plan (Min.VROM, 1999)

Sector	Estimated emissions in unchanged policy (Mton CO_2-eq in 2010)	Projected reduction (Mton CO_2-eq in 2010)
Industry	89	10.0
Energy companies	61	8.0
Agriculture	28	2.0
Transportation	40	3.0
Households	23	2.3
Other	18	1.0

1.2. Mobility policy

The Dutch National Traffic and Transport Plan does not contain explicit policy aims to combat greenhouse gas emissions or to reduce energy use. It does point to the Climate Policy Implementation Plan for the measures to reduce emissions from the transportation sector. However, the Traffic and Transport Plan offers the framework for the expansion of the transport infrastructure. If expansion is carried out in such a way that all traffic jams in the Netherlands are resolved, an environmental benefit will form of 0.3 Mton CO_2 emissions prevented (Veurman et al., 2000). This computation has been a secondary result of a research that did not include the emissions of the road construction and road maintenance nor the formation of generated traffic or any adverse modal shift. Generated traffic is the traffic that results from the attractive influence of improved road capacity on road transport demand. The existing policy intentions in the two mentioned policy plans make it difficult to judge the desirability of the impact of specific road construction projects. Instead, this paper offers a mathematical framework to quantify the energetic impacts of road network expansion by comparing the energy consumption of construction works to the energy use related to fuel consumption. More over, in none of the policy measures any account is given of the relationship between infrastructure expansion and vehicle use. Table 2 shows the policy measures that are in effect or are to become in effect before 2010.

The fuel use of vehicles causes most of the emissions in the total transportation sector (Bos, 1998). Most sources of greenhouse gas emissions in the transportation sector are therefore mobile sources. It is generally believed that mobile energy consumers are more difficult to shift into renewable energy consumption than static ones. For that and other practical reasons, this paper does not look at the emissions patterns, but looks at the total energy consumption of the transportation sector. Since 90% of the emission of the transportation sector is a result of road transportation, the examples in this paper are examples in the road sector.

Table 2 Policy measures to reduce greenhouse gases in the Dutch transportation sector with its contributions. Source: Climate Policy Implementation Plan (Min.VROM, 1999).

Policy measure	Estimated effect (CO_2-eq savings in 2010)
Efficiency improvement of new vehicles	0-0.4
Changes in vehicle ownership tax (on efficiency grounds)	0.6
Tax on kilometer use (levy)	0.2
Changes in tax system for commuter and business traffic	0.1-0.3
Stricter upkeep of speed limits	0.3
Increase the use of fuel measuring devices in vehicles	0.5
Increase of tire pressure	0.3
Miscellaneous plans	0.2-0.3
Lowering N_2O emissions of combustion catalysts	0.5

1.3. Research framework

This paper fits into a broader PhD-research project conducted at the University of Groningen on analyses of the energy use of transportation systems. That project aims to perform a life cycle analysis of the transportation systems of road and rail traffic. The main subsystems are for each modality the infrastructure and the vehicles. This paper only looks at the modality of road traffic and does not take every life stage of every subsystem into account. Figure 1 shows which system elements are included in this research document.

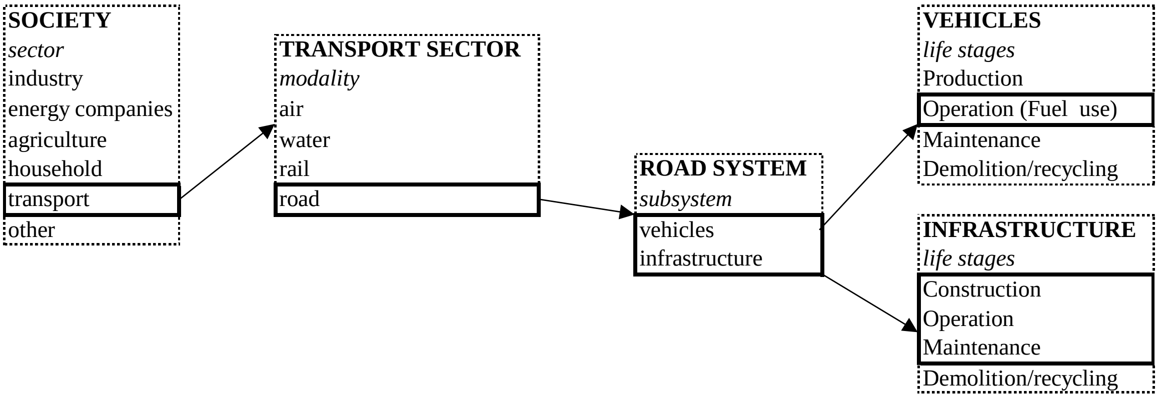

Figure 1 Inclusion of the specific elements in this research project. The energy use of the transport sector is the general topic of the PhD-project, while this paper specifically looks at the road system. From the subsystem of the vehicles, the life stages that are directly connected to the amount of vehicles are ignored, since it is assumed that there is no immediate relation between the amount of existing vehicles and the construction of specific road project. From the subsystem of the vehicles, the demolition phase is not looked at, since the relative contribution of this phase to the total energy use of the infrastructure is small and the demolition phase is mostly far in the future.

The PhD-work aims to study in depth questions on the allocation of energy resources for the expansion of the Dutch road infrastructure in the past and in the future; and into questions on the timing and desirability of applying capacity improving measures, like new roads and their location, and capacity improving measures on existing road sections.

This project is conducted as part of the Young Scientist Summer Program 2002. Its contents fits to the current research aim of the Dynamic Systems group, both to improve the environmental context of the research and to focus on developing optimization methods for large scale systems. The contents of this research project can also be seen as part of a PhD-project on “analyzing the energy and material use of future transport infrastructure expansion.”

2. Problem definition

2.1. Research question

The project researches the possibilities to minimize the energy resource requirements in infrastructure expansion. It will both include the energy requirements necessary for the physical infrastructure (construction, maintenance and operation), as well as energy requirements for the fuel consumption due to the use of that infrastructure by the vehicles. Given the relations between them, is it possible to use an optimal control model to minimize the sum of energy needed for construction, operation, maintenance and use during lifetime? If so, the model can be applied to the question: how and how fast should infrastructure expand, given the expected traffic growth, on an energetic criterion.

More general questions that may be answered by applying the results of this project, include the question under what circumstances – like expected traffic growth rate – will it be energetically beneficial to implement measures that increase capacity only slightly?

The underlying PhD-project looks specifically at Dutch transport infrastructure. This project narrows that focus to Dutch state roads. Therefore, any case in this project will also be subjected to this focus. Briefly stated, two conditions are valid for the Netherlands that cannot automatically be extrapolated to any other country. The first is the lack of major grades on state roads due to the absence of hills or mountains. The second is the already dense road network, which makes the construction of new roads through virgin land very rare. Still, the general concept that will be applied in this work can be made valid for cases in other countries as well.

2.2. Translating the problem into management terms

The direct effect of constructing new roads or improving existing roads is the increase in road capacity. Since a vehicle in a traffic jam uses more energy than in unperturbed traffic (Veurman et al., 2000) and since most traffic jams are caused, directly or indirectly, by insufficient capacity as shown by table 3, the energy use is treated as function of, among others, the capacity.

Table 3 Reported causes for traffic jam formation in the Netherlands in 2000. The bottleneck jams can also be seen as a capacity related cause. Source: Ministry of Transport and Public Works (Min.V&W, 2001).

Reporting cause	Number of traffic jam reports in 2000
Intensity vs. capacity	12%
Bottleneck	69%
Maintenance works	4%
Accidents	12%
Jams caused by watchers to accidents	1%
Other causes	2%

The dependence of energy use on the available road capacity is important. Thus, changes in capacity will influence the energy use. Supplying adequate capacity will therefore minimize energy use.

2.3. Applying optimal control to find a suitable solution

One can minimize the energy use, or manage the system, by controlling the capacity. The problem can therefore be regarded as a management problem. On one hand we have a relation between the capacity x and the applied control u. It is of the form:

$$\dot{x} = f_1(u,...) .$$

On the other hand, the energy use J depends in its turn on the capacity x:

$$J = f_2(x,...) .$$

The definitions of the various parameters and variables are given in section 2.4. In short, one should supply at a certain moment in time a certain capacity, guaranteeing the minimum use of energy. Finding the optimal control u is therefore a part of this research.

2.4. Concepts and definitions

The time unit in this research is hour. This corresponds to the practice in traffic management. The capacity of a road tells something about the number of vehicles that can use the road in a certain amount of time. A distinction is possible between the point capacity or flux – the number of vehicles possible passing one point in a certain amount of time – and the section capacity. The latter stands for the highest possible transport performance on a road section in a certain time span. Where the first is measured in veh/h, the latter is measured in vehkm/h.[1]

Furthermore, the capacity can indicate either the theoretically highest value, or the highest value given the specific circumstances like weather, road conditions and other traffic conditions. In this paper, the capacity x is the highest possible transport performance on a certain road section in a certain amount of time, with dimension veh·km/h. In this capacity x it is assumed that the traffic is moving at the optimal velocity and that no exogenous parameters – like the weather – have any influence.

Another important parameter that presumably influences the energy use is the actual traffic y. For compatibility with future studies, this paper uses the term 'transport demand'. The transport demand is principally only the *demand* for transportation, thus the transportation that people want to see performed. If other connections are left out of consideration, and thus route choice problems are ignored, one can state that the actual traffic is less than or equal to the transport demand. This paper supposes them to be equal, with the same dimensions as the capacity: veh·km/h.

The overall energy use J of the system, of which the specific boundaries are described in section 3.1.3, is the combined energy use of the infrastructure under consideration and the vehicles using that infrastructure. The infrastructure system needs energy for the construction of new roads and the maintenance of existing roads, thus for construction activities u. But it also requires energy, mostly electricity, for operation activities, like lighting, opening bridges and electronic signaling. The energy use of the vehicles is predominantly determined by the fuel consumption. The energy is measured in MJ.

[1] veh stands for 'vehicles'. The transport performance can have – depending on its context –dimensions in ton·km (cargo), pass·km (passenger) or veh·km (vehicles).

3. Background

3.1. Expansion of infrastructure

The expansion of an infrastructure system, in this paper specifically the road infrastructure, is synonymous with the enlargement of the capacity of an infrastructure system. The infrastructure system fundamentally is a collection of connections. Such a system can be improved both by enlarging the collection of connection, and by improving the functionality of one existing connection. In other words, the capacity of the infrastructure can be improved by building new roads and by improving existing roads.

Consider a road network between three cities *A*, *B* and *C*. Two roads connect these cities: road *1* from *A* to *C*, and road *2* from *A* to *B*. The respective roads have capacities ${}^{1}x$ and ${}^{2}x$. Figure 2 shows two ways of improving the network.

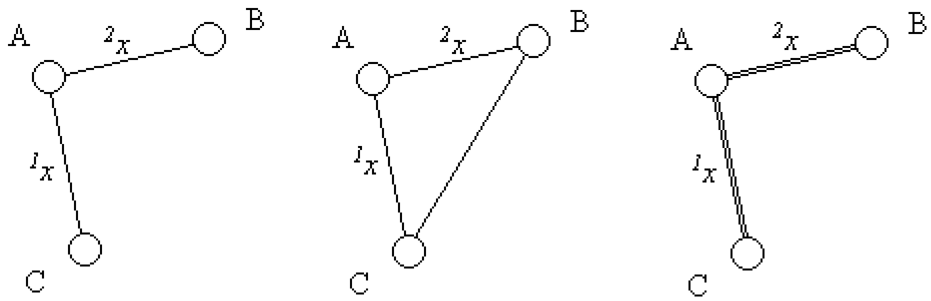

Figure 2 The left picture shows the initial state of the road system x_0. The system can either be improved through building a new road between cities B and C (middle picture), or through improving the existing roads 1 and/or 2 (right picture).

The complications in the form of complex system analyses in the case of the construction of new roads are not essential for the theoretical work of this paper. Therefore, it is considered to be sufficient to look only at the improvement of existing connections.

3.2. Dynamics of expansion relation between infrastructure and use

The initial state of the system x_0 in figure 2 is defined by the capacities of the roads. Let the capacities be symmetrical in direction, thus the capacity of the road from A to C is equal to the capacity of the road from C to A: $x^{A,C}=x^{C,A}=½\cdot({}^{1}x)$. The total capacity of the system can therefore be represented by:

$$x = \begin{bmatrix} - & x^{B,A} & x^{C,A} \\ x^{A,B} & - & 0 \\ x^{A,C} & 0 & - \end{bmatrix}$$

The transport demand *y*, averaged over a certain period of time, can also be assumed to be symmetrical, as shown by:

$$y = \begin{bmatrix} - & y^{B,A} & y^{C,A} \\ y^{A,B} & - & y^{C,B} \\ y^{A,C} & y^{B,C} & - \end{bmatrix}$$

Transport between the cities B and C is not directly possible, since there is no direct road between these two cities ($x^{B,C}$=0). Therefore, the load on the other connections between A and B, and between A and C, will be greater than the theoretical transport demand between those cities. In a complex system, it is far from clear which routes will be used to accommodate the transport demand. Finding the best composition of the matrix *x* to accommodate the matrix *y* with the lowest amount of energy possible requires therefore a solid routing model. In this paper, this problem will be avoided by only looking at upgrading a single road connection.

3.3. System boundaries

An important issue in transportation science is the phenomenon of generated traffic. "Generated traffic is the additional vehicle travel that results from road improvement. Generated traffic consists in diverted traffic (trips shifted in time, route and destination, and induced vehicle travel (shifts from other modes, longer trips and new vehicle trips" (Litman, 2001). Particularly estimations of the induced travel are hard to quantify.

Many of these problems are not addressed in this paper, due to the chosen system boundaries. The focus on only a single connection makes the research insensitive for changes in route or destination choice. As far as this research will average out the temporary peaks in transport demand, shifts in time form no complication. The research also looks only at a specific modality, the road transport. Therefore, modal shifts are not included. Although all these phenomena play an important role, they are not included in this research. The concept that is laid out in this paper does enable the future inclusion of these effects.

This paper looks at the improvement of a single road connection, where the transport demand is supposed to be constant in place, mode and time of the day. The road section should be of considerable length, so that any expansion can be considered continuous in time. The amount of latent transport is also considered absent.

It is assumed that the absence of grades in the research topic will have the following consequences: in hilly regions the relative fuel use of vehicles and the production energy for new infrastructure will be underestimated, while in mountainous areas the capacity will be overestimated as well.

3.4. Types of expansion

The capacity of a road section is not only dependent upon the number of lanes and the width of the road, but also on velocity, velocity distribution and lighting. Therefore, more measures exist to improve the capacity of a road than merely by 'laying down more asphalt'. Out of a list of 31 published capacity improving measures, a selection is shown in table 4. This selection is only meant to give an impression of the kinds of measures possible.

Table 4 Capacity improving measures as are under consideration in the Netherlands. Source: Ministry of Transport and Public Works as published in (Alberts, 2002).

Measure	Implementation level	Jam type
Dynamic change in number of lanes	Road section	Intensity/Capacity
Dynamic change in speed limits	Road section	Intensity/Capacity
No overtaking for trucks	Road section	Intensity/Capacity
Shoulder use in rush hour	Road section	Intensity/Capacity
Closing junctions	Network	Intensity/Capacity
Dynamic Route Information Panels	Network	Intensity/Capacity
Incident management	Road section	Accident
Additional measures at Work in Progress	Road section	Maintenance

It should be noted that every measure requires a measure-specific amount of energy for construction and maintenance. The environmental cost-benefit ratio can be no means assumed to be equal for all measures. In this paper, the measure investigated is the one in which the capacity of a road section will be improved in the most straightforward way, by laying down more asphalt and creating additional lanes.

3.5. The Netherlands

Both the length of the road network and the transport demand is rising in the beginning of the 21st century, but the growth rate is slowly decreasing. It is possible to estimate a saturation level of both network length and transport demand (Grübler and Nakićenović, 1991). Figure 3 shows that the saturation level for traffic in the Netherlands is not yet achieved. Figure 4, however, indicates that the road system hardly expands at all in mere length[2]. So, improving the network in other ways than increasing the length is in the last decades mostly responsible for accommodating the growth in traffic. This is also some qualitative justification of the choice to look only at road improvements.

---

[2] The length of the road is the length of the physical connection. It should therefore be noted that in the context of this paragraph, the addition of new lanes to the road does not lead to an increase in length.

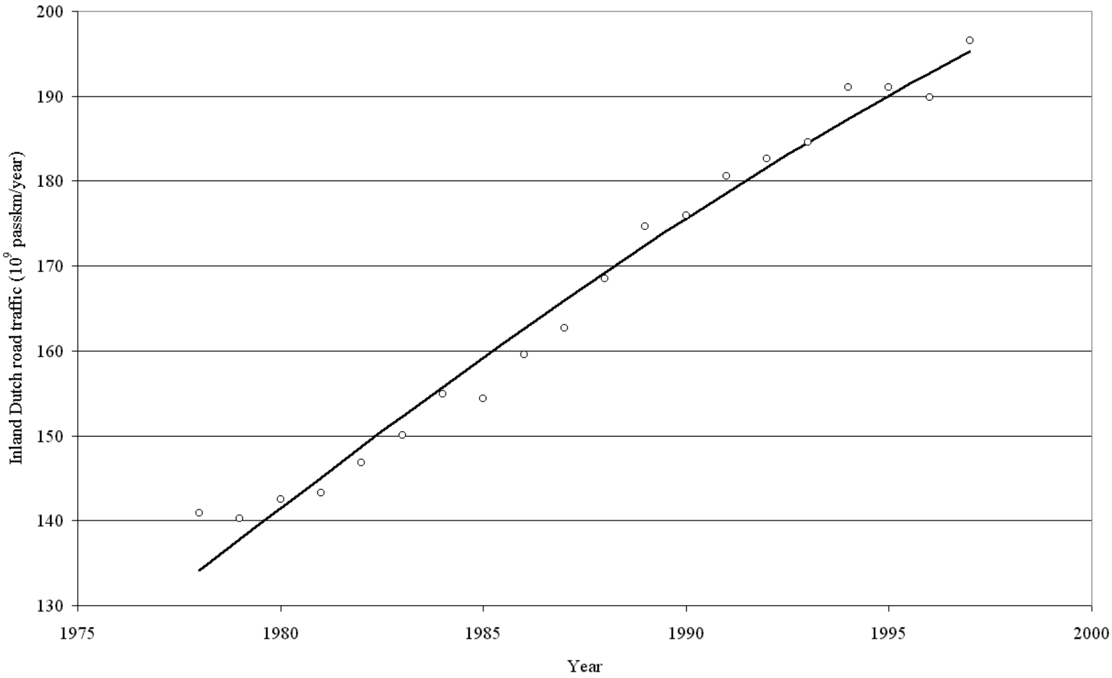

Figure 3 The Dutch road traffic is increasing fast, but it seems that the rate of growth is slowing down in the last decade. Using the least squares method to fit a logistic curve to the data, the saturation level can be determined at approximately 250·10^9 pass·km/year.

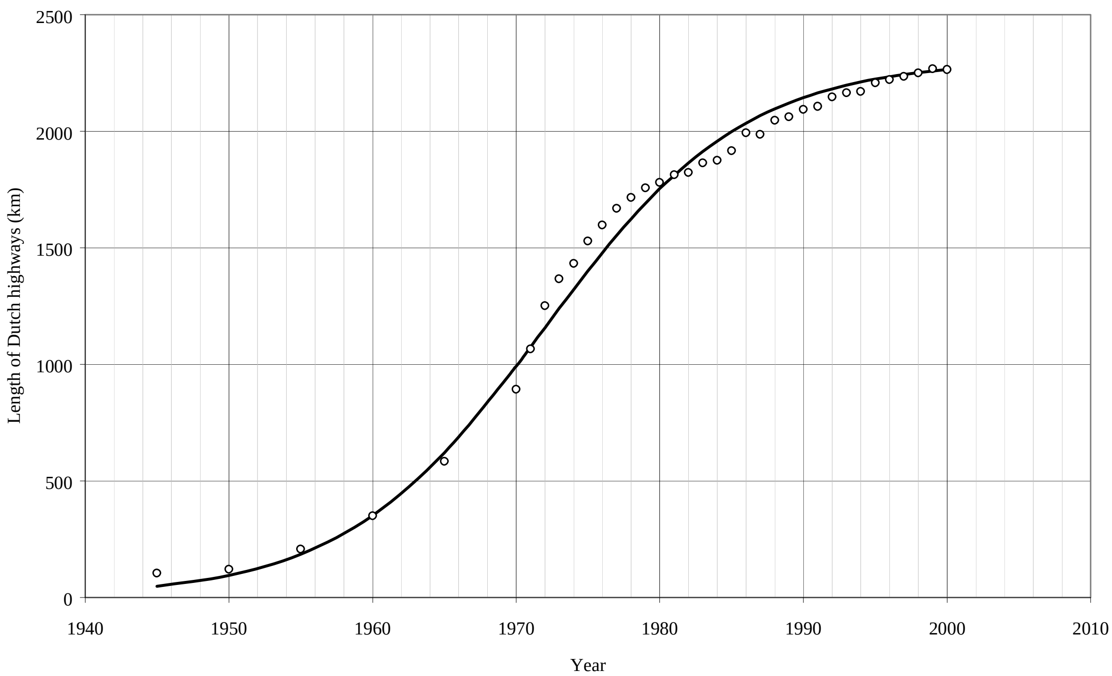

Figure 4 The length of the Dutch state roads has almost reached saturation levels. The logistic curve is fitted to the data with an additional constraint that the curve should intersect with the latest data of the year 2000. Should this constraint not have been imposed, than the saturation level would have been below the length of the year 2000.

3.6. Life cycle analysis

The concept of life cycle analysis is well developed. The main idea is to assess the impacts of a product from cradle to grave. In essence, the impacts from all the different life stages of the product are added together, commonly – most products have a relative short lifetime – without time discounting. There is no default set of impacts that are assessed in the analysis; the impacts can in principle range from economical and sociological to environmental ones. As discussed previously, this paper looks specifically at the energy use. It is therefore better to talk about 'energy analysis' (IFIAS, 1974).

The life stages of any product are the production phase, the usage phase and the discard phase. This chain is depicted in figure 5.

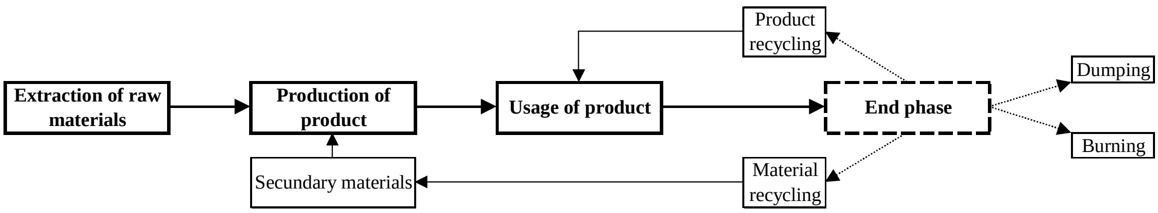

Figure 5 The life chain of a product. Some feedback loops exists in this chain. They represent the recycling options.

For an energy analysis, one has to collect data about the amount of materials needed for production of the product, and energy needed for assembly; for the usage phase as well, one needs to know the amount of materials and direct energy use. The materials represent a certain amount of energy needed to extract, to manufacture and to transport the materials. This specific energy use is called the Gross Energy Value of a material, commonly expressed in MJ/kg, and values for GER can be found in literature (Kok et al., 2001). These include assumptions on recycling rates. By knowing the amount and type of materials needed, one can calculate the total embodied energy in the materials. For transportation systems, this calculation can be used as a good approximation of the total production energy. This can be deduced from the thesis of Bos (Bos, 1998).

An energy analysis on the transportation system follows basically the same scheme as figure 5. The mobility system should, however, be seen as a system consisting of several rather independent products. Firstly, any transportation system consists of an infrastructure system and a collection of vehicles. Secondly, the mobility system consists of several modalities. In most cases, the infrastructure and vehicles are not shared between modalities. A single modality system, as is the case in this paper, is the simplest form of a transportation system. The infrastructure can in this case be seen as a collection of roads. The usage phase of the system requires special attention. Usage includes both the direct energy consumption and the maintenance requirements, like repair materials. For the infrastructure subsystem, the energy in the usage phase is mostly needed for the asphalt requirements for repair and electricity use for operation (bridges, lighting and electronic devices). For the vehicles subsystem, the fuel consumption of the vehicles is the dominant process in energy consumption.[3]

[3] A golden rule for many products is that the direct energy requirements form 85% of the total life cycle energy requirements. The indirect requirements (for production, maintenance and removal) equal 15%.

4. Theory

4.1. State equation

The state of the transportation system under consideration is represented by the capacity of it. The capacity x is therefore describing the state of the system. Suppose the infrastructure can be treated as a production-inventory system. For a description see (Sethi and Thompson, 2000).

The system can be improved by new production with a production rate p: ${}_1\dot{x}(t) = p(t)$. This production rate depends on time. The system also looses quality at a constant rate δ in time: ${}_2\dot{x}(t) = -\delta \cdot x(t)$. The total deterioration depends linearly on the existing capacity. The deterioration rate δ is constant in time. The autonomous deterioration can be, at least partially, counteracted by conducting maintenance at rate $m(t)$: ${}_3\dot{x}(t) = m(t) \cdot x(t)$. Combining these effects together, we get a first form of the state equation:

$$\dot{x}(t) = p(t) - (\delta - m(t)) \cdot x(t) . \tag{4.1}$$

As the capacity is measured in vehkm/h, the dimensions of p are vehkm/(yr·h), and those of δ and m are yr^{-1}. The initial state of the system is given by $x(0) = x_0 > 0$.

Other constraints on the system are:

$p(t) \geq 0$ for all $t > 0$; $\delta > 0$; $0 \leq m(t) \leq \delta$ for all $t > 0$. This upper bound on m is necessary, since the maintenance (i.e. repair of damage) cannot lead to completely new capacity.

The total construction effort $u(t)$ is defined as the sum of the production and maintenance activities: $u(t)=p(t)+m(t)\cdot x(t)$. Substituting in equation (4.1) gives a relation with a single control parameter u:

$$\dot{x}(t) = -\delta \cdot x(t) + u(t) . \tag{4.2}$$

The construction effort is limited by the maximum construction effort:

$$0 \leq u(t) \leq u_{max} .$$

4.2. Production and maintenance

It is difficult to give a precise definition of both production and maintenance. Both are construction works, but often the construction work for new production is combined with necessary maintenance. A uniform classification of maintenance does not exist, but some distinction in road maintenance and improvement work is possible (Paterson, 1987):

Routine maintenance	*m*	Localized repairs (typically less than 150m in continuous length) of pavement and shoulder defects, and regular maintenance of road drainage, side slopes, verges and furniture.
Resurfacing	*m*	Full-width resurfacing or treatment of the existing pavement or roadway (inclusive of minor shape correction, surface patching or restoration of skid resistance) to maintain surface characteristics and structural integrity for continued serviceability.
Rehabilitation	*m* or *p*	Full-width, full-length surfacing with selective strengthening and shape correction of existing pavement or roadway (inclusive of repair of minor drainage structures) to restore the structural length and integrity required for continued serviceability.
Improvement	*p*	Geometric improvements related to width, curvature or gradient of roadway, pavement, shoulders, or structures, to enhance traffic capacity, speed or safety; and inclusive of associated “rehabilitation” or “resurfacing” of the pavement.
Reconstruction	*p*	Full-width, full-length reconstruction of roadway pavement and shoulders mostly on existing alignment, including rehabilitation of all drainage structures generally to improved roadway, pavement and geometric standards.
New construction	*p*	Full-width, full-length construction of a road on a new alignment, upgrading of a gravel or earth road to paved standards, and provision of additional lanes or carriageways to existing roads.

Choosing the rehabilitation to be part of *m* implies that $m(t)=\delta$ in normal maintenance conditions. The Dutch administrative maintenance practice is to make a distinction between continuous preventive maintenance *m* and discreet rehabilitation maintenance would be included in *p.* In this case $m(t) < \delta$. The former has theoretical, mathematical advantages, while the latter has the benefit of the possibility to research decision to let additional capacity construction coincide with necessary rehabilitation.

In the case that it is possible to use one expression for all the construction works, thus the usage of the construction effort $u(t)$, the choice in classification between $p(t)$ and $m(t)$ becomes arbitrary.

4.2. Deterioration

The following formulas are taken from a publication of the World Bank (Paterson, 1987). The deterioration of the road surface can be measured in the International Roughness Index (IRI). A specific definition of IRI can be found in the World Bank publication. Roughness itself can be defined as "the deviations of a surface from a true planar surface with characteristic dimensions that affect vehicle dynamics, ride quality, dynamics loads and drainage." An empirical formula to predict the roughness R is:

$$R(t) = \left[R_0 + 725 \cdot (1+S)^{-5.0} \cdot L_4(t)\right] \cdot e^{0.0153t} \ . \qquad (4.3)$$

The roughness $R(t)$ in m/km IRI, at age t in year since construction depends on two major parameters (R_0 is typically between 1 and 3 for new roads):

$L_i(t)$ is the cumulative traffic loading at time t, in million ESA (assuming that the load damage increases with power i). Mostly it is predicted that i=4. ESA is the number of equivalent 80 kN single axle load.

S is the so-called modified structural number of pavement strength. It can be calculated using the formula:

$$S = 0.04 \sum_i a_i h_i + 3.51 \cdot \ln(B) - 0.85 \cdot \ln^2(B) - 1.43$$

a_i : material and layer strength coefficients;

h_i : layer thickness in mm ($\Sigma h \leq 700$ mm);

B: in situ California Bearing Ratio of subgrade in %.

Table 5 Emperical values for a_i and CBR to be used in formula above.

Pavement layer	Strength coefficient a_i
Surface course	
Asphalt concrete	0.30-0.45
Base course	
Granular materials	0.0-0.14
Cemented materials	$0.075+0.039 \cdot UCS-0.00088 \cdot UCS^2$
Subbase and subgrade layers	
Granular materials	$0.01+0.065 \cdot \ln(B)$
Cemented materials UCS>0.7 Mpa	0.14

UCS: unconfined compressive strength in MPa after 14 days.

Typical values for S are between 2 and 6. Let, for argument sake, $S=2.4$; $R_0=1.5$. Formula (4.3) would then lead to (with t in years):

$R(t) \approx [1.5+1.60 \cdot L_4(t)] \cdot e^{0.0153t}$.

The cumulative traffic loading L_4 can be computed as: $L_n = \sum_a N_a \cdot \left(\frac{a}{80}\right)^n$ with N_a the number of passing axle loads a. For n the figure commonly used is n=4. The commonly used dimension is ESA.

Also, a relation needs to be established between *v* and *R*. (Paterson, 1987) gives a graphical representation of such a relation. The velocity is a slowly decreasing function of *R*, see figure 6.

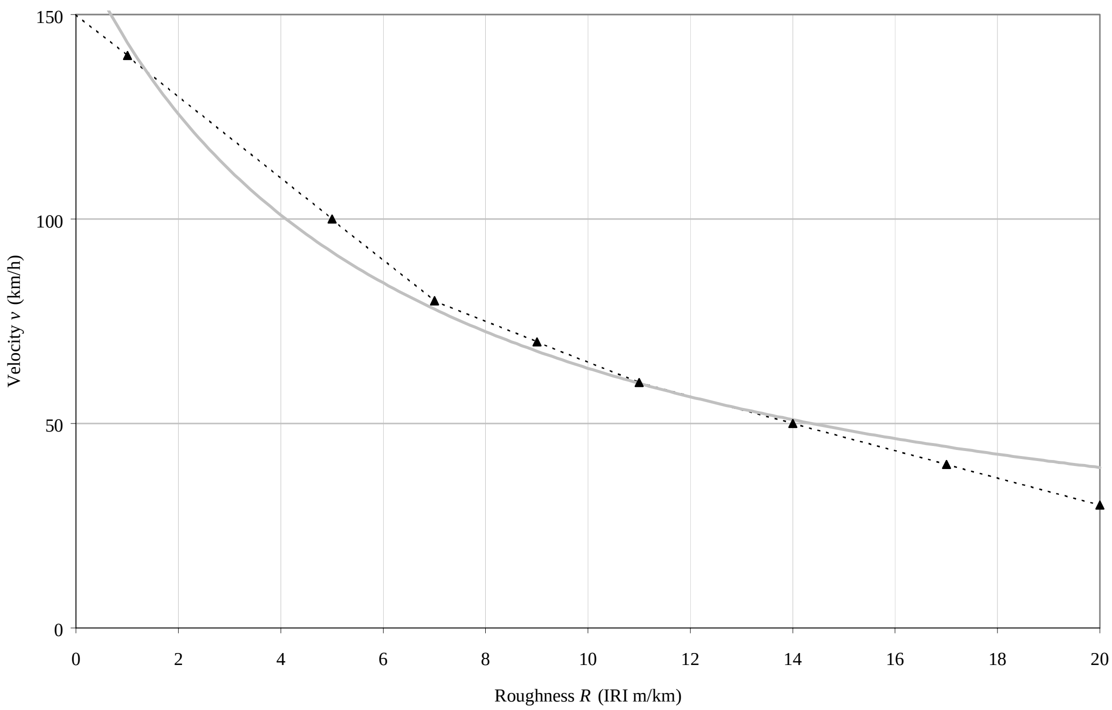

Figure 6 The relation between *R* and *v*. The function shown in gray is given by: *v(R)=166.5/(1+0.162·R(t)).*

The assumption that the capacity of a road decreases proportionally to the velocity allows an estimation of the autonomous deterioration of the capacity possible. So: $x(t)/x_0=v(t)/v(0)$. It follows that:

$$x(t) = \frac{166.5}{1+0.162(R_0+1.6\cdot L_4\cdot t)e^{0.0153t}}\cdot\frac{1}{v(R_0)}\ x_0.$$

If the transport demand *y* is constant in time, than $L_4(t)$ is constant in time. Now, using the state equation (4.2) and setting *p=0* and *m=0*, it is possible to determine the autonomous deterioration rate δ, since $\dot{x}(t) = -\delta\cdot x(t) \Rightarrow \delta = -\frac{1}{x(t)}\cdot\frac{\mathrm{d}x(\mathrm{t})}{\mathrm{d}t}$. In figure 7 this is numerically determined for R_0=1.5 m/km and L_4 = 1 ESA. It shows that *R* is not constant in time, but – for traffic densities common in the Netherlands – mostly ranges between 0 and 0.3.

The maintenance standards in the Netherlands state that roads with IRI<2.6 do not require maintenance. Roads where 2.6 ≥ IRI <3.5 need maintenance planning, since roads with IRI ≥ 3.5 require immediate maintenance. In the Netherlands, 0.2% of the state roads have IRI ≥ 3.5, while 98.7% have IRI < 2.6 (MinV&W, 1999).

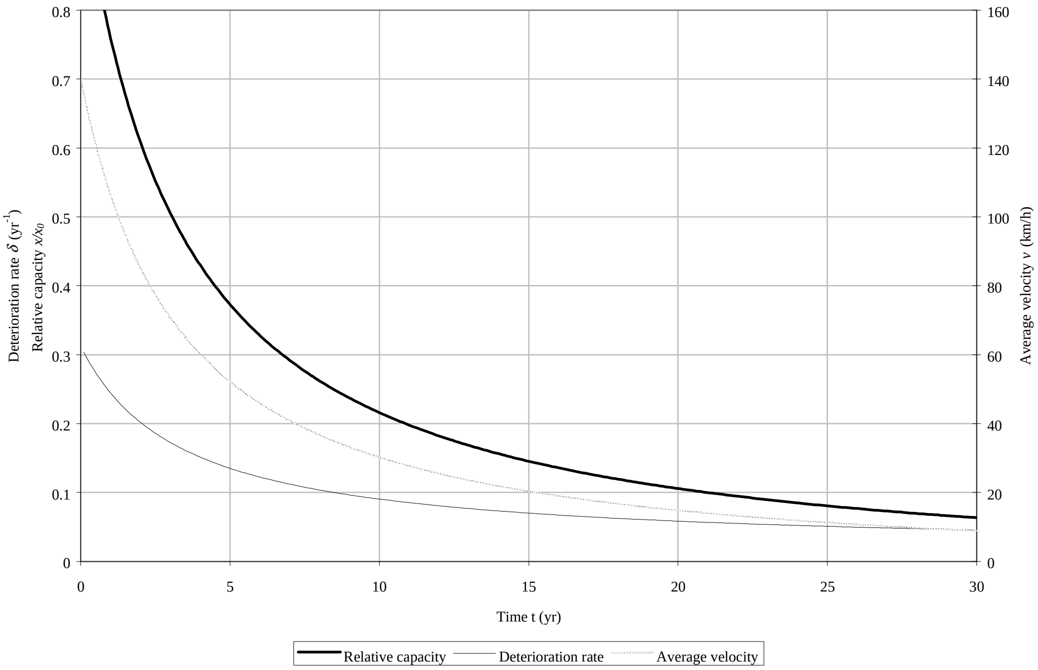

Figure 7 The deterioration rate δ declines as a function of time. Regular maintenance in the Netherlands is conducted every 6 to 8 years (Alberts, 2002).

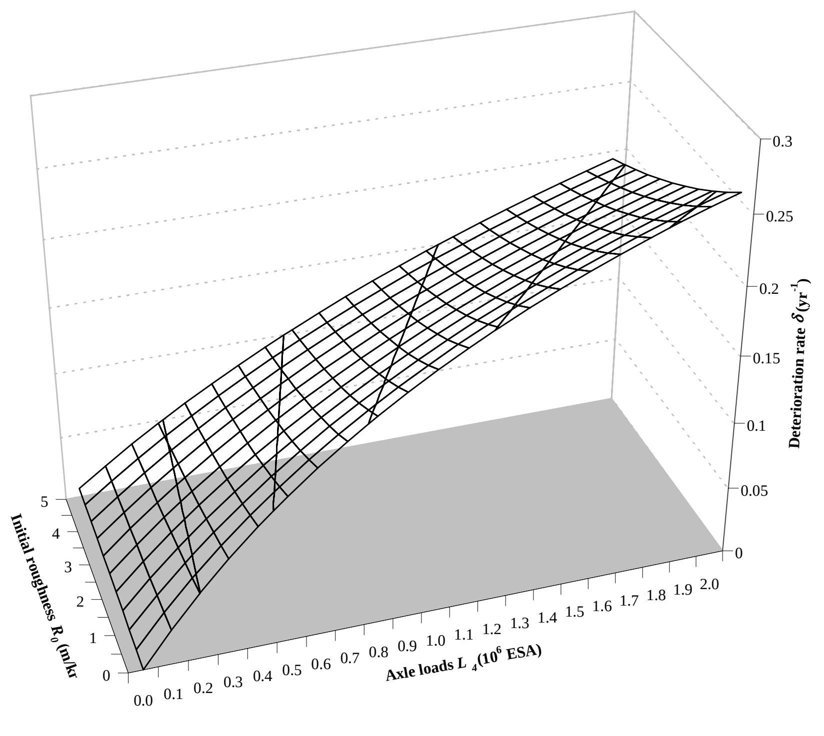

Figure 8 The deterioration rate after 2 years for several cumulative axle loads L_4 (horizontal axis) and values of roughness R_0 (vertical axis).

As maintenance is conducted every 6 to 8 years, then it follows, if maintenance *m* is defined such that $m(t) \approx \delta$, that 0.13 yr$^{-1} < \delta <$ 0.17 yr^{-1} for normal Dutch road conditions. The conclusion is that on average: δ=0.15 yr^{-1}.

4.3. Objective function

The objective function J representing the time discounted life cycle energy use of the system is in its core the summation of the energy use of the different life stages:

$$J = \int_0^\infty e^{-\rho t} \cdot [E(p) + H(m, x) + F(x, y)]dt \,. \tag{4.4}$$

The three utility functions E, H and F represent the energy use of – respectively – the production phase of the infrastructure E, the maintenance and operation phase of the infrastructure H, and the operation phase of the vehicles F. These functions are all positive: $E(p)\geq 0$ for all $p\geq 0$; $H(m,x)\geq 0$ for all $m\geq 0$, $x\geq 0$; $F(x,y)\geq 0$ for all $x\geq 0$, $y\geq 0$.

$E(p)$ is a function representing the energy use for the construction of new capacity. Every capacity increasing measure on a road has a specific influence on the increase in capacity, with a specific energy requirement. All capacity increasing actions exist of one or more distinctive engineering measures. For most measures i, the relationship will be of a linear type: $E_i(p)=\alpha_i \cdot p$. The parameter α_i is assumed to be constant, but might decrease slowly in time as technology improves. For some measures, the relation between the energy use E and the production rate p might be less than linear in p, as initial installation costs may be high (lighting, electronic traffic regulation). For asphalt construction, is it presumably more than linear in p. The reason for the latter is that a third lane on a highway has less effect than a second one, and a fourth less then a third, etcetera, while the energy needs are largely determined by the amount of asphalt that is equal for every lane. Similar relationships will also exist in maintenance requirements: $H_i(m,x)=ß_i \cdot m \cdot x$. The parameter $ß_i$ has a similar function as parameter α_i in the production energy function. There does not exist a clear correlation between α_i and $ß_i$. Some measures with low α will have a high $ß$. (One should think about measures which require large continuous electricity supply in the operation phase).

So it is assumed that most relations can be characterized by:

$$E(p(t))=\alpha_i \cdot p(t),\ \alpha_i > 0;$$

$$H(m(t),x(t))=ß_i \cdot m(t) \cdot x(t),\ ß_i>0.$$

A thorough analysis for the whole life cycle of infrastructure is carried out by Bos (Bos, 1998). This study results in energy for total production of a standard freeway of $64 \cdot 10^6$ MJ/km. This accounts for $\alpha_{road} = 8.0 \cdot 10^3$ MJ·h·veh^{-1}·km^{-1}. The materials are accountable for most of the energy requirements, both for the construction and the maintenance phase. If one assumes that at one point in time all the materials will have to be replaced (*), the assumption $\alpha \approx ß$ is valid. However, the current data suggests that maintenance requirements are substantially lower than construction requirements. The current data looks at the material requirements during the functional lifetime of the road. That is, until the road subbase (sand bed and lower asphalt layers) and the concrete artworks (bridges) need replacement. Therefore, only asphalt renewal in the top layers is included. Taking data from current Dutch studies (Bos, 1998; Alberts, 2002), one can estimate $ß$ at 150-500 MJ·h·veh^{-1}·km^{-1}. For the reason (*) mentioned above, this value of $ß$ is an underestimation. Therefore, with the assumption $\alpha=ß$ and only taking traditional road construction into consideration, part of the objective function can be expressed in terms of the total construction effort: $E(p)+H(m,x)=\alpha \cdot p(t)+ß \cdot m(t) \cdot x(t)=\alpha \cdot u(t)$.

4.4. Main utility function

4.4.1. Estimating transport demand

The energy function that relates to the vehicular fuel consumption *F(x(t),y(t))*, however, is determined by a more complicated relationship. Suppose an exogenous function *y(t)* exists that forecasts the transport demand. See figure 9 for an example. The flux of traffic Φ at one point is, on average, given by: $\overline{\Phi}(t) = y(t)/\ell$. With the case of figure 9, in which the length ℓ=54.3 km, the average throughput for 2010 is given by: ${}^{A12}\overline{\Phi}_{2010} = {}^{A12}y_{2010} \big/ {}^{A12}\ell \approx \frac{6.6 \cdot 10^6}{54.3} = 120 \cdot 10^3 \text{ veh/day} = 5065 \text{ veh/h}$. Note that this formula does not include rush-hour peak traffic. In this paper, the transport demand is considered constant in time: *y(t)=y*. For the answer to the question to which extent the transport infrastructure should grow, is it best to use the value of $y(t)=y_{max}$ in the equations.

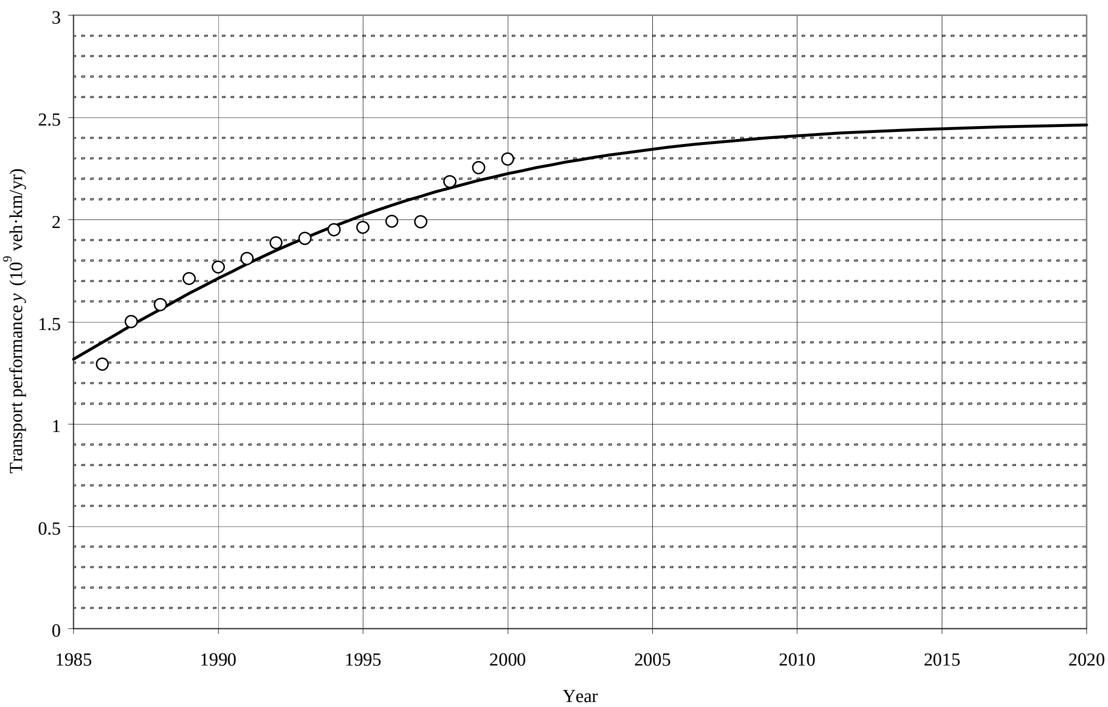

Figure 9 The transport performance on the Dutch highway A12, The Hague-Utrecht, including an extrapolation until 2020. The baseline of *t* in years is *t*=0 for the year 1986. The extrapolation curve is given by ${}^{A12}y(t) = {}^{A12}y_{max}/(1+e^{-0.136 \cdot t+0.258})$, with y_{max}=2.48·10^9 vehkm/year. The curve is fitted using numerical least squares methods, but it should be noted that the form of the outcome of the figure is subjective to the chosen fit curve.

4.4.2. Determining flux for two traffic states

The flux $\Phi(t)$ represents the actual number of vehicles that are passing one point in a certain amount of time. The flux cannot exceed the point capacity, or:

$$\Phi(t) \leq x(t)/\ell.$$

The actual flux $\Phi(t)$ on a road section determines the velocity $v(\Phi)$. For this purpose an experimental function is created to establish a relationship between the velocity and the flux. It should be noted that the traffic flow on a road can exists in two different regimes: the 'normal' free flow state, and the 'congested' forced flow state. The system can almost instantaneously jump from one state to the other. An article by Wahle et al. shows an example of such an occurrence (Wahle et al., 1999).

The maximum possible flux approaches the theoretical capacity of a road at the optimum velocity $\hat{v}$: $\max(\Phi(v)) = x / \ell$ or $\Phi(\hat{v}) = x / \ell$. According to literature, the optimum velocity lies commonly between the 50 km/h and 75 km/h, depending on the architecture of the road (Kreuzberger and Vleugel, 1992).

For the forced flow system, it is assumed that all the vehicles are queued. The vehicles in this case have to maintain a safe distance between them to avoid collision. This safe distance is $\Delta_{safe}=c_1+c_2 \cdot v+c_3 . v^2$. The flux is given by:

$$\Phi(v) = \frac{v \cdot w}{c_1 + c_2 v + c_3 v^2} \text{ for } v < \hat{v} . \tag{4.5a}$$

The variable w is the width of the road in the number of lanes. For low speeds, the velocity and the average distance between vehicles determine the flux. c_1 is the minimum distance between vehicles, set at $7.5 \cdot 10^{-3}$ km/veh (Wahle et al., 1999); c_2 is the reaction time of an individual driver, thus $c_2 \cdot v$ is the approximate safe distance to avoid collisions. c_3 is a higher order term, since the braking distance increases slightly more than linear with velocity. c_3 is determined by stating that $\mathrm{d}\Phi_{max} / \mathrm{d}v = 0$ at the turning point of $\hat{v}$=60 km/h; this implies that:

$c_3 = c_1 / \hat{v}^2 = \frac{7.5 \cdot 10^{-3}\text{ km}}{(60\text{ km/h})^2} = 2.08 \cdot 10^{-6}\text{ h}^2/\text{km}$. c_2 is set by the constraint on the highest possible flux of a single lane of $\Phi(\hat{v}) = 2000$ veh/h. It results in an average reaction time of $0.25 \cdot 10^{-3}$ h = 0.9 seconds, if w=1 and y/ℓ=2000 veh/h. See figure 10 for a graphical representation.

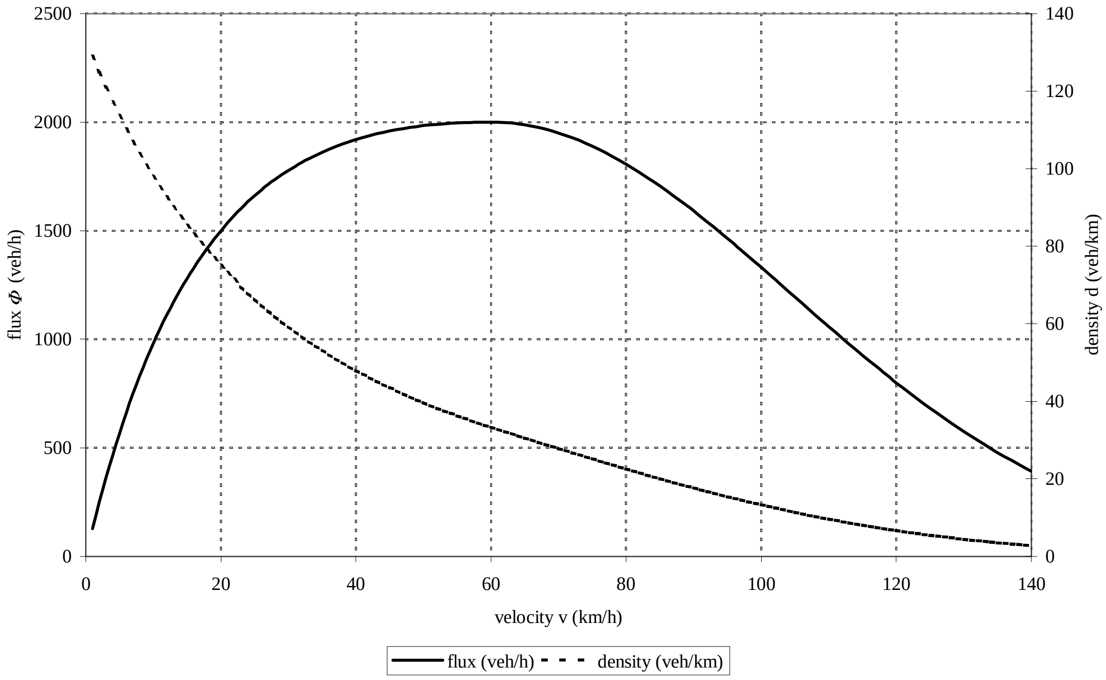

Figure 10 The solid line represents the highest possible flux at given velocity. The dashed line indicates the density that results from given combinations of flux and velocity.

The second part of the graph is part of a Gauss-curve. The idea is that the more vehicles are driving on a road, the likelier it is that they will interact resulting in reduction of velocity. So, it is chosen that this should be a probabilistic curve of form:

$$\Phi(v) = c_4 \cdot \mathrm{e}^{(v-c_5)^2/c_6} \text{ for } v \geq \hat{v}. \qquad (4.5b)$$

The parameters of this curve are $c_4=\Phi_{max}=x/\ell=2000$ veh/h (for a single lane road); $c_5=\hat{v}=60$ km/h and c_6 that performs a similar function as the deviation in the standard distribution curve. $c_6=3929$ by demanding that $\Phi_{max}(120)=800$ for a single lane road, or $\Phi_{max}(\tilde{v}) = \tilde{f} \cdot \frac{x}{\ell}$ with $\tilde{f} = \frac{2}{5}$ in general. These figures are valid for Dutch highways with a maximum speed allowed of 120 km/h.

4.4.3. Determing velocity

For the free flow system $y/\ell<\Phi(\hat{v})=2000$ veh/h, the velocity is – following equation (4.5b) – given by: $v = c_5 + \sqrt{\ln(\Phi/c_4)\cdot c_6}$. For high transport demands $y/\ell \geq 2000$ veh/h, it is assumed that all cars want to move at the optimum velocity of $\hat{v}=60$ km/h. Now the amount of vehicles per kilometer, or the density *d* (veh/km) is determining the velocity. A certain flux *Φ* implies, at 60 km/h, a necessary average density of cars of $d=\Phi/60$ (veh/km). Thus, higher throughputs imply higher densities. However, judging by figure 10, a certain density correlates to a specific velocity. As one can see, at a flux of more than the maximum of 2000 vehicles per hour, the velocity will drop below the optimum velocity. Therefore, the road will accommodate even less than 2000 vehicles. The remaining vehicles will either have to change route, or will be put on a ‘waiting list’. In this paper, it is assumed that they will be accommodated elsewhere on the network with the same relative energy efficiency. Let $d^*=d/(w \cdot \hat{v})$ denote the theoretical density. For $\Phi \geq 2000$ veh/h, the velocity is given by:

$$d^* = \frac{1}{c_1 + c_2 v + c_3 v^2}.$$

Since $d^* > 0$ it follows: $c_3 v^2 + c_2 v + c_1 - (d^*)^{-1} = 0$ and thus:

$$v = \frac{-c_2 \pm \sqrt{c_2{}^2 - 4c_3(c_1 - (d^*)^{-1})}}{2c_3}.$$

Since $(d^*)^{-1} > c_1$, or the cars cannot move closer together than the minimum distance, it goes that $c_1 - (d^*)^{-1} < 0$, therefore $\sqrt{c_2{}^2 - 4c_3(c_1 - (d^*)^{-1})} > c_2$, thus:

$$v(\Phi) = \begin{cases} c_5 + \sqrt{\ln(\Phi/c_4) \cdot c_6} & \text{for } 0 < \Phi < x/\lambda; \\ \dfrac{-c_2 + \sqrt{c_2{}^2 - 4c_3(c_1 - \frac{\hat{v}}{\Phi})}}{2c_3} & \text{for } x/\lambda \leq \Phi < \hat{v}/c_1. \end{cases}$$

It is now possible to rewrite some constants following the explanations of those constants as mentioned in section 4.4.2. Equation (4.6) shows the final formula to determine the velocity.

$$v(t) = \begin{cases} \hat{v} + \sqrt{(\tilde{v} - \hat{v})^2 \cdot \ln(\frac{y(t)}{x(t)}) \cdot \ln^{-1}(\tilde{f})} & \text{for } 0 < y < x; \\ \hat{v} + \dfrac{\ell w \hat{v}^2}{2c}\left(\sqrt{x(t)^{-2} + \dfrac{4c}{\hat{v}\ell w}\left(y(t)^{-1} - x(t)^{-1}\right)} - x(t)^{-1}\right) & \text{for } x \leq y < \dfrac{\hat{v}\ell w}{c}. \end{cases} \tag{4.6}$$

The parameters used in (4.6) with their default values in this paper:

c, $c>0$	Effective length of vehicle at rest	$7.5 \cdot 10^{-3}$	km·veh^{-1}
$\hat{v}$, $0<\hat{v}<\tilde{v}$	Optimum velocity (velocity with highest capacity)	60	km·h^{-1}
$\tilde{v}$, $\tilde{v}>\hat{v}$	Arbitrary velocity: $\Phi(\tilde{v}) = \tilde{f} \cdot x \cdot \ell^{-1}$	120	km·h^{-1}
$\tilde{f}$, $0<\tilde{f}<1$	Arbitrary fraction: $\Phi(\tilde{v}) = \tilde{f} \cdot x \cdot \ell^{-1}$	0.4	
$\ell>0$	Length of road (network)		km
$w \geq 1$	Width of road (network)		*(# lanes)*

4.3.2. Relation between energy consumption and velocity

The velocity *v* can be used to calculate the average fuel consumption of the traffic. The curve in figure 11 is a result of the equations of section 4.3.1, and shows the dependence of velocity on flux.

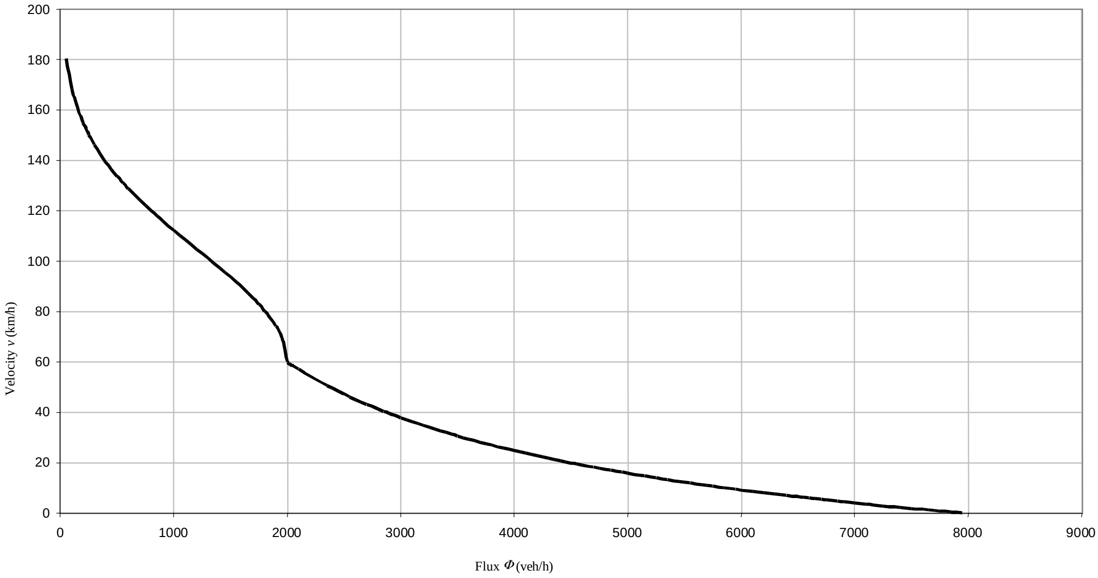

Figure 11 This curve shows the relation between flux Φ and velocity *v*. The discontinuity in the first derivative at Φ=2000 veh/h is where the system jumps from the free flow state to the forced flow (jam) state. In reality, there exists a range of fluxes in which there is a certain probability that the system will jump between the states. Velocities higher than legally permitted should be excluded of course.

As mentioned earlier, the actual flux at low speeds differs from the theoretical flux, since at low speeds not all throughputs can be accommodated. As said, an approximation is to assume that those throughputs beyond capacity will be accommodated at some other place without interfering with any other transport or traffic system, but with the same efficiency as the throughput that is accommodated. Thus, the energy use of the traffic *F(x(t),y(t))* would be given by the relative energy use *g(v)* and the traffic *y(t)*:

$$F(x(t),y(t))=g(v)\cdot y(t).$$

The formula of *g(v)* can be approximated in interpolating some empirical data of the actual energy consumption in congested traffic in the Netherlands (Veurman et al., 2002). The empirical data was divided into classes. Between the classes, the interpolation is made. See figure 12 for the resulting curve.

The curve in figure 12 is presented by $g(v)=-0.730\cdot(1-9.29\cdot e^{-0.0101v}-1.86\cdot 10^{-6}\cdot v^3)$. The third order is related to the aerodynamic resistance of vehicles at high speeds. In the low regions, a fit with an exponential curve is made. The total energy use of all the vehicles is thus:

$$F(x,y)=y(t)\cdot\{-0.730\cdot(1-9.29\cdot e^{-0.0101v(x,y)}-1.86\cdot 10^{-6}\cdot v(x,y)^3)\}. \quad (4.7)$$

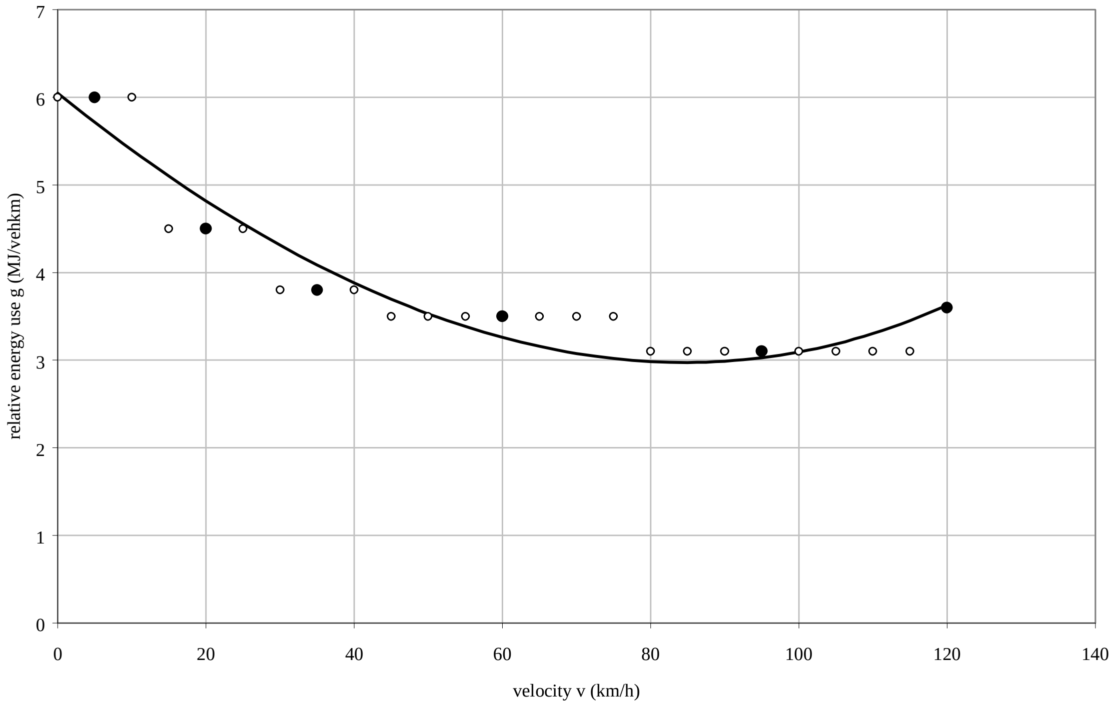

Figure 12 An interpolation is made of the empirical class data as presented by the solid dots. The open dots indicate the range of the classes. At the high end, velocities above 90 km/h, a third order power becomes dominant. The fit is made using the least squares method.

In the next graph, figure 13, the total energy consumption of traffic is illustrated.

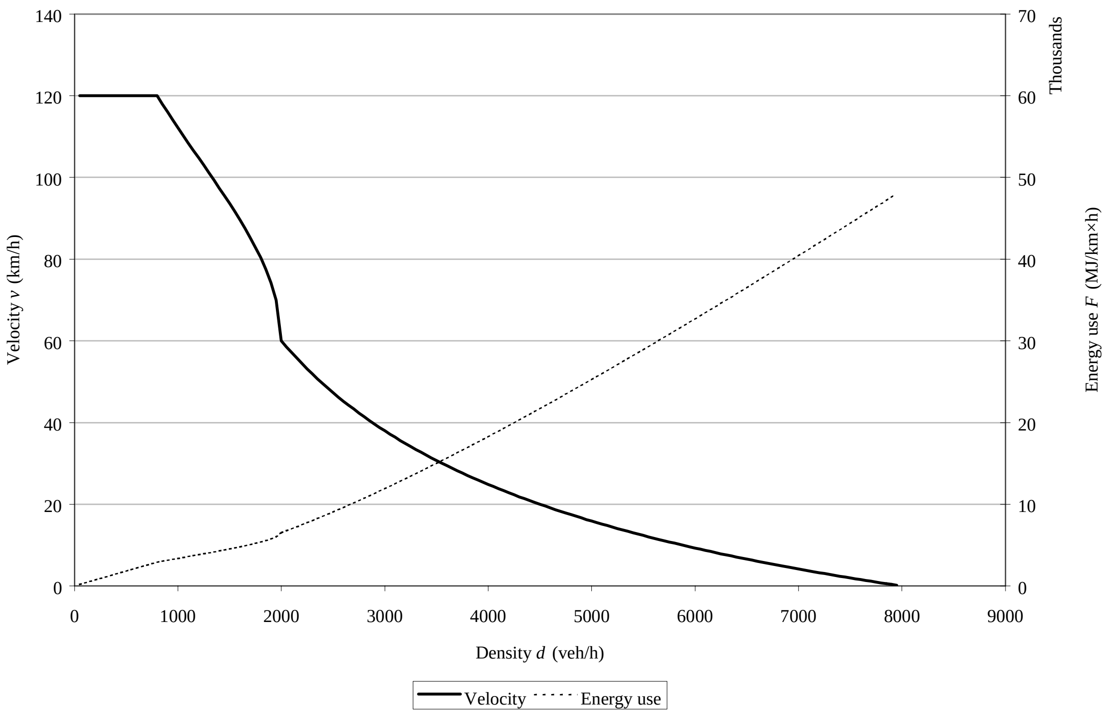

Figure 13 The final energy curve is presented by the dashed line; the energy use is shown in MJ per km·h. The velocity curve is the solid line. Note that the velocity is limited to 120 km/h.

4.3.3. Discount rate ρ

The objective function has a discount factor of $e^{-\rho t}$. This model chooses the discount rate such that ρ^{-1} the mean expected lifetime of a road is. Since it is not clear how the infrastructure will function after the end of the roads existence, there is reason to include a discount factor. It is a factor that is associated with uncertainty in time. Discount rates in an economic context describe "the inter-temporal preference structure of the economic agents." (Haurie, 2001). In this case, the discount rate ρ associated with uncertainty about the functionality of the infrastructure during the usage period, induces the discounting process. In the case of a discount factor of $e^{-\rho t}$, one could consider ρ a killing rate, since $\rho\, dt$ is the elementary probability that "death occurs" in the elementary time interval $[t,t+dt]$, given that the construction survived up to time t. A discount rate of 1.25% (t in years) corresponds to a random life duration with expected value 1/0.0125=80 years.

Only for later consideration, this paper suggests another type of discount factor. An alternative discount factor could be formulated on ground of allowed greenhouse gas emissions. It would be a factor that is associated with uncertainty in sustainable energy consumption. Considering that the transportation sector is fully dependent on fossil fuels throughout the present century, and considering that the emissions of greenhouse gases have to be reduced significantly in the same period, one could formulate a discount factor, which is based on the highest emission rate allowed. That would be of a form $\Delta = (1+e^{t^*-t})^{-1}$, with t* the time at which half of the emissions should be reduced. This would be a rising discount rate.

A proposed discount factor of $\Delta = e^{-\rho t}$ can be based on the speed of transition to clean energy in the transportation sector. The discount rate ρ would then typically be defined as the inverse of the time at which half of the transportation energy is clean energy. The combination of both factors: $\Delta = \frac{e^{-\rho t} \cdot r}{1+e^{t^*-t}}$, r being the reduction rate to be achieved at time t^*, would include both a reduction path for greenhouse gas emissions and a transition path to clean energy. The function meets the requirement of:

$$\lim_{t\to\infty} \Delta = \lim_{t\to\infty} \frac{e^{-\rho t} \cdot r}{1+e^{t^*-t}} = 0,\ \text{for all } \rho>0,\ 0<r<1,\ t^*>0.$$

Nevertheless, the current paper uses the discount rate of $e^{-\rho t}$ in which ρ corresponds with the average lifetime of the construction.

4.3.4. Summary

The objective function of the life cycle energy consumption of a specific road should be minimized. The final form of the function was deduced in the following steps:

$$J = \int_0^\infty \mathrm{e}^{-\rho t} (E(p(t)) + H(m(t), x(t)) + F(x(t), y(t)) dt \to \min;$$

$$J = \int_0^\infty \mathrm{e}^{-\rho t} (\alpha \cdot p(t) + \beta \cdot m(t) \cdot x(t) + F(x(t), y(t)) dt \to \min;$$

$$J = \int_0^\infty \mathrm{e}^{-\rho t} (\alpha \cdot u(t) + F(x(t), y(t)) dt \to \min.$$

The energy consumption of the vehicles is described by:

$$F(x(t),y(t))=y(t)\cdot\{-0.730\cdot(1-9.29\cdot \mathrm{e}^{-0.0101v(x,y)}-1.86\cdot 10^{-6}\cdot v(x,y)^3)\}$$

in which the velocity is dependent on the traffic system being in forced-flow or free-flow mode:

$$v(x(t), y(t)) = \begin{cases} \hat{v} + \sqrt{(\tilde{v} - \hat{v})^2 \cdot \ln(\frac{y(t)}{x(t)}) \cdot \ln^{-1}(\tilde{f})} & \text{for } 0 < y < x; \\ \hat{v} + \dfrac{\ell w \hat{v}^2}{2c} \left(\sqrt{x(t)^{-2} + \dfrac{4c}{\hat{v}\ell w}\left(y(t)^{-1} - x(t)^{-1}\right)} - x(t)^{-1} \right) & \text{for } x \le y < \dfrac{\hat{v}\ell w}{c}. \end{cases}$$

One aims to minimize energy use. The minimization of the objective function J leads to a minimization of the energy use of one road during lifetime. Consider now the superscript j that indicates that the variables are related to road j. For a road system, or infrastructure consisting of N roads, thus $1 \le j \le N$, the objective becomes:

Minimize ($J = \int_{t=0}^{\infty} e^{-\rho t} \cdot \sum_{j=1}^{N} \left[\alpha^j u(t) + F(^j x(t), ^j y(t))\right] \mathrm{d}t$).

Due to possible interrelations between $^i x(t)$ and $^j y(t)$ for all $0<i\le N$, $0<j\le N$, the optimal value of the objective function J for the whole system is not necessary the same as the sum of optimal values of objective functionals,

$$J_j\left(x(t), u(t)\right) = \int_0^\infty e^{-\rho t}\left[\alpha_j u + F(^i x(t), ^j y(t))\right] dt \to \min ,$$

minimized independently.

For practical purposes, this project looks at one specific road, while any interaction with the surrounding road system is neglected. The road should be long enough to be able to regard production and maintenance energies E and H as continuous functions in time.

5. Optimal control methodology

5.1. Optimal control problem

Consider the following optimal control problem (P):

$$\dot{x} = u - \delta \cdot x \quad , \quad u \in [0, a]; \tag{5.1}$$

$$x(0) = x_0; \tag{5.2}$$

$$J\left(x(t), u(t)\right) = \int_0^\infty e^{-\rho t}\left[\alpha u + F(x,t)\right] dt \to \min. \tag{5.3}$$

Here $x \in R^1$, $u \in R^1$, $\delta > 0$, $a > 0$, $x_0 > 0$, $\rho > 0$, $\alpha > 0$; $F(x,t)$ is a continuously differentiable function on $[0,\infty) \times [0,\infty)$; The class of admissible controls consists of all piecewise continuous[4] functions $u(t)$ for all $t \in [0,\infty)$ with $u(t) \in [0,a]$. The trajectory corresponding to the admissible control $u(t)$ is a piecewise continuously differentiable solution $x(t)$ of (5.1) with initial condition (5.2). $u(t)$, $x(t)$ is an admissible pair. The function $F(x,t)$ must meet several assumptions:

(H1) $\left|F(x,t)\right| \le \kappa_1$ for all $x \in [0,\infty)$ and all $t \in [0,\infty)$;

(H2) $\dfrac{\partial F(x,t)}{\partial x} < 0$ for all $x \in [0,\infty)$ and all $t \in [0,\infty)$;

(H3) $\left|\dfrac{\partial F(x_*,t)}{\partial x}\right| < \kappa_2$ for all $t \in [0,\infty)$;

(H4) $\left|\dfrac{\partial F}{\partial x}(0,0)\right| > \alpha(\rho + \delta)$;

(H5) $\dfrac{\partial^2 F}{\partial x^2}(x,t) > 0$ for all $x \in [0,\infty)$ and all $t \in [0,\infty)$;

(H6) $\lim\limits_{x \to \infty} \dfrac{\partial F(x,t)}{\partial x} = 0$ for all $t \in [0,\infty)$.

Physical parallels of these assumptions are the next: (H1) tells that the total energy consumption of the traffic cannot exceed a certain maximum. (H2) says that only too small capacities are energetically inefficient, but too large capacities do not cause greater fuel consumptions. In reality, large capacities will lead to higher and thus more inefficient velocities and might cause the real function to rise again slightly at large capacities. As for (H5), as velocities are already at their peak value – and the velocities cannot increase further – there is no change in fuel consumption if the capacity is already much greater than necessary.

---

[4] Piecewise continuous functions have a limited number of discontinuities on each finite time interval $[0,T]$, $T > 0$. It is assumed that in each point of discontinuity, the function is continuous from the left.

It is easy to see that due to the system state equation (5.1) and the assumption $x_0>0$, the state variable *x(t)* is strictly positive and bounded for all $t \geq 0$. Thus there exists a constant $b > 0$ such that $0 < x(t) < b$ for all $t \geq 0$.

Now, it is proven that in a problem like problem (P), there exists an optimal control $u_*(t)$. See theorem 3.6 in (Balder, 1983).

5.2. Pontryagin's maximum principle

The main tool in the study of problem (P) involves looking for the necessary optimality conditions in the form of the Pontryagin maximum principle (Pontryagin et al., 1962). The theory involves two closely related functions:

$\mathcal{H}(t,x,u,\Psi) = (u - \delta x)\Psi - e^{-\rho t}\left[\alpha u + F(x,t)\right]$ is the Hamilton-Pontryagin function.

$H(t,x,\Psi) = \max\limits_{u\in[0,a]} \mathcal{H}(t,x,u,\Psi)$ is the Hamiltonian.

The next result is a specialized version of the Pontryagin maximum principle for problem (P).

Theorem 1. *Let $u_*(t)$, $x_*(t)$ be an optimal pair in (P). Then there exists an adjoint function $\tilde{\Psi}(t)$, with the following properties:*

1) $\dot{\tilde{\Psi}} = \delta \cdot \tilde{\Psi} + e^{-\rho t}\dfrac{\partial F}{\partial x}(x_*(t),t);$

2) $H(t,x_*(t),\tilde{\Psi}(t)) = \mathcal{H}(t,x_*(t),u_*(t),\tilde{\Psi}(t));$

3) $\tilde{\Psi}(t) > 0$ *for all* $t\in[0,\infty)$;

4) $\tilde{\Psi}(t) \to 0$ *as* $t \to \infty$.

The proof of this result is similar to the proof of Theorem 2 from (Aseev et al., 2001) and Theorem 3.2 from (Aseev et al, 2002).

Corollary. $x(t)\tilde{\Psi}(t) \to 0$ as $t \to \infty$.

Proof. As *x(t)* is bounded, there is a constant *b*: $b > 0$, such that $x(t)<b$ for all $t>0$. Hence, we have $x(t)\tilde{\Psi}(t) < b\tilde{\Psi}(t)$ for all $t>0$. According to condition 4) of Theorem 1, we can write:

$$\lim_{t\to\infty} x(t)\tilde{\Psi}(t) \leq \lim_{t\to\infty} b\tilde{\Psi}(t) = b\lim_{t\to\infty}\tilde{\Psi}(t) = 0.$$

■

Corollary. *There exists a constant $\kappa_3>0$ such that* $x_*(t)\tilde{\Psi}(t) \leq \kappa_3 e^{-\rho t}$ *for all* $t\in[0,\infty)$.

Proof. To find an upper bound for $x_*(t)\tilde{\Psi}(t)$ one differentiates, remembering the state equation $\dot{x}(t) = -\delta \cdot x(t) + u(t)$ and using condition 1) of Theorem 1:

$$\frac{d}{dt}\left[x_*(t)\tilde{\Psi}(t)\right] = -\delta x_*(t)\tilde{\Psi}(t) + u_*(t)\tilde{\Psi}(t) + \delta x_*(t)\tilde{\Psi}(t) + x_*(t)e^{-\rho t}\frac{\partial F}{\partial x}(x_*(t),t) =$$
$$= u_*(t)\tilde{\Psi}(t) + x_*(t)e^{-\rho t}\frac{\partial F}{\partial x}(x_*(t),t).$$

Since $u_*(t) \geq 0$ and $\tilde{\Psi}(t) \geq 0$, it follows that $u_*(t)\cdot\tilde{\Psi}(t) \geq 0$;

The combination of these equations and condition (H2) gives:

$$\frac{d}{dt}\left[x_*(t)\tilde{\Psi}(t)\right] \geq F'(x_*(t))x_*(t)e^{-\rho t} \text{ on } [t,T]:$$
$$x_*(T)\tilde{\Psi}(T) - x_*(t)\tilde{\Psi}(t) \geq \int_t^T F'(x_*(s))x_*(s)e^{-\rho s}ds \geq \kappa_3\left(e^{-\rho T} - e^{-\rho t}\right)$$
$$x_*(t)\tilde{\Psi}(t) \leq x_*(T)\tilde{\Psi}(T) + \kappa_3 e^{-\rho t} - \kappa_3 e^{-\rho T}, \text{ where } \kappa_3 = \kappa_2 b/\rho.$$

Now for $T \to \infty$ and remembering the transversality condition 4), the upper bound is:

$$x_*(t)\tilde{\Psi}(t) \leq \kappa_3 e^{-\rho t} \text{ for all } t \in [0,\infty).$$

■

Let us introduce a new adjoint variable: $\Psi(t) = e^{\rho t}\tilde{\Psi}(t)$.

Then the following result is a reformulation of Theorem 1 in terms of $\Psi(t)$.

Theorem 2. *Let $u_*(t)$, $x_*(t)$ be an optimal pair in (P). Then there exists an adjoint function $\Psi(t)$ such that:*

1) *$\Psi(t)$ is a solution of the adjoint system* $\dot{\Psi} = (\rho+\delta)\Psi + \frac{\partial F}{\partial x}(x_*(t),t)$;
2) $u_*(t)\cdot\left(\Psi(t)-\alpha\right) = \max\limits_{u\in[0,a]} u\cdot\left(\Psi(t)-\alpha\right)$;
3) $\Psi(t) > 0$ *for all $t \in [0,\infty)$;*
4) $\Psi(t)x_*(t) \leq \kappa_3$ *for all $t \in [0,\infty)$.*

5.3. Construction of the associated Hamiltonian system

We consider (starting from this place) only the case of $F(x,t) \equiv F(x)$. From the maximum condition 2) of Theorem 2, we have:

$$u_*(t) = \begin{cases} 0 & \text{, if } \Psi(t) < \alpha; \\ a & \text{, if } \Psi(t) > \alpha; \end{cases}$$

$u_*(t) \in [0,a]$, if $\Psi(t) = \alpha$.

Let us introduce sets G, G_0, G_1, G_2 in R^2.

$G = [0,\infty)\times[0,\infty);$
$G_0 = [0,\infty)\times[0,\alpha);$
$G_1 = [0,\infty)\times(\alpha,\infty);$
$G_2 = [0,\infty)\times\alpha.$

Obviously $G = G_0 \cup G_1 \cup G_2$.

Let us introduce a multi-valued function *r(x, Ψ)* and a scalar function *s(x, Ψ)* by the following way:

$$r(x,\Psi) = \begin{cases} -\delta\cdot x & \text{, if } (x,\Psi)\in G_0; \\ a-\delta\cdot x & \text{, if } (x,\Psi)\in G_1; \\ [0,a]-\delta\cdot x & \text{, if } (x,\Psi)\in G_2; \end{cases}$$

$$s(x,\Psi) = (\rho+\delta)\Psi + \frac{\partial F}{\partial x}(x).$$

Lemma. *Let $u_*(t)$, $x_*(t)$ be an optimal pair in problem (P) and Ψ(t) be an adjoint function corresponding to $u_*(t)$, $x_*(t)$ due to Theorem 2. Then $x_*(t)$, Ψ(t) solve the following Hamiltonian system by inclusion:*

$$\dot{x} \in r(x,\Psi); \tag{5.4}$$

$$\dot{\Psi} = s(x,\Psi). \tag{5.5}$$

Moreover, $\Psi(t) > 0$ for all $t\in[0,\infty)$ and $\Psi(t)x_*(t) \le \kappa_3$.

Proof. Indeed due to Theorem 2:

$$\dot{\Psi}(t) = (\rho+\delta)\Psi(t) + \frac{\partial F}{\partial x}\big(x_*(t)\big)$$

and

$\dot{x}_*(t) = -\delta\cdot x_*(t)$ if $\big(x_*(t),\Psi(t)\big)\in G_0$;
$\dot{x}_*(t) = a-\delta\cdot x_*(t)$ if $\big(x_*(t),\Psi(t)\big)\in G_1$
and $\dot{x}_*(t)\in[0,a]-\delta\cdot x_*(t)$ if $\Psi(t)=\alpha$ (i.e. if $\big(x_*(t),\Psi(t)\big)\in G_2$).

Hence $x_*(t)$, Ψ(t) is a solution of the Hamiltonian system (5.4), (5.5).

Due to Theorem 2 $\Psi(t) > 0$ for all $t\in[0,\infty)$ and $\Psi(t)x_*(t) \le \kappa_3$.

■

Remark. Obviously the Hamiltonian system (4),(5) is single-valued in G_0 and G_1. In G_0 it takes the form:

$$\begin{cases} \dot{x} = -\delta\cdot x; \\ \dot{\Psi} = (\rho+\delta)\Psi + \dfrac{\partial F}{\partial x}(x). \end{cases}$$

In G_1 it takes the form:

$$\begin{cases} \dot{x} = a - \delta \cdot x; \\ \dot{\Psi} = (\rho + \delta)\Psi + \dfrac{\partial F}{\partial x}(x). \end{cases}$$

Consider the set G_2={ (x,α) : $x > 0$ } and introduce a function $\gamma(x)$:

$$\gamma(x) = -\frac{1}{\rho + \delta}\frac{\partial F}{\partial x}(x). \tag{5.6}$$

Due to assumptions (H1)-(H4) $\gamma(x)$ is a decreasing function and

$\gamma(0) > \alpha;$

$\gamma(x) \to 0 \text{ if } x \to \infty.$

Hence there exists a unique x^0 for which $\gamma(x^0) = \alpha$ i.e. $(x^0, \alpha) \in G_2$.

5.4. Necessary and sufficient conditions for optimality

The following result is the Arrow (Arrow and Kurz ,1970) type sufficient condition of optimality.

Theorem 3. *Let $u_*(t)$, $x_*(t)$ be an admissible pair satisfying to the condition of Theorem 2 with an adjoint function $\Psi(t)$. Then the pair $u_*(t)$, $x_*(t)$ is optimal.*

Proof. Due to the convexity of $F(x)$, see assumption (H5), the following inequality takes place for arbitrary $t \geq 0$:

$$\frac{\partial F}{\partial x}(x_*(t)) \cdot (x_*(t) - x(t)) \geq F(x_*(t)) - F(x(t)),$$

where $u(t)$,$x(t)$ is an arbitrary admissible pair.

Indeed, for arbitrary x_* and x, and all $0 < \eta < 1$, we have:

$$F(x_* + \eta(x - x_*)) = F(\eta x + (1-\eta)x_*) \leq \eta F(x) + (1-\eta)F(x_*).$$

Hence,

$$\frac{F(x_* + \eta(x - x_*)) - F(x_*)}{\eta} \leq F(x) - F(x_*).$$

If we let $\eta \to 0$, we derive

$$\frac{\partial F}{\partial x}(x_*) \cdot (x_* - x) \geq F(x_*) - F(x).$$

Due to condition 1) of Theorem 2, we have

$$\frac{\partial F}{\partial x}(x_*(t)) = \dot{\Psi}(t) - (\rho+\delta)\Psi(t).$$

Hence, $\dot{\Psi}(t)(x_*(t)-x(t)) - (\rho+\delta)\Psi(t)(x_*(t)-x(t)) \geq F(x_*(t)) - F(x(t)).$

Due to the system dynamics (5.1), we have

$$\dot{x}_*(t) = u_*(t) - \delta x_*(t) \Rightarrow x_*(t) = \frac{u_*(t)-\dot{x}(t)}{\delta}.$$

Analogously $x(t) = \dfrac{u(t)-\dot{x}(t)}{\delta}$.

Now we can write the inequality as

$$\dot{\Psi}(t)(x_*(t)-x(t)) - \rho\Psi(t)(x_*(t)-x(t)) - \Psi(t)(u_*(t)-u(t) + \Psi(t)(\dot{x}_*(t)-\dot{x}(t)) \geq$$
$$\geq F(x_*(t)) - F(x(t)).$$

Due to the maximum condition 2) of Theorem 2 we have

$$u_*(t)\Psi(t) - u_*(t)\alpha \geq u(t)\Psi(t) - u(t)\alpha.$$

Hence, $(u_*(t)-u(t))\Psi(t) \geq \alpha(u_*(t)-u(t)).$

Hence, $\dot{\Psi}(t)(x_*(t)-x(t)) + \Psi(t)(\dot{x}_*(t)-\dot{x}(t)) - \rho\Psi(t)(x_*(t)-x(t)) - \Psi(t)(u_*(t)-u(t)) \leq$

$$\leq \frac{d}{dt}\left[\Psi(t)(x_*(t)-x(t))\right] - \rho\Psi(t)(x_*(t)-x(t)) - \alpha(u_*(t)-u(t)).$$

Hence, $\dfrac{d}{dt}\left[\Psi(t)(x_*(t)-x(t))\right] - \rho\Psi(t)(x_*(t)-x(t)) \geq$

$$\alpha(u_*(t)-u(t) + F(x_*(t) - F(x(t)).$$

Hence, $e^{-\rho t}\left[\Psi(t)(x_*(t)-x(t))\right] - \rho e^{-\rho t}\Psi(t)(x_*(t)-x(t)) \geq$

$$\geq e^{-\rho t}\left[\alpha u_*(t) + F(x_*(t))\right] - e^{-\rho t}\left[\alpha u(t) + F(x(t))\right].$$

Hence, $\dfrac{d}{dt}\left[e^{-\rho t}\Psi(t)(x_*(t)-x(t))\right] + e^{-\rho t}\left[\alpha u(t) + F(x(t))\right] \geq$

$$\geq e^{-\rho t}\left[\alpha u_*(t) + F(x_*(t))\right].$$

By integrating on the time interval [0, *t*], we have

$$e^{-\rho t}\Psi(t)\left[x_*(t)-x(t)\right] - \Psi(0)\left[x_*(0)-x(0)\right] + \int_0^t e^{-\rho s}\left[\alpha u(s) + F(x(s))\right]ds \geq$$

$$\geq \int_0^t e^{-\rho s}\left[\alpha u_*(s) + F(x_*(s))\right]ds.$$

Condition 3) of Theorem 2 tells that for all $t>0$: $\Psi(t)x(t)>0$. The inequality can therefore be written as

$$e^{-\rho t}\Psi(t)x_*(t)-e^{-\rho t}\Psi(t)x(t)+\int_0^t e^{-\rho s}\left[\alpha u(s)+F(x(s))\right]ds \geq \int_0^t e^{-\rho s}\left[\alpha u_*(s)+F(x(s))\right]ds.$$

Hence, $e^{-\rho t}\Psi(t)x_*(t)+\int_0^t e^{-\rho s}\left[\alpha u(s)+F(x(s))\right]ds \geq \int_0^t e^{-\rho s}\left[\alpha u_*(s)+F(x(s))\right]ds.$

From condition 4) of Theorem 2 follows:

$e^{-\rho t}\Psi(t)x_*(t) \leq e^{-\rho t}\kappa_3 \to 0$ for $t \to \infty$

Taking the integrals now over infinite time, we have:

$$\int_0^\infty e^{-\rho s}\left[\alpha u(s)+F(x(s))\right]ds \geq \int_0^\infty e^{-\rho s}\left[\alpha u_*(s)+F(x_*(s))\right]ds;$$

$$J(x(t),u(t)) \geq J(x_*(t),u_*(t))$$

The admissible pair $x_*(t),u_*(t)$ is therefore the optimal pair that minimizes the objective function. ■

As a consequence of Theorems 2 and 3 we have the following

Corollary. *The maximum principle (Theorem 2) is a necessary and sufficient condition of optimality in problem (P).*

6. Results

6.1. Stable curves in Hamiltonian system

Formula (5.6) described the form of the function *γ(x)* in terms of the first derivative of *F(x)*. The theoretical formula for *F(x)* that was deduced in section 4 does, in the form as it is presented in section 4.3.4., not meet the criteria for *F(x)* that are laid down in section 5. The first derivative of *F(x)* shows a discontinuity at *x=y* that has to be resolved first, see figure 14.

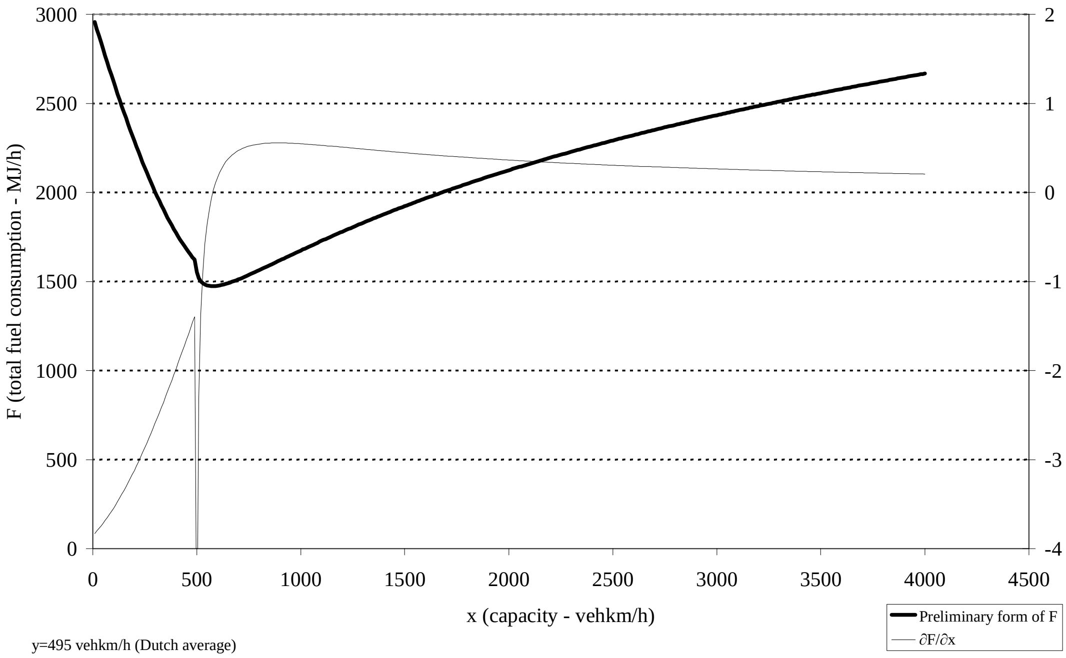

Figure 14 Example of the curve of the total energy use for fuel consumption *F(x)*. The thick black line shows the main function *F(x)*, the thin line of *F'(x)* exhibits a spike at *x=y=495*.[5]

To resolve the problem of the spike in *F'(x)* at *x=y* – or at the point where the system jumps between the free flow and the forced flow regime – let us compare the curve of the free flow system (at $x \geq y$) with the theoretical curve of the forced flow (for all *x*). In the interval $495 \leq x \leq 1050$ (vehkm/h), the change in curves is no more than 5.5% for *y*=495. Due to the low thus induced error, the curves are interchangeable, compare with figure 15.

---

[5] The total area of Dutch state roads (67.75 km^2) and their total length of 5678 km give an average width of 11.9 m (one-directional road for freeways, bi-directional for other state roads). An assumed average lane width of 3.5 m (also valid for the shoulder), gives and average number of lanes of 2.4. Therefore, the Netherlands have approximately $13.7 \cdot 10^3$ km of single lanes. The total transportation of $5.15 \cdot 10^6$ vehkm/h gives an average use of 376 vehkm/h. An expected rise in transportation of 30% (see figure 3) gives a final lane occupancy of 495 vehkm/h. This figure is only used for illustrative purposes. No sensitivity analysis is carried out. (MinV&W, 1999; CBS, 1995)

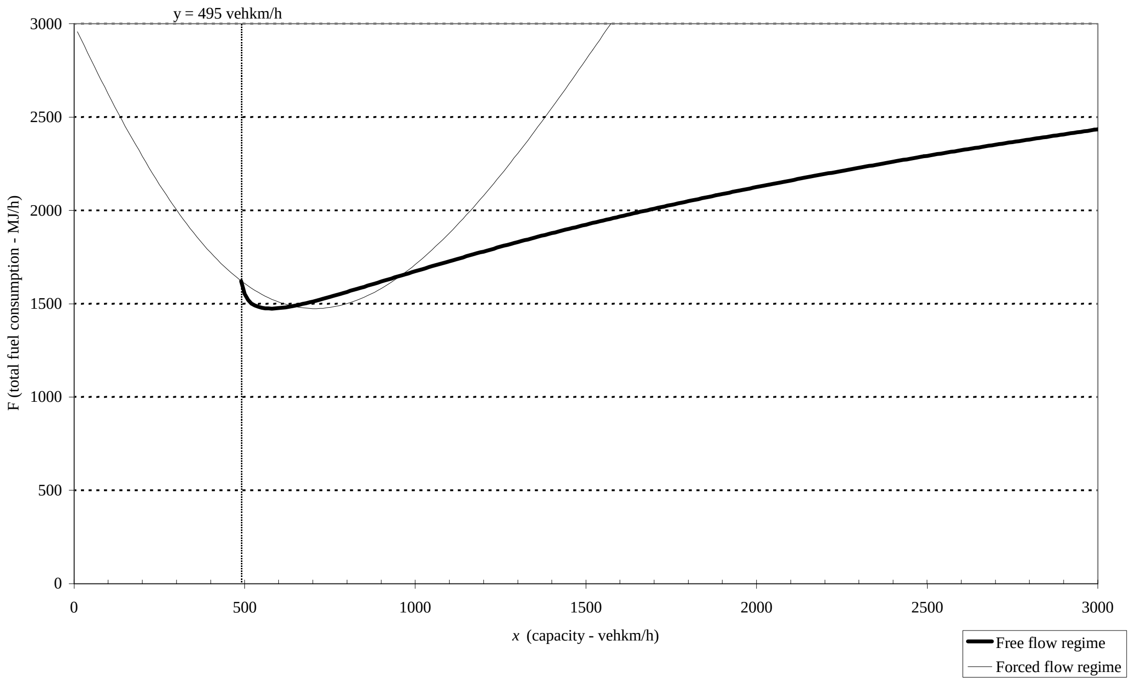

Figure 15 The difference between both curves is relatively small in the region between 495 and 1050 vehkm/h (≤5.5%). The first attempt is to use from the left the (thin line) forced flow curve until the point of equal and positive derivatives. From that point on, the (thick line) free flow curve is used.

Consider the two curves of figure 15. Denote F_{free} as the free flow curve, defined for $x>y$ and denote F_{forced} the forced flow curve, defined for all x. The point x_{jump} is the 'jump point': $F=F_{forced}$ for $x \leq x_{jump}$ and $F=F_{free}$ for $x>x_{jump}$. Suppose this jump point is defined as: $\frac{\partial F}{\partial x}(x_{jump}) > 0$ and $\frac{\partial F_{forced}}{\partial x}(x_{jump}) = \frac{\partial F_{free}}{\partial x}(x_{jump})$, then the function $F(x)$ will have a form as depicted in figure 16.

To prevent any conflict with condition (H2), also positive derivatives have to be excluded. Therefore, in regions with a thus far positive slope, the function is kept almost constant at the lowest level. It cannot be fixed at a constant level, as this is excluded by assumption (H5). Suppose a capacity level x_{min} where the minimum of F in x lies:

$\frac{\partial F}{\partial x}(x_{\min}) = 0$. This point, as can be seen in figure 16, is defined as following the formula of F_{forced} in above rules. The final form of F is now defined as:

$$F(x) = \begin{cases} F_{forced}(x) \text{ for } x \leq x_{\min}; \\ F_{forced}(x_{\min}) + \varepsilon(x) \text{ for } x > x_{\min}. \end{cases}$$

The very small and positive function $\varepsilon(x)$ meets the assumption (H2), (H3) and (H5) on F concerning the derivatives. The version of F used in the final part of this paper, does not always meet all the conditions for $x<x_0$. It can be proven that this does not compromise the result.

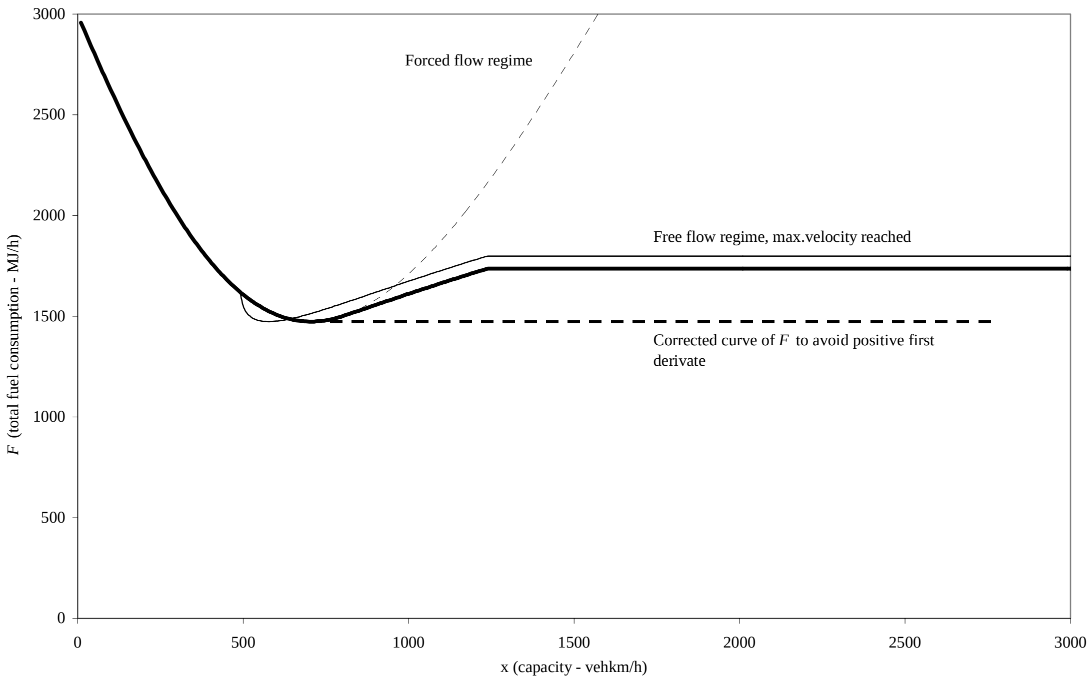

Figure 16 Schematic overview of construction of the final form of *F(x)*. Note that in this graph the option of velocities greater than 120 km/h is disallowed. The approximation to avoid positive derivatives is indicated by the dotted line.

6.2. Defining regions with different control

Three different sets, linked to three regimes, were defined in section 5.3: G_0, G_1 and G_2. Figure 17 shows the curve $\gamma(x)$, the stable line in Ψ, a arbitrary vertical line at $x=a/\delta$ representing the highest maintainable capacity level possible, and a horizontal line at $\Psi=\alpha$, where the control variable can take values in the range [0, *a*]. Above the line $\Psi=\alpha$, the regime *V* connects to set G_1. Below this line, thus $\Psi < \alpha$, the regime *W* links to set G_0.

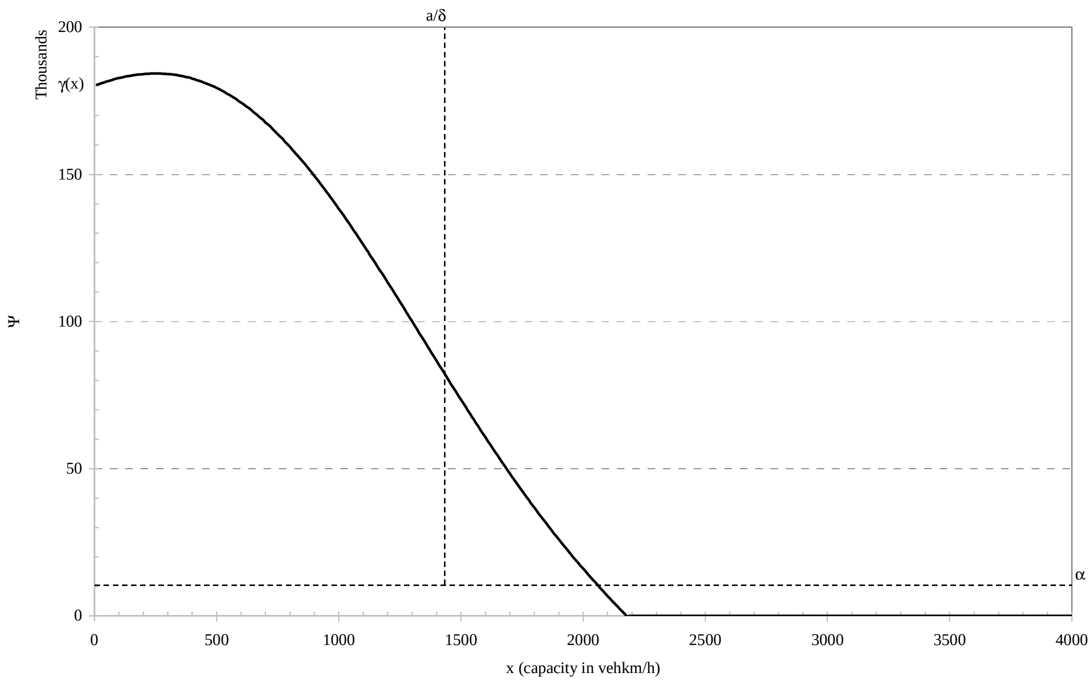

Figure 17 Graph representing the stable lines in the vector diagram. The thick black curve is $\gamma(x)$, along which $d\Psi/dt = 0$. This graph is created for y=1500 vehkm/h.

Inside this chart, it is possible to indicate the vector field of the Hamiltonian system by the direction of the vectors. The notation has the form: $<regime>^{<change\ in\ x><change\ in\ \Psi>}$. For example, the vector field in the top right corner (high x, high Ψ) can be indicated by V^{-+}, since $\dot{x} < 0$ and $\dot{\Psi} > 0$. Figures 18 and 19 use this notation.

6.3. Optimal control values

The solution to finding optimal control values follows from Theorem 2. The final optimal steady state can be deduced used the theory of section 5.3. The rest points in the Hamiltonian system can exist at the intersections of either $\gamma(x)$ and the line $\Psi=a$, or $\gamma(x)$ and the line $x=a/\delta$. Which intersection is representative of the equilibrium situation depends on the value of the maximum construction effort a. Figure 18 shows a case with limited resource (a/δ=1500) and figure 19 shows one with sufficient resource (a/δ=3000). In both figures the optimal transition paths towards these rest points are depicted by back casting, using the rest points as starting points.

As example, let us look at figure 18. One of an infinite number of possible trajectories starts at (x_0=10, $\Psi(0)=100\ 000$). Initially, this system moves in figure 18 to the right, thus with an increase in x. At the same time, it also shows a tendency to move downwards. This behaviour continues until the system reaches the line $x=\alpha$, from where x will be decreasing. Shortly after, the system drops below the line $\Psi=0$, and the system will violate condition 3) of Theorem 2. Therefore, a trajetory starting at (x_0=10, $\Psi(0)=100\ 000$) is not an optimal trajectory. Also, the paths starting at large x (e.g. x_0>2000) and large Ψ (e.g. $\Psi(0)$>100 000) would rise sharply, and thus conflict with condition 4) of Theorem 2. Therefore, those paths are also not optimal.

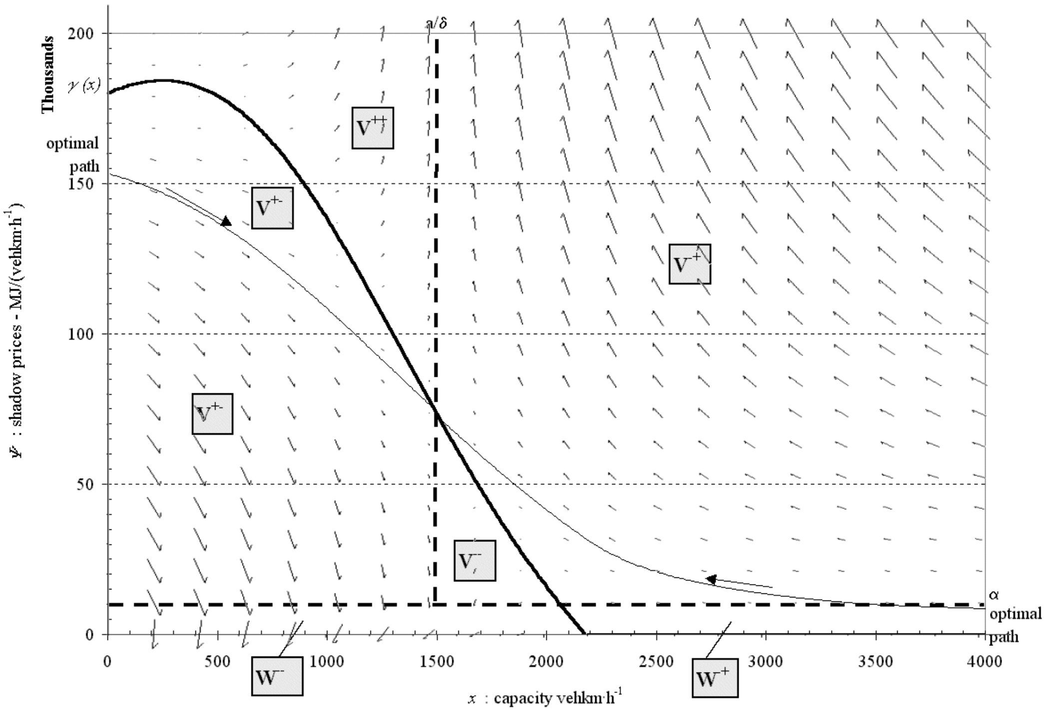

Figure 18 Hamiltonian system of the low resource case. The rest point is at *x=a/δ*. The vector field shows that the intersection of *Ψ=α* is not a stable point, since at that point *dx/dt*<0.

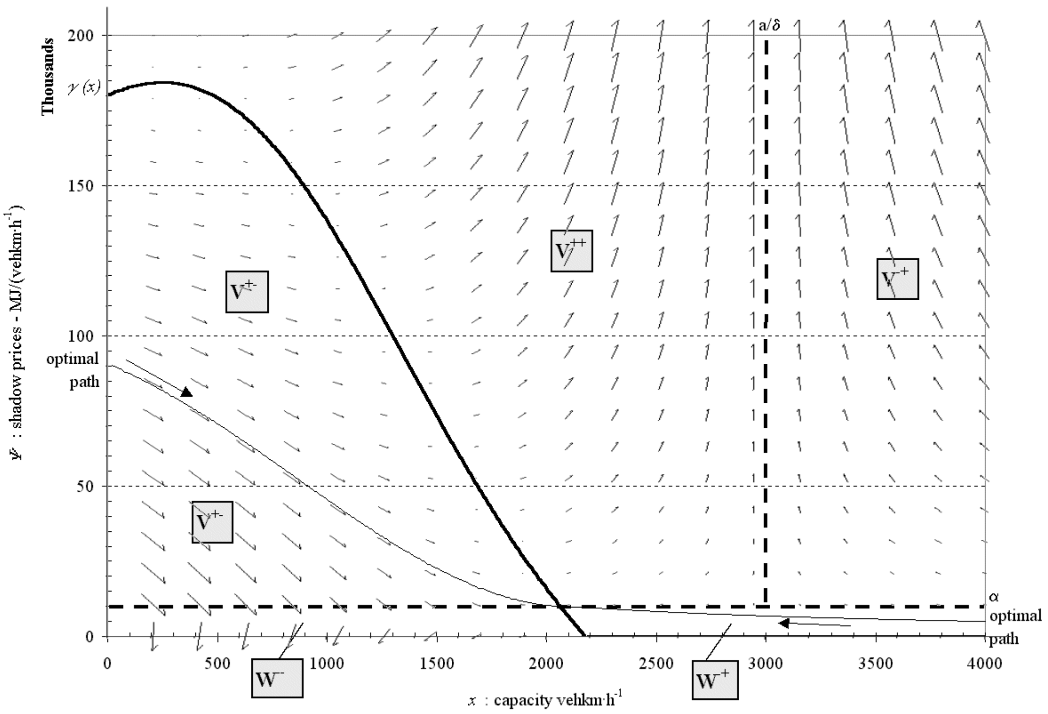

Figure 19 The Hamiltonian system for a case with sufficient resources. The rest point is at *γ(x)=α*. The capacity x^0 of this point can be maintained since x^0 < *a/δ*. As is proven in chapter 5, the transition path is unique.

6.4. Case of average Dutch highway

The current expenditure on road infrastructure management is taken as starting point. In 2000, $1.0 \cdot 10^9$ /year was spent on production of new state road capacity, while $0.5 \cdot 10^9$ /year on state road maintenance. The conversion rate from monetary values to energy, the energy intensity *I*, is calculated for the Dutch construction sector by (Kok et al., 2001): $I_{road} = 9.06$ MJ/ . The energy expenditure for the Dutch state roads was:

$\alpha \cdot p = 9.06 \cdot 10^3$ TJ/yr;

$ß \cdot m \cdot x = 4.53 \cdot 10^3$ TJ/yr.

According to Bos (Bos, 1998), the total indirect energy costs on freeways amount to 90 TJ/km, of which 26 TJ/km are used for 50 years of maintenance. The 64 TJ/km are for the production. Since a default freeway has 2x2 lanes, one kilometer of it has therefore a capacity of approximately $x = 8000$ vehkm/h. The relative energy costs for production is therefore $\alpha = 8000$ MJ/(vehkm·h^{-1}).

The 26 TJ/km are for 50 years maintenance; it implies an life time average energy expenditure of 0.52 TJ/yr, or $ß \cdot m \cdot x = 0.52$ TJ/yr; assuming $m = \delta = 0.15$ yr^{-1}, and remembering $x = 8000$ vehkm/h, it is possible to resolve the value of *ß*: $ß = 433$ MJ/(vehkm·h^{-1}). This still defies the assumption of section 4.3, since $\alpha \neq ß$. However, that might still be because high energy requirements for maintenance at roads of high age are not included in the current maintenance expenditures (basically, because the roads are not old enough). Nevertheless, these values of *α* and *ß* can be used for the calculation of the maximum construction effort at current expenditure levels.

The value of the production effort is $p = 9.06 \cdot 10^9 / 8000$ vehkm/(yr·h) = $1.13 \cdot 10^6$ vehkm/(yr·h). A default freeway has, as said, a capacity of $8.0 \cdot 10^3$ vehkm/h. Therefore, the production rate is comparable to 142 km of new freeway per year. Figure 4 showed an increase in total length of roughly 20 km/year; therefore most of this production is used for broadening of existing roads.

Similarly, the maintenance effort is $m \cdot x = 4.53 \cdot 10^9$ / 433 vehkm/(yr·h) = $10.5 \cdot 10^6$ vehkm/(yr·h). The current production effort is thus $u = p + m \cdot x = 11.6 \cdot 10^6$ vehkm/(yr·h). With $x = u/\delta$, we get: $x = 77.3 \cdot 10^6$ vehkm/h. The current length of the network is $13.7 \cdot 10^3$ km(lane). Very roughly, for one kilometer of lane the maximum construction effort at current expenditure levels is around the a/δ = 5600 vehkm/h.

Finally, one has to determine the transport demand of the average Dutch highway. This has been done and – including an expected rise of 30% - estimated at 495 vehkm/h. In reality, the transport demand shows daily recurring fluctuations. Figure 20 shows two curves of $\gamma(x)$: $\gamma_1(x)$ shows the curve at which $d\Psi/dt = 0$ for the case of the transport demand evenly distributed over the day; $\gamma_2(x)$ shows the curve where a distinction is made between transport demand during day times, night times and rush hours. For this, the following assumptions are made: rush-hours last for 4 hours per day with an intensity of 4 times the day intensity; night traffic last for 8 hours per day with an intensity of 25% of the day intensity. Therefore, the daytime intensity lasts for 12 hours per day.

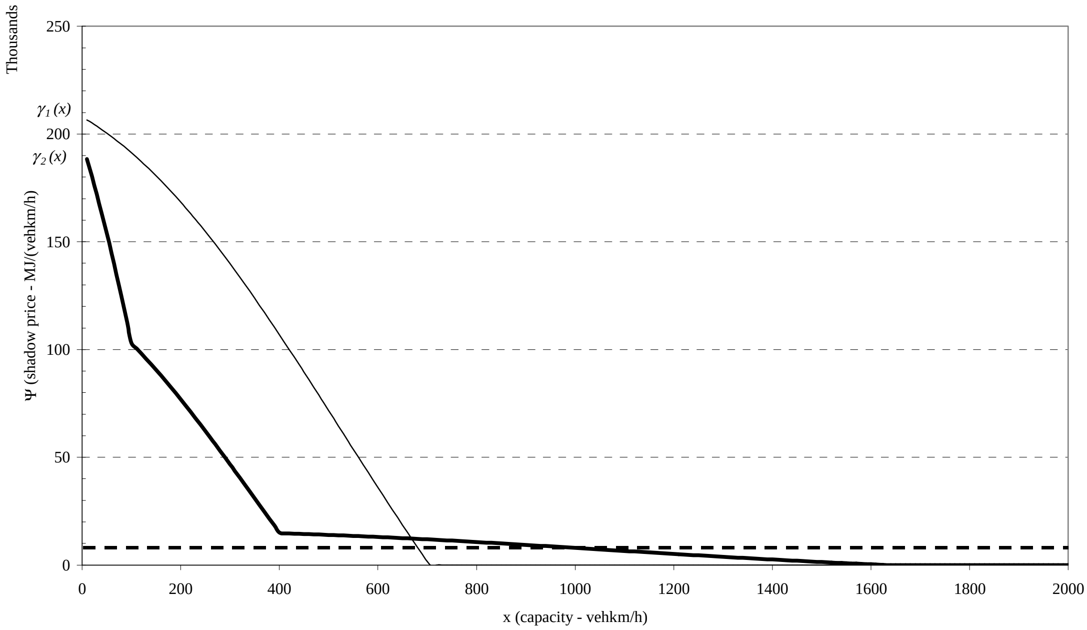

Figure 20 Hamiltonian system for one kilometer-lane of average Dutch state road. The thin line represents the situation with a constant transport demand throughout the day. The black line is more in agreement with reality, since it distinguishes into quiet night hours, normal day hours and busy rush hours. The stable intersection point shifts from approximately 700 vehkm/h to 1000 vehkm/h due to the energy inefficiency during rush hours. The line of maximum capacity lies above the *x*=5000 vehkm/h, so far to the right of the graph.

The stable capacity level lies at 1000 vehkm/h. By definition, the capacity of one kilometer of single lane is 2000 vehkm/h. Therefore, even with inclusion of the rush hour inefficiency, the Dutch state road infrastructure is *on average* overdimensioned from an energetic point of view.

6.5. Examples of other possible cases

Although the Dutch road system is overdeveloped on average, it does not mean that no specific road project should be undertaken anymore. A more detailed analysis per road section might reveal other conclusions. Consider for example the road section of the freeway A4 between Roelofarendsveen and intersection Burgerveen. In the year 2000, this section saw more than 100 traffic jams. Traffic is expected to increase significantly in the future on this section as is shown in figure 21. It can be shown that even from an energy point of view, this section needs to be expanded – if one does not take network effects into account. The methodology proposed in this paper can in this case provide an adequate answer to which extent this road section should be widened.

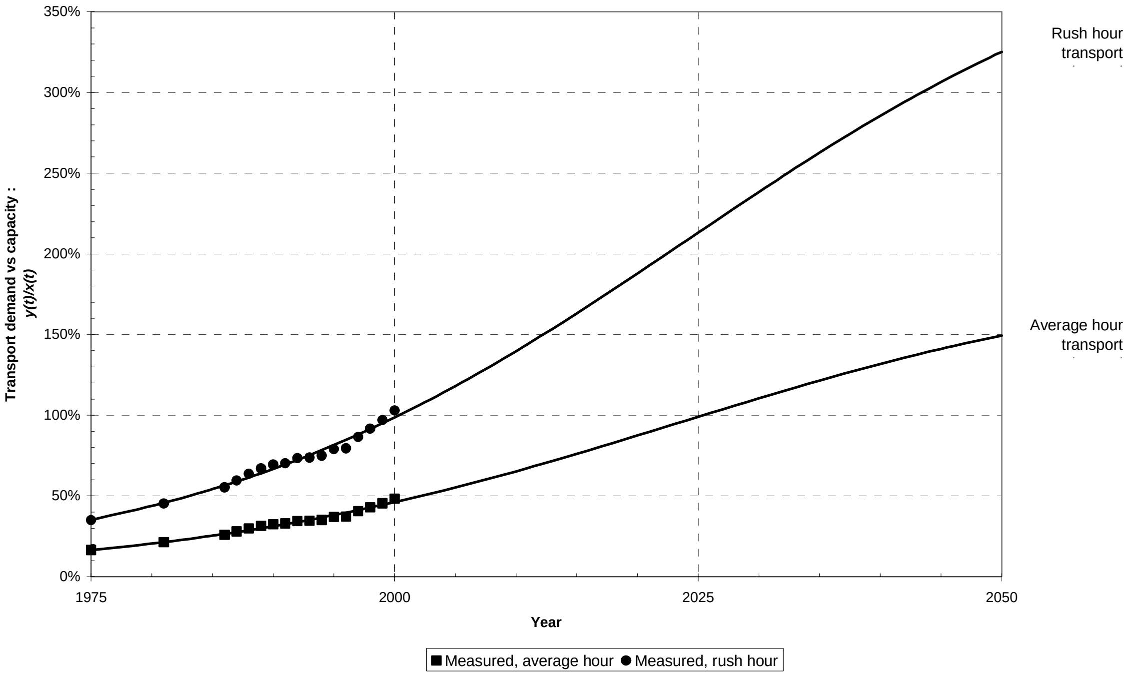

Figure 21 Recorded and expected transport demand on a section of the freeway A4, Roelofarendsveen-Burgerveen. In 2025 the transport demand will reach on average to capacity. At least from a transportation point of view, the section needs to be upgraded.

Another interesting case might be one in which there are insufficient funds to increase the capacity of the infrastructure. Two situations are easily identifiable. One is the case in which a small municipality does not have enough funds to improve the local transportation system adequately, the other relates to a poor country, in which the total road network system cannot be expanded and maintained at an appropriate level due to insufficient funds again.

7. Conclusion

7.1. Applicability of methodology

The methodology of optimal control theory can successfully be applied to the problem of minimizing the energy use of the transportation sector. There exists a unique transition path that leads towards the optimal solution. Theoretically, two stable situations exist, depending on the available resources. If analysis shows that the infrastructure has to expand, then the theory shows that the infrastructure has to expand as fast as possible either to the optimal capacity or to the highest reachable capacity level. Should, however, the transport demand not be constant in time, then further analysis is necessary in other to determine the transition speed.

To use the described methodology, it was necessary to construct a function that represents the fuel consumption of traffic as function of the transport demand and the road capacity. It shows that transportation inside traffic jams at low velocities uses more energy than normal traffic on highways.

The methodology is applied towards the issue of reducing energy consumption. Other utility functions, for example those who describe economic damage due to traffic jams, can help formulate solutions on how far infrastructure should expand on to economic criteria.

7.2. Relevance of case situations

The two cases, one for sufficient resources and another for insufficient resources, are theoretically deduced. These resources can be bound by financial or material constraints, but also a resource limitation due to available workforce or permitted road closures is possible. Road authorities in the Netherlands will presumably have sufficient financial and material resources. For the case of the average Dutch road network, the methodology is not suitable. On average, the road network is overdimensioned. The averaging over all kind of roads makes the provided solution too straightforward.

The presented methods can be applied internationally. It is conceivable that some countries do not have sufficient means to expand the network to the desired level. Also congested small communities in Western countries can face problems of insufficient resources. To define an upper bound for the available resources has been found to be difficult if not impossible.

8. Discussion

A crucial element in the beginning of this project was the development of a formula to describe the adverse environmental effect of a traffic jam. It is chosen to take the capacity of the road as starting point and to find a relation between the energy consumption of road traffic and the available relative capacity. As road traffic is a very dynamic phenomenon, one can argue about the validity of such a function. The function presented here is primarily used to enable the investigation into the applicability of the methodology of optimal control theory. For future research it should be noted that some distinction into vehicle types, like passenger cars and freight vehicles, is recommended. This might lead to a stochastic approach of defining the utility functions.

The utility function *F(x)* does not meet all the conditions everywhere as required by the maximum principle. Future research might try both to make the conditions less strict, and to approximate *F(x)* carefully in such a way that it does meet all the conditions everywhere.

This paper used a constant value for the deterioration rate of the infrastructure. For a case of the Netherlands, where maintenance is provided at an adequate and frequent level, this is true by approximation. One can consider to make the deterioration rate also time dependent by including a differential equation for the deterioration.

It is not uniformly clear which physical actions are considered production activities and which are maintenance activities. One can either establish a clear definition of those, or redefine them on a case-by-case basis. The relative energy use for production activities and the relative energy use for maintenance are, for most capacity improving measures, not equal. The redefinition of production and maintenance can for a specific case rescale those relative energy investment figures. Another option to avoid the problem of these non-equal relative energy investment figures might be to see the maintenance no longer as control variable. The statement that the maintenance rate must equal the deterioration rate, would resolve into a different set of equations.

Future research should try to include a time dependent transport demand, which gives a more detailed insight in the speed at which the transport network should grow. More geographical detail in the case studies is necessary, but calls for inclusion of network relations. Not only does a change in capacity of one road effect the traffic on other roads, but also might a capacity increase lead to induced traffic. For a single road, it is still possible to modify the utility function only slightly to include this latter phenomenon.

Literature

1. A.Alberts, Energie- en materiaalaspecten van leefbaarheids-, bereikbaarheids- en veiligheidsmaatregelen op het hoofdwegennet, Tauw bv, report 3957292, Deventer, June 2002.
2. K.J.Arrow, M.Kurz, Public investment, the rate of return and optimal fiscal policy, The John Hopkins Press, Baltimore, 1970.
3. S.M.Aseev, A.V.Kryazhimskii, A.M.Tarasyev, First order necessary optimality conditions for a class of infinite horizon optimal control problems, IIASA interim report IR-01-007, Laxenburg, February 2001.
4. S.M.Aseev, G.Hutschenreiter, A.V.Kryazhimskii, A dynamical model of optimal allocation of resources to R&D, IIASA interim report IR-02-016, Laxenburg, March 2002.
5. E.J.Balder, An existence result for optimal economic growth problems, J. Of Math.Analysis and Applications, Vol.95 pp 105-213, 1983.
6. A.J.M.Bos, Direction Indirect, thesis University of Groningen, Groningen, 1998.
7. CBS, Statline, internet database, http://www.cbs.nl, 2002.
8. CBS, Zakboek verkeer en vervoer 1995, report N41/1995, ISBN 9035716728, Voorburg/Heerlen, 1995.
9. A.Grübler, N.Nakićenović, Evolution of transport systems: past and future, IIASA, report 91-8, Laxenburg, 1991.
10. A.Haurie, Integrated Assessment Modeling for Global Climate Change, NCCR-WP4 working paper draft version, December 2001.
11. IFIAS, 1974. Energy analysis. Workshop on methodology and conventions. Report no. 6. International Federation of Institutes for Advanced Study, Stockholm.
12. R.Kok, R.M.J.Benders, H.C.Moll, Energie-intensiteiten van de Nederlandse consumptieve bestedingen anno 1996, IVEM research report 105, Groningen, March 2001.
13. E.Kreutzberger, J.M.Vleugel, Capaciteit en benutting van infrastructuur: capaciteitsbegrippen en infrastructuurgebruik in de binnenvaart en het lucht-, rail- en wegvervoer, University of Delft, ISBN 90-6275-739-1, Delft, 1992.
14. T.Litman, ITE Journal, vol 71, no 4, Institute of Transportation Engineers, April 2001, pp38-47.
15. Ministerie van Verkeer en Waterstaat, AVV, Verkeersgegevens jaarrapport 2000, ISSN 0926-6011, Rotterdam, Oktober 2001.
16. Ministerie van Verkeer en Waterstaat, Jaarrapport weggegevens 1998, report P-DWW-99-046, ISBN 90-369-3751-5, Delft, September 1999.
17. Ministerie van Verkeer en Waterstaat, Nationaal Verkeer en Vervoerplan, Den Haag, 2001 (National traffic and transport plan).

18. Ministerie van VROM, Uitvoeringsnota klimaatbeleid, Den Haag, 1999 (Climate Policy Implementation Plan).

19. W.D.O.Paterson, Road deterioration and maintenance effects: models for planning and management, World Bank, ISBN 0-8018-3590-9, Washington, 1987.

20. L.S.Pontryagin, V.G.Boltyanskii, R.V.Gamkrelidze, E.F.Mishchenko, The Mathematical Theory of Optimal Processes, Interscience Publishers, John Wiley&Sons, New York, 1962.

21. S.P.Sethi, G.L.Thompson, Optimal control theory, second edition, ISBN 0-7923-8608-6, Kluwer, Dordrecht, 2000.

22. J.Veurman, I.Wilmink, R.Gense, H.Baarbé, Files zorgen vooral lokaal voor milieueffecten: effecten van congestie op brandstofverbruik en luchtkwaliteit, Verkeerskunde 2 (2002), pp 32-38.

23. J.Wahle, L.Neubert, M.Schreckenberg, Modeling and simulation of traffic flow, Computer Physics Communications 121-122 (1999), pp 402-405.