Stellingen behorende bij het proefschrift

A stochastic heat equation for freeway traffic flow

van

Ello Weits

1. Zij T_t een contractie halfgroep op een complexe Hilbertruimte H. Veronderstel dat T_t voldoet aan de eis dat $T_t^* T_t = T_t T_t^*$, d.w.z. T_t is normaal. Dan geldt dat de sterke limiet van $T_t^* T_t$ voor $t \to \infty$ een projectie is (noem deze Q). Bovendien heeft de orthogonale decompositie van H: $H = QH \oplus QH^{\perp}$ de eigenschap dat T_t is unitair op QH en volledig niet-unitair op $QH^{\perp}$.

   E.B. Davies, *One-parameter semigroups*, Academic Press, London, 1980.
   Ello Weits, *Terugkeer naar evenwicht*, doctoraalscriptie natuurkunde, 1985.

2. Zij $\{X_n\}$ een reëelwaardig stochastisch proces ($n \geq 0$) gegeven door: $X_n \sim N(0, V)$ en $V_{nm} = \phi^{|n-m|} + \phi^{n+m}$, met $0 < \phi < 1$. $\{X_n\}$ kan worden gerepresenteerd als een autoregressief proces (een AR(1)-proces):

$$X_n = a_n + \phi X_{n-1},$$

   waarbij $\{a_n\}$ een verzameling van i.i.d. stochasten is met $a_n \sim N(0,1)$. Het onderscheid met de stationaire variant, in welk geval $V_{nm} = \phi^{|n-m|}$, schuilt alleen in de reconstructie van de eerste storingsterm a_0. In plaats van $a_0 = \sqrt{1-\phi^2} X_0$ stellen we nu

$$a_0 = \sqrt{\frac{1-\phi^2}{2}} X_0.$$

3. Het volgende stelsel evolutievergelijkingen wordt wel gebruikt om een verkeersstroom op een autosnelweg te beschrijven:

$$\begin{aligned} \frac{\partial \rho}{\partial t} &= -\frac{\partial (\rho v)}{\partial x} \\ \frac{\partial v}{\partial t} &= \frac{1}{\tau}(V(\rho) - v) - \frac{\nu}{\rho}\frac{\partial \rho}{\partial x} - v\frac{\partial v}{\partial x}, \end{aligned}$$

   waarbij $\rho(t,x)$ de dichtheid van de verkeersstroom op tijd t en plaats x is en $v(t,x)$ de snelheid; τ en ν zijn positieve constanten; $V(\rho)$ is een functie, die voor elke waarde van ρ de bijbehorende 'evenwichtssnelheid' geeft.

   In tegenstelling tot de bewering van Reinhart Kühne kent dit stelsel geen eenduidige schokgolfoplossingen. De oorzaak hiervan is dat met de tweede vergelijking geen behoudswet correspondeert, zodat er geen voorschrift beschikbaar is om het gedrag van de oplossing in de discontinuïteitspunten vast te leggen.

Reinhart D. Kühne, Macroscopic Freeway Model for Dense Traffic — Stop-start Waves and Incident Detection, *Proceedings of the Ninth International Symposium on Transportation and Traffic Theory*, VNU Science Press, pp, 21–42, 1984.
Reinhart D. Kühne, Fernstraßenverkehrsbeeinflussung und Physik der Phasenübergänge, *Physik in unserer Zeit*, 15 (1984), nr. 3, pp. 84–93.
G.B. Whitham, *Linear and Nonlinear waves*, Wiley, New York, 1974.

4. Zij H een reële separabele Hilbertruimte en X_t een H-waardig proces, dat wordt beschreven door

$$dX_t = KAX_t\,dt + dB_t,$$

waarbij K een positieve constante is, A een operator op H die een geschikte halfgroep genereert en B_t een H-waardige Brownse beweging. De schatter voor de parameter K, die Koski en Loges tooien met de naam 'minimum contrast schatter' (overigens niet ten onrechte), is in feite niets anders dan een momentenschatter.

Koski & Loges, On Minimum-Contrast Estimation for Hilbert Space-Valued Stochastic Differential Equations, *Stochastics*, vol. 16 (1986), pp. 217–225.

5. Veronderstel dat σ een vermenigvuldigingsoperator is, afgeleid van een Lipschitz-continue functie op $\mathbf{R}$. Dan heeft de oneindig dimensionale stochastische differentiaalvergelijking

$$dX_t = \frac{d^2}{dx^2}X_t\,dt + \sigma(X_t)\,dB_t, \qquad 0 \leq t \leq T$$

zowel in het geval dat X_t een functie is op $[-M, M]$ als in het geval dat X_t een functie is op geheel $\mathbf{R}$ een unieke oplossing. Laten we deze X_t^M en X_t noemen. Beide gevallen worden meestal met iets verschillende wiskundige instrumenten behandeld. Als X_t^M en X_t beschouwd worden als $C([0,T] \times [-S, S])$-waardig processen (voor vaste S), dan convergeert de oplossing X_t^M in verdeling naar X_t.

6. Het verstrekken van OV-jaarkaarten aan grote groepen van de (beroeps)bevolking levert geen wezenlijke bijdrage aan de oplossing van de mobiliteitsproblematiek.

7. Alleen al de mogelijkheid om kinderen te krijgen is voor een vrouw op de arbeidsmarkt een handicap.

8. De verhouding tussen de filosofie en de theologie vertoont overeenkomsten met de verhouding tussen de wiskunde en de natuurkunde.

A stochastic heat equation for freeway traffic flow

A stochastic heat equation for freeway traffic flow

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus,
prof. drs. P. A. Schenck,
in het openbaar te verdedigen
ten overstaan van een commissie
aangewezen door het College van Dekanen
op donderdag 17 mei 1990 te 16.00 uur

door

Ello Aart Gijsbert Weits

geboren te Emmen,
doctorandus in de natuurkunde.

Dit proefschrift is goedgekeurd door de promotor
prof. dr. P. Groeneboom.

Samenstelling promotiecommissie:

de Rector Magnificus
prof. dr. P. Groeneboom (Technische Universiteit Delft)
prof. dr. M. S. Keane (Technische Universiteit Delft)
prof. dr. R. D. Gill (Rijksuniversiteit Utrecht)
prof. dr. C. L. Scheffer (Technische Universiteit Delft)
prof. dr. L. de Haan (Erasmus Universiteit Rotterdam)
prof. dr. ir. J. H. A. de Smit (Universiteit Twente)
dr. A. W. J. van der Weide (Technische Universiteit Delft)

opgedragen aan mijn ouders

Preface

This thesis reports the main results of research carried out in behalf of the project 'Statistical analysis of freeway traffic flow', which has been supported by the Netherlands Foundation for Technical Research (STW). Prof. P. Groeneboom has been the project leader.

The project has grown out of a consultation of prof. P. Groeneboom for the Transportation and Traffic Engineering Division (DVK) of the institute 'Rijkswaterstaat' of the Dutch Ministry of Transport.

The central point of interest in the consultation as well as in my research has been the modelling of freeway traffic flow. A succesful characterization can be used as a basis for the design of measures of the homogeneity of traffic.

I have enjoyed the hospitality of the Centre for Mathematics and Computer Science in Amsterdam during the first two years of the project and of the department of Mathematics and Computer Science of the Technical University of Delft during the last two years. Both institutes have provided me an inspiring environment. Scientific literature, computer facilities etc. were readily available and also I had opportunity to discuss all kinds of problems with colleagues.

Mathematical research directed towards application needs the input of data. These were supplied by DVK of Rijkswaterstaat. Rijkswaterstaat also participated in the so-called 'gebruikerscommissie'. This background committee saw to the overall progress of the project. Furthermore, Rijkswaterstaat will investigate in the near future the possibilities of applying the results of the research.

The thesis has been printed at the Centre for Mathematics and Computer Science.

I express my gratitude to all, who have contributed to the writing of the thesis and to my joy of living while working on it.

Delft, April 1990 Ello Weits

Contents

Chapter 1

Introduction

1.1 Traffic problems

Traffic has increasingly become the subject of discussion and criticism in the Netherlands. Recently much attention has been given to the issue of air pollution caused by traffic. Other problems are the noise produced, the amount of space taken up and the lack of safety. Furthermore, the increase of traffic intensity causes many daily traffic jams in the more densely populated areas of the country.

It is widely recognized that traffic and mobility meet deeply rooted needs of Dutch society. Hence the solutions to the problems mentioned above are in the first place a matter of political and economical decision making.

Nevertheless, it may be expected that technical measures will be part of any policy. Such measures concern for example construction of vehicles that consume less energy and produce less pollution as well as less noise, design of efficient traffic networks around and between cities, improvement of public transport and also a more efficient use of the freeway capacity.

This thesis aims at a mathematical description of traffic flow on a freeway[1], thereby identifying some useful flow characteristics. Estimating these characteristics and taking appropriate action may lead to enhancement of the homogeneity of the flow. The capacity of a particular freeway is not entirely independent of the traffic stream; improving the homogeneity of the flow enlarges what might be called the 'effective'

[1]Although the text is written in English, we always use the American word 'freeway' instead of the English word 'motorway', because 'freeway' is standard terminology in transportation science.

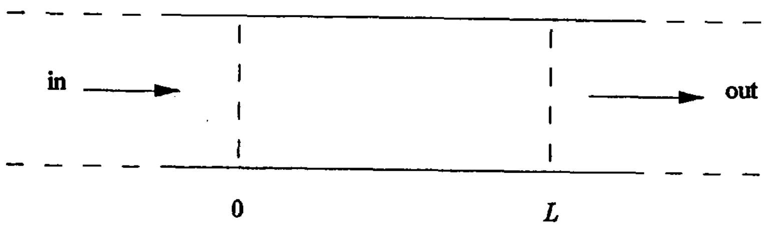


Figure 1.1: The traffic is observed on a stretch of a freeway of fixed length (L).

capacity.

While much congestion is due to accidents and the existence of bottlenecks (especially at junctions near the larger cities), many other occurrences of congestion can be ascribed to lack of freeway capacity. When the intensity of the traffic stream increases above some 'saturation value', a small disturbance may cause a collapse of the traffic flow.

Figure 1.1 schematically shows the typical situation. Entrances and exits are not taken into account; we concentrate upon the behaviour of the traffic flow between junctions.

1.2 High density stationary freeway traffic

The attempt to model freeway traffic flow is usually undertaken for practical reasons. As mentioned above, the application we have in mind is in the first place an efficient use of the freeway capacity. Therefore, the interest is primarily directed at high density multilane freeway traffic as low density freeway traffic is not problematic, except when questions of safety are discussed.

The high density assumption already is part of a vocabulary that exploits the analogy between traffic flow and fluid flow. Often high density multilane freeway traffic is described in terms of density (number of vehicles per kilometer), mean velocity (kilometers per hour) and 'flow' (number of vehicles passing some point per hour). The analogy is quite obvious; nevertheless there are also differences. The level of aggregation is rather low in the case of traffic flow (hundreds or maybe thousands of vehicles compared to billions of fluid particles). Furthermore, the freeway traffic flow cannot be classified into a few classes of identical vehicles, whereas precisely this classification ensures the validity of many calculations concerning fluid flows.

In Chapter 2 it is argued that these deficiencies of the analogy are one reason to add a stochastic term to the equation describing the behaviour of the traffic flow. Another reason is the fact that drivers exhibit a nonconstant, stochastic driving style.

In order to keep the model mathematically tractable, the non-linear equations describing the flow are linearized. This simplication is valid as long as the resulting model is confined to a stationary high density multilane freeway traffic flow. The restriction to a stationary flow is not very serious. Every traffic flow model has a limited area of application. (Or, alternatively, every type of traffic requires its own mathematical model.) For example high density (multilane) traffic subject to bottlenecks such as traffic lights might be best described by retaining the non-linearity of the basic equations without inclusion of a stochastic term.

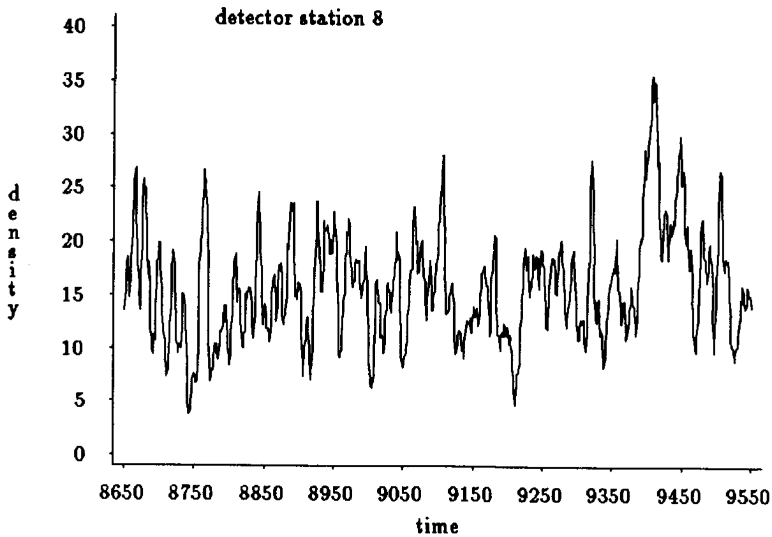
The linear model describes a stationary high density freeway traffic flow in terms of some fixed mean density and stochastic fluctuations which occur around this mean. A flow is said to be homogeneous if the fluctuations are small. The capacity of the freeway can be defined as the maximum 'flow' (or intensity, measured in number of vehicles per hour) that can pass along the freeway. This maximum usually corresponds to a maximum density if the flow is uncongested. Obviously, the mean density should be below the maximum density. The margin, however, is determined by the maximum amplitude of the fluctuations. If the sum of the mean density and the instantaneous fluctuation exceeds the maximum density, a breakdown of the flow is likely to occur. It should be stressed that, due to linearization of the equations, this also implies a breakdown of the model. Therefore, the model is intended to describe the characteristics of the traffic flow that give rise to the breakdown (the onset of congestion), but not what happens after the breakdown.

Figure 1.2 shows the evolution of the density at a particular site of the freeway A13 near Delft, as measured on September 27, 1989 between 15.30 h. and 16.00 h.

In order to make the notion of characteristics somewhat clearer, it is useful to write down here the linearized equation describing the evolution of the density, $R(t,x)$, where t is the time and x the space variable. The equation is an example of a stochastic heat equation.

$$\frac{\partial R}{\partial t} = K\frac{\partial^2 R}{\partial x^2} - c_0\frac{\partial R}{\partial x} + \sigma\frac{\partial B}{\partial t}. \tag{1.1}$$

K and c_0 are positive constants. K is a parameter that determines how strong the smoothing tendency of the process is that counteracts the



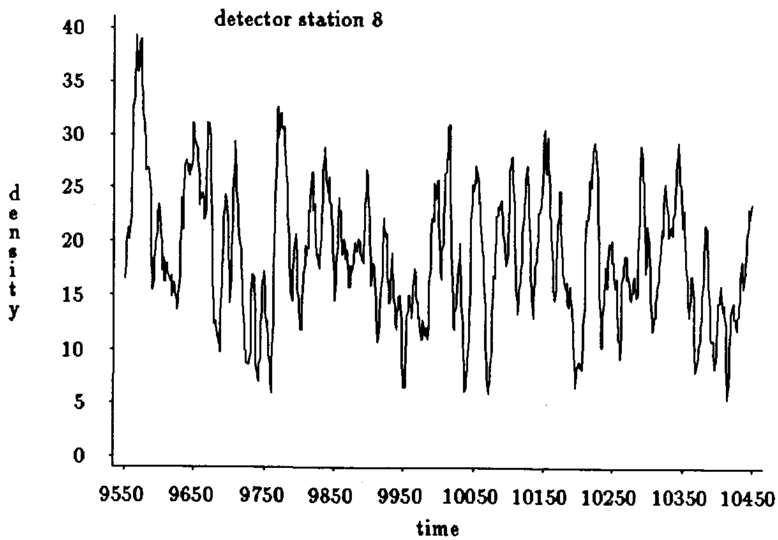
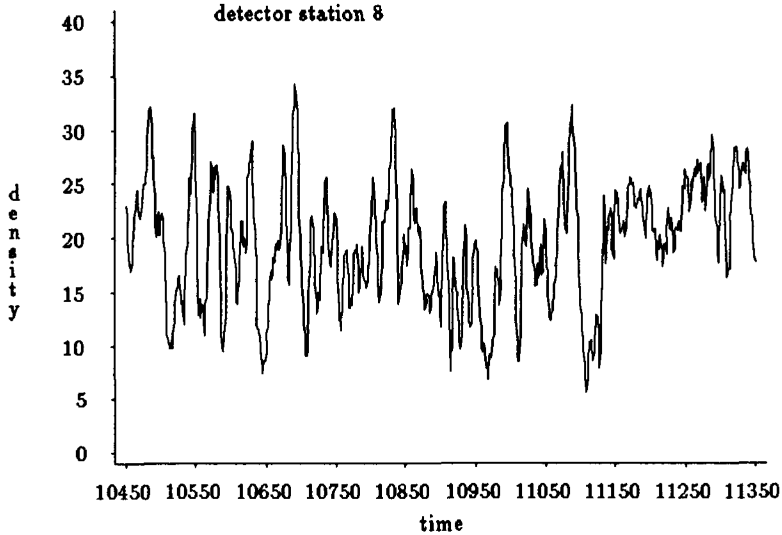


Figure 1.2: The evolution of the density (vehicles per km per lane) at detector station nr. 8 situated on the western carriageway of the freeway A13 (see for further description of these data Chapter 7). The time is given in seconds.



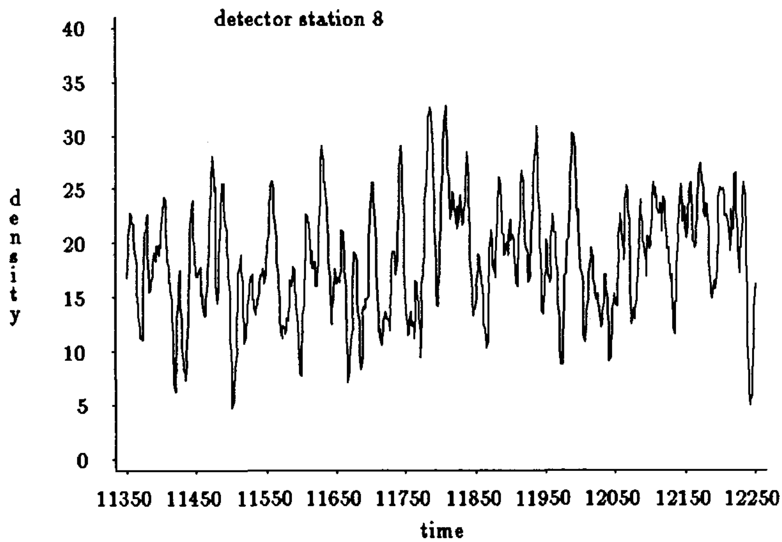


Figure 1.2: Continued.

influence of the noise term; c_0 is a velocity. The noise term, $\sigma\,\partial B/\partial t$, contains the stochastic disturbances of the traffic stream. This term not only depends on σ, but also (inplicitly) on another (real-valued, positive) parameter S; σ determines the amplitude of the disturbances and S their range (so the mean density is effectively the mean of the density over a stretch of the freeway that is long compared with S). K, σ and S may be called the characteristics, since they determine the degree of (non-)homogeneity of the flow.

It should be emphasized that equations such as (1.1) require careful interpretation. The first part of the thesis is therefore devoted to this subject.

1.3 Outline of the following chapters

In Chapter 2 a survey of several freeway traffic flow models is given. Furthermore, the choice of the 'macroscopic continuum model', which already has been sketched above, is motivated. Chapters 3 and 4 contain the mathematical (probabilistic) background of the model. In Chapter 5 we summarize the discussions of Chapters 2 to 4 presenting once more the ideas of these chapters, without entering into the mathematical details. Chapter 6 describes the statistical analysis of the model. The results of the comparison of the model with real data are reported in Chapter 7. Finally, Chapter 8 contains a summary, conclusions and also some remarks concerning the utilization of the research.

Chapter 2

Models of freeway traffic flow: a survey

Research on the subject of freeway traffic flow (and on other problems of traffic theory) started some thirty years ago, in particular in the Unites States. A key paper was written by Lighthill and Whitham in 1955 [19]. During these three decades various different approaches have been proposed, but no particular model seems to be superior to the others. There are at least two obvious reasons for this.

(i) A model can be microscopic or macroscopic; individual vehicles are observed in practice, but is it judicious to incorporate individual behaviour into the model?

(ii) In microscopic as well as in macroscopic models there are a lot of factors that cannot be modelled exactly, but have to be viewed as random disturbances. As it is by no means clear how this should be done and as in most cases introducing stochastic components greatly complicates the analysis of the model, should one decide for a stochastic or for a deterministic model?

These two questions are unresolved in general. Therefore, we can readily distinguish between four types of model, each of which has its own advantages and disadvantages. These four types are shown and categorized in Table 2.1.

There is a fifth type which does not fit in with this classification, namely the models based on a kind of a 'Boltzmann equation'; with regard to both aspects of the classification they occupy an intermediate position. We will discuss each type of model separately. Our primary

	deterministic	stochastic
microscopic	car following theories	headway models, simulation
macroscopic	continuum models	stochastic continuum models

Table 2.1: Classification of freeway traffic flow models.

goal is to show how the specific assumptions of each type relate to a specific area of application. In the last section we give an outline of the stochastic continuum model, that we concentrate on in the rest of this thesis (see Chapter 5 for a complete description of the model).

2.1 Microscopic models

2.1.1 Car following theories

Car following theories are the typical representatives of the class of microscopic, deterministic models. One considers a long (possibly infinite) sequence of vehicles which are numbered $1, 2, \ldots$; the vehicle in front has number 1, the one following number 2 etcetera. At time t the position of vehicle i is given by $y_i(t)$. Usually one denotes its velocity by $\dot{y}_i(t)$. A suggested equation of motion is [12,38]

$$\dot{y}_i(t+T) = G(y_{i-1}(t) - y_i(t))$$

or, equivalently, assuming G to be differentiable,

$$\ddot{y}_i(t+T) = [\dot{y}_{i-1}(t) - \dot{y}_i(t)]\, G'(y_{i-1}(t) - y_i(t));$$

G is a particular function of the headway and T is the time lag of the driver-vehicle system. Car following theories have been applied mainly to single lane traffic, especially in tunnels (see for example [25]). This is, of course, due to the fact that phenomena such as passing and lane-changing cannot be described easily in this setting. Sometimes a stochastic term is added to the (second) differential equation representing the so-called 'acceleration noise', the discrepancy between the actual acceleration and the 'ideal' acceleration [24].

2.1.2 Headway models and simulation

Headway models (see for example [11] and [4]) usually deal with the time headway between successive vehicles measured at some fixed point of a given lane. It is assumed that the time headways are independent and identically distributed according to some probability distribution. This probability distribution is constructed as follows: a distinction is made between 'leaders' and 'followers', followers are vehicles driving closely behind their predecessors (i.e. within a following distance or following time), whereas leaders are out of reach of their predecessors. A follower's behaviour depends strongly on the vehicle ahead, whereas a leader's behaviour does not have this property. As a consequence of this distinction the time headways of a follower and a leader are taken from different probability distributions.

Headway models introduce probability into the class of microscopic models. Nevertheless they retain the assumption that all vehicles are essentially the same. Removing this assumption to some extent would greatly complicate the model. Furthermore the models do not incorporate dynamics. And, again, they apply only to single lane traffic. Simulation models offer opportunities to overcome these restrictions. (See for example [20,32].) Letting a computer do the calculations allows us to build complex, but realistic models, in which vehicles, driving behaviour, road and weather conditions, etc. can be specified in detail. A price has to be paid for this freedom: simulation yields little insight in the crucial properties of a traffic stream. In general the relation between the microscopic specifications of the model and the macroscopic properties remains unclear.

If we want to allow for multilane traffic and for the variety of vehicle characteristics and still have tractable models, it seems best to start with a macroscopic model. The next section introduces continuum models as the prime example of a macroscopic model.

Sometimes one distinguishes microscopic and macroscopic simulation. What we have discussed here concerns microscopic simulation. Macroscopic simulation is in fact nothing else but an application of a (discretized) macroscopic model: the subject of the next section.

2.2 Continuum models

Continuum models deal with traffic streams in terms of aggregate variables. This macroscopic approach results in a limited number of equations which are relatively easy to handle. Since continuum models view the traffic as a continuous stream, they are obviously especially suited for high density traffic. These models are not rigorously built from microscopic 'principles', instead the information about the vehicles and the dynamics is macroscopic from the beginning: it consists largely of field measurements and some heuristic reasoning.

The three basic aggregate variables are: the flow $q(t,x)$ (vehicles per hour), the density $\rho(t,x)$ (vehicles per kilometer) and the velocity $v(t,x)$ (kilometers per hour), where t and x denote time and place, respectively. Here both t and x are taken to be continuous variables, though often place or place and time are discretized in order to facilitate simulation and comparison with experimental data. (See for example [21].) The range of t and x is usually specified afterwards, together with appropriate initial and boundary conditions.

If we assume that the aggregate variables are differentiable functions of t and x, then we have two exact relations:

$$q = \rho v \tag{2.1}$$

$$\frac{\partial \rho}{\partial t} = -\frac{\partial q}{\partial x}. \tag{2.2}$$

The equation (2.1) is obvious; (2.2) states 'the conservation of vehicles'.

To get a complete description of the dynamics we need a third model equation. There are at least two possibilities:

(a) assume that q is a (differentiable) function of ρ, i.e. $q = Q(\rho)$;

(b) derive an equation describing the evolution of the velocity.

The first possibility yields, writing $dQ/d\rho = c(\rho)$,

$$\frac{\partial \rho}{\partial t} + c(\rho)\frac{\partial \rho}{\partial x} = 0. \tag{2.3}$$

This non-linear, first order partial differential equation was introduced by Lighthill and Whitham [19]. The solutions to this equation are waves that develop into shock waves [38, pp. 68–77]. These shock waves bear resemblance to phenomena observed in traffic streams. Note that assuming q to be a function of ρ is equivalent to assuming that v is a function of ρ: $v = V(\rho) = Q(\rho)/\rho$.

The second possibility to complete the description exists in assuming a more complicated relation between v and ρ. Several evolution equations for the velocity have been suggested. We mention two of them. Payne [28] (see also [36,38]) proposed the following equation:

$$\begin{aligned} \frac{dv}{dt} = \frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} &= \frac{1}{\tau}((V(\rho) - v) - \frac{\nu}{\rho}\frac{\partial \rho}{\partial x}) \\ &= \frac{1}{\tau}(V(\rho) - v) - \frac{\nu}{\tau\rho}\frac{\partial \rho}{\partial x}, \end{aligned} \tag{2.4}$$

where dv/dt is the total time derivative of v for a 'moving observer'. A moving observer observes the traffic stream while moving along with the stream at the same (variable) speed v. The total time derivative decomposes into the true time derivative plus the so-called convection term. This convection term arises, because the two terms on the right hand side (which are called the relaxation term and the anticipation term, respectively) are conceived as effects acting on a moving traffic stream (and thus observable by a moving observer). The relaxation term describes a tendency of the traffic stream to adjust its velocity to a value V that matches the density. Note that in the Lighthill-Whitham model this adjustment takes place instantaneously, i.e. $\tau \downarrow 0$ (for $\nu = 0$). Finally, the anticipation term represents the idea that a traffic stream also anticipates near-future situations, which announce themselves via a density gradient. The parameters τ and ν are in most cases assumed to be positive constants.

In particular the anticipation term has been subject to criticism. The thesis of S. A. Smulders [36, Chapters 2 and 3] offers a thorough discussion of this term and some alternatives. Here yet another alternative is suggested:

$$\begin{aligned} \frac{dv}{dt} = \frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} &= \frac{1}{\tau}((V(\rho) - v) - \mu\frac{\partial v}{\partial x}) \\ &= \frac{1}{\tau}(V(\rho) - v) - \frac{\mu}{\tau}\frac{\partial v}{\partial x}, \end{aligned} \tag{2.5}$$

where μ is some constant. The idea behind this choice is that drivers in a traffic stream are more likely to anticipate velocity changes than density changes, as the first more directly entail possible danger.

Up till now no randomness was included in the continuum model. There are arguments for introducing randomness (by adding a random term to one of the equations) and also arguments against it. Let us first list some arguments in favour of introducing a stochastic term. It appears that experimental data exhibit lack of small scale 'regularity' in spite of certain large scale effects. This irregularity seems to be due to the large variation in vehicle characteristics and to the limited level of aggregation (i.e. the number of 'particles' per unit of distance or time is relatively small, compared with fluid flows). Furthermore, driver characteristics are not wholly constant in time nor entirely predictable (cf. the 'acceleration noise' in Subsection 2.1.1). Finally, a stochastic term might compensate for modelling errors in the deterministic terms.

The prime disadvantage of adding a stochastic term is the increasing complexity of the model. The only way to avoid this seems to be linearization of the model equations. Fortunately, linearization not only reduces the complexity of the model, but also eliminates the dilemma posed by the presence of various possible anticipation terms. It turns out that the Payne-model (2.4) as well as the equation (2.5) reduce to the same linearized equation, viz.

$$\frac{\partial r}{\partial t} = K\frac{\partial^2 r}{\partial x^2} - c_0\frac{\partial r}{\partial x}. \tag{2.6}$$

Here $r(t,x)$ denotes the deviation of the density around some 'mean' value ρ_0; c_0 and K depend only on ρ_0, so that they can be treated as constants as long as the linearization is valid. We note that in general (2.6) has stable solutions if and only if K is positive. The expressions for K are $K = \nu - \tau[V'(\rho_0)\rho_0]^2$ and $K = -[\mu + \tau V'(\rho_0)\rho_0]V'(\rho_0)\rho_0$ for the Payne model and (2.5), respectively. Thus K is positive if $\nu/\tau > [V'(\rho_0)\rho_0]^2$ or if, in the alternative model, $\mu/\tau + V'(\rho_0)\rho_0 > 0$. (It is assumed that $V(\rho)$ is strictly decreasing in ρ, so that $V'(\rho_0) < 0$.) In both cases this condition can be interpreted as the requirement that the anticipation effect should be 'strong' enough to compensate for the instability induced by the nonzero $V'(\rho_0)$; it is worth remarking that this nonzero derivative is also responsable for the existence of shock wave solutions for (2.3). In both models c_0 equals $V'(\rho_0)\rho_0 + V(\rho_0)$.

Our interim conclusion is that the model equation of a 'linearized stochastic continuum model' might read as follows:

$$\frac{\partial R}{\partial t} = K\frac{\partial^2 R}{\partial x^2} - c_0\frac{\partial R}{\partial x} + \text{noise term}, \tag{2.7}$$

where we have written R instead of r to indicate that now the density(-fluctuation) is a stochastic function (of t and x). As will be shown later (see Chapter 4), this equation permits calculations that are not too complex, provided a suitable choice of the noise term is made. The remaining question is whether the linearization does not limit the area of application too much. We will return to this question in the last section of this chapter. In the meantime we will have a look at the 'unclassified' models, which are based on a kind of Boltzmann equation.

2.3 Models based on a Boltzmann equation

2.3.1 The basic idea

As always one starts by considering an ideal situation, in this case one carriageway of a freeway of infinite length lacking entrance and exit ramps. The traffic on the carriageway is described by a distribution function $f(x, v, t)$: at time t the expected number of vehicles present at a location between x and $x + dx$ and having a velocity between v and $v + dv$ equals $f(x, v, t)\, dx\, dv$. Note that $f(x, v, t)$ is an expected value: the real value fluctuates around this expectation. The use of this kind of distribution function (not to be confused with the distribution function in statistics) has been borrowed from statistical physics, where it is used to describe what happens in a dilute gas. Change of $f(x, v, t)$ is due to the fact that drivers increase or decrease their velocity and to the so-called 'convection'. The word convection simply denotes the movement of the traffic and its changing state as a consequence thereof.

Prigogine and Herman [30] assumed two reasons for velocity-change of individual drivers:

(a) drivers react to each other (interaction);

(b) drivers wish to drive at some desired velocity (relaxation).

These basic considerations give the following 'Prigogine-Boltzmann equation'

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + v\frac{\partial v}{\partial x} = \left[\frac{\partial f}{\partial t}\right]_{rel} + \left[\frac{\partial f}{\partial t}\right]_{int},$$

where df/dt is the total time derivative for a moving observer, which decomposes into the true time derivative plus a term due to the convection. The resulting evolution equation for the distribution function $f(x, v, t)$ is

$$\frac{\partial f}{\partial t} = \left[\frac{\partial f}{\partial t}\right]_{rel} + \left[\frac{\partial f}{\partial t}\right]_{int} - v\frac{\partial v}{\partial x}. \tag{2.8}$$

The terms of the right hand side of (2.8) bear the obvious names: relaxation term, interaction term and convection term, respectively.

Much now depends on the choice of the relaxation and interaction terms in (2.8). We will comment briefly on this matter in the next subsections.

2.3.2 The interaction term

In this subsection we first present the essentials of the derivation of the interaction term (taken from an article of Munjal and Pahl [22]) and secondly make some comments concerning the underlying assumptions (taken from an article by Paveri-Fontana [27]).

Prigogine ([29] and [30]) proposed the following interaction: when a vehicle catches up with another one, there is a probability P that it is able to pass, in which case it does not alter its velocity; with probability

$(1 - P)$ it cannot pass and slows down adjusting its velocity to the velocity of the vehicle just in front of it. The expected number of vehicles having velocity between w and $w + dw$ that interacts during a timespan of length dt at time t and at place x with a particular vehicle having velocity v $(v < w)$ equals $f(x,w,t)\,dw\,(w - v)\,dt\,(1 - P)$. Note that the incoming 'flux' equals $f(x,w,t)\,dw\,dx$, where $dx = (w - v)\,dt$, i.e. dx is the maximum distance that an approaching vehicle can cover relative to the 'slow' vehicle in front.

Integration with respect to w $(w > v)$ and multiplication by the expected number of vehicles present in the volume element $dx\,dv$ at the 'point' (x, v) yields the increase of the expected number of vehicles in the same volume element:

$$\left[\frac{\partial f}{\partial t}\right]^{+}_{int} dx\,dv\,dt = (1 - P)\,f(x,v,t)\left[\int_{v}^{\infty} f(x,w,t)(w - v)\,dw\right] dx\,dv\,dt.$$

In an analogous way we can treat the interaction between vehicles of velocity v and slower vehicles $(w < v)$. We get

$$\left[\frac{\partial f}{\partial t}\right]^{-}_{int} dx\,dv\,dt = (1 - P)\,f(x,v,t)\left[\int_{0}^{v} f(x,w,t)(v - w)dw\right] dx\,dv\,dt.$$

The total interaction term is obtained by summing these contributions:

$$\begin{aligned}\left[\frac{\partial f}{\partial t}\right]_{int} &= \left[\frac{\partial f}{\partial t}\right]^{+}_{int} - \left[\frac{\partial f}{\partial t}\right]^{-}_{int} \\ &= (1 - P)\,f(x,v,t)\left[\int_{0}^{\infty} f(x,w,t)(w - v)\,dw\right] \\ &= (1 - P)\,f(x,v,t)\,(\bar{v} - v)\,\rho(x,t),\end{aligned}$$

where $\rho(x,t) \equiv \int_0^\infty f(x,v,t)\,dv$ is the density and $\bar{v} \equiv \int_0^\infty f(x,v,t)v\,dv$ the mean velocity. It is assumed that P only depends on the density.

Paveri-Fontana [27] has listed the assumptions underlying this derivation:

(i) if a 'fast' vehicle passes a 'slow' one, its velocity is not affected;

(ii) the slow vehicle does not change its velocity during the interaction;

(iii) vehicle lengths are neglected;

(iv) the event of slowing down is, if it occurs, instantaneous;

(v) only two vehicle interactions are taken into consideration;

(vi) the assumption of 'vehicular chaos' is valid; in the derivation given above we implicitly assumed that the expected number of pairs of vehicles having velocity v and w respectively (at the same place and at the same time) is simply the product of $f(x,v,t)\,dx\,dv$ and $f(x,w,t)\,dx\,dw$: $f_2(x,v,x,w,t) = f(x,v,t)\,f(x,w,t)$, where $f_2(x,v,y,w,t)\,dx\,dv\,dy\,dw\,dt$ denotes the expected number of pairs of vehicles such that one vehicle is in dx (around x) and dv (around v) and the other in dx (around x) and dw (around w); without this assumption the expression for $[\partial f/\partial t]_{int}$ would be

$$\left[\frac{\partial f}{\partial t}\right]_{int} = (1-P)\int_0^\infty f_2(x,v,x,w,t)(w-v)\,dw.$$

Munjal and Pahl [22] as well as Paveri-Fontana [27] rightly assert that assumption (vi) is likely to break down in the case of high density traffic. Since the velocities of vehicles that are close to each other tend to be positively correlated, it is obvious that $f_2(x,v,x,v,t) > f(x,v,t)^2$. Furthermore, as soon as queues have appeared, these queues interact as a whole with other queues or with individual vehicles. Thus assumption (v) seems to be questionable as well in the case of high density traffic. Assumptions (i) to (iv) also lack plausibility in this situation. It is for example quite clear that vehicle lengths become crucial when the density is approaching the critical value at which a jam is likely to appear.

2.3.3 The relaxation term

Prigogine's [30] proposal is

$$\left[\frac{\partial f}{\partial t}\right]_{rel} = -\frac{f(x,v,t) - f_0(x,v,t)}{T},$$

where T denotes the relaxation time and $f_0(x,v,t)$ the desired velocity distribution function. The relaxation time is the typical time needed by a vehicle to adjust its velocity to the desired velocity. Since T is chosen to depend on the passing probability P, it can be interpreted as the average time a fast vehicle is stuck behind a slow one (see for example [2]). The desired velocity distribution function has been subject of many investigations (see for example [27]).

2.3.4 Concluding remarks with respect to the Prigogine-Boltzmann models

The 'elementary' Prigogine-Boltzmann equation now reads

$$\frac{\partial f}{\partial t} = -\frac{f(x,v,t) - f_0(x,v,t)}{T} + (1-P)f(x,v,t)(\bar{v} - v)\rho(x,t) - v\frac{\partial v}{\partial x},$$

where P depends only on $\rho(x,t) \equiv \int_0^\infty f(x,v,t)\,dv$ and T depends only on P. If f_0 factors as follows: $f_0(x,v,t) = \rho(x,t)p_0(v)$, where $p_0(v)$ is the probability density according to which preferred velocities are distributed among vehicle drivers in general, a steady-state, position-independent solution can be calculated [27]. (There is, however, an obvious weakness in the conjecture concerning f_0: it overlooks the fact that fast vehicles tend to crowd at places and at times such that passing is difficult, i.e. in situations where the density $\rho(x,t)$ is high.)

The elementary model can be extended by considering a stretch of a freeway of finite length, including entrance and exit ramps, and by making more refined assumptions concerning the desired velocity distribution function and the parameters P and T. But, in spite of all these possible modifications, the applicability remains limited to low density traffic situations. This restriction is implied by the assumptions made in the derivation of the interaction term. In fact, the restriction is already hidden in the basic idea of using a kind of Boltzmann equation to describe freeway traffic flow.

2.4 Conclusions concerning the choice of the type of model

Recall from Chapter 1 the objective of modelling high density freeway traffic flow. The concentration on high density traffic situations disqualifies the Prigogine-Boltzmann models for reasons mentioned above. The next classes of models to be dismissed are the car following theories and the headway models, because their usefulness is restricted to single lane traffic. Extension of these models to multilane traffic enormously complicates the analysis.

The first conclusion, therefore, is that macroscopic continuum models seem the best choice for our purposes. But, as we have seen, the inclusion of a stochastic term in the model equations is at the same time a desirable and complicating feature of such a model. Linearization of the model equations offers a way to avoid too complex models, even when a stochastic term is added. The question is: does linearization of the model equations restrict the applicability of the model too much or is it an acceptable limitation?

In our view the handicap is not too serious. It seems that all models that have been designed up till now have their own limited area of application, because each model incorporates the characteristic features of only some specific traffic situations. The linearized stochastic continuum model that will be presented in Chapter 5 applies specifically to high density stationary freeway traffic flow. In case stationarity is a too restrictive requirement alternative models might be preferable. If for example large deterministic effects are dominating the flow — one may think of accidents, which, once occurred, have deterministic impact, and reduction of the number of lanes — a deterministic continuum model like the one proposed by P. Ross [34] is a good candidate. If one wishes to retain randomness and non-linearity at the same time, discretization of the space variable into a relatively small number of road sections is an alternative — in some cases even one single section [36].

If, however, the questions we ask concern the capacity of freeways and the phenomenon of spontaneous congestion — a congestion for which there is no clear cause except the high density — the assumption of stationary high density traffic flow may very well be an appropriate one.

Of course, the obvious objection can be made that the appearance of congestion ruins the stationarity of the traffic flow and therefore cannot be described by a model assuming stationarity. However, the onset of congestion can be seen as a limit or a breakdown of the model, so that a linearized model does have something to say on the subject of congestion.

Chapter 3

Stochastic integration in separable Hilbert spaces

3.1 Preliminaries

Before we consider stochastic integrals in separable, real Hilbert spaces in Section 3.2, we briefly discuss some of the ingredients needed there. These ingredients are: general Gaussian families, standard one-dimensional Brownian motion and Itô's definition of the (one-dimensional) stochastic integral. First we cite some facts about Gaussian families from Hida's book on Brownian motion [13].

Definition 3.1 *A Gaussian family is a collection of real valued random variables $\{X_\lambda : \lambda \in \Lambda\}$, such that every finite linear combination of elements of $\{X_\lambda\}$ has a Gaussian distribution. Λ is an arbitrary index set.*

By analogy with finite families of random variables we call $\{m(\lambda) = E(X_\lambda)\colon \lambda \in \Lambda\}$ the mean vector of the Gaussian family and $\{V(\lambda,\mu) = E(X_\lambda - m(\lambda))(X_\mu - m(\mu))\colon \lambda,\mu \in \Lambda\}$ its covariance matrix.

Theorem 3.1 (Theorem 1.10 in [13]) *Given a set $\{m(\lambda)\colon \lambda \in \Lambda\}$ and a real positive definite 'matrix' $\{V(\lambda,\mu)\colon \lambda,\mu \in \Lambda\}$, there exists a unique Gaussian family $\{X_\lambda : \lambda \in \Lambda\}$, the mean vector and covariance matrix of which coincide with $\{m(\lambda)\}$ and $\{V(\lambda,\mu)\}$, respectively.*

Proposition 3.1 (Proposition 1.10 in [13]) *Let $\{X_\lambda : \lambda \in \Lambda\}$ be a Gaussian family. Then a necessary and sufficient condition for $\{X_\lambda :$*

$\lambda \in \Lambda\}$ to be independent is $V(\lambda,\mu) = 0$ for every $\lambda, \mu \in \Lambda$ $(\lambda \neq \mu)$. A necessary and sufficient condition for a member X_μ of the family to be independent of $\{X_\lambda : \lambda \in \Lambda,\ \lambda \neq \mu\}$ is $V(\mu,\lambda) = 0$ for all $\lambda \neq \mu$.

Next, we define a standard one-dimensional Brownian motion with respect to a given stochastic set-up. Let $(\Omega, \mathcal{F}, P)$ be a complete probability space with a right-continuous increasing family of sub-σ-algebra's, $\{\mathcal{F}_t : t \geq 0\}$, each containing all P-null sets. We say that the set-up $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$ satisfies the usual conditions.

Definition 3.2 *A collection of random variables $\{b(t) : t \geq 0\}$ is called a standard one-dimensional $\{\mathcal{F}_t\}$-Brownian motion, if*

(i) $\{b(t) : t \geq 0\}$ is a real-valued process and $b(0) = 0$ almost surely;

(ii) $\{b(t) : t \geq 0\}$ is $\{\mathcal{F}_t\}$-adapted, i.e. for every t it holds that $b(t)$ is $\mathcal{F}_t$-measurable;

(iii) for all s and t such that $0 \leq s \leq t$ it holds that $b(t) - b(s)$ is independent of $\mathcal{F}_s$ and $b(t) - b(s)$ is a Gaussian random variable with mean 0 and variance $t - s$.

The existence of the standard one-dimensional Brownian motion is guaranteed by Theorem 3.1. If we take $\Lambda = [0, \infty)$, $m(t) = 0$ and $V(s,t) = s \wedge t$, Theorem 3.1 provides the corresponding Gaussian family, which we denote by $\{b(t) : t \geq 0\}$. Defining, for all t, $\mathcal{F}_t$ as the completion of $\sigma(b(s) : 0 \leq s \leq t)$ with respect to the probability measure of the family, we can verify property (iii) of Definition 3.2. An immediate consequence of property (iii) is that the increments of the Brownian motion are independently distributed.

Definition 3.2 is phrased in terms of a more general probability space than the one that is given by Theorem 3.1, because often the Brownian motion is not the only 'source of randomness'.

It can be shown that each $\{b(t) : t \geq 0\}$ has a sample-continuous version (i.e. all sample-paths are continuous). Henceforth we will assume that we only deal with such sample-continuous versions.

The third issue of this section is the definition of the one-dimensional stochastic integral. As usual $L^2([0,T] \times \Omega)$ denotes the Hilbert space of all real measurable processes $f : [0,T] \times \Omega \longrightarrow \mathbf{R}$ that satisfy $\|f\|^2 \equiv E\int_0^T f^2(t)\,dt < \infty$. T is some constant. The associated inner product is of course given by: $\langle f, g\rangle = E\int_0^T f(t)g(t)\,dt$. In the spirit of Ikeda and Watanabe [14], we define $A^2([0,T] \times \Omega)$ (or A^2) as the subspace of $L^2([0,T] \times \Omega)$ that consists of all processes f that are adapted to $\{\mathcal{F}_t : t \geq 0\}$. A^2 is a closed subspace of the space $L^2([0,T] \times \Omega)$. Further we denote by A_0^2 the subspace of A^2 consisting of all 'stepfunctions', i.e. each $f \in A_0^2$ can be written as:

$$f(t,\omega) = \sum_{i=0}^{m-1} f_i(\omega) 1_{(t_i,t_{i+1}]}(t),$$

where $\{t_i\}$ is a partition of $[0,T]$ such that $t_0 = 0$ and $t_m = T$ for some integer m and for all i $(0 \leq i \leq m-1)$ f_i is an $\mathcal{F}_{t_i}$-measurable square integrable random variable. Without proof we mention that A_0^2 is dense in A^2 with respect to the norm $\|\cdot\|^2$ [14, p. 46]. We define for $f \in A_0^2$

$$\int_0^T f(t,\omega)\, db(t) \equiv \sum_{i=1}^{m-1} f_i(\omega)[b(t_{i+1}) - b(t_i)].$$

It is obvious that this definition does not depend on the partition. Let g be another element of A_0^2. If we take a partition $\{t_i\}$ that includes the partitions corresponding to f and g, then it is easy to see that

$$E \int_0^T f(t)\, db(t) \int_0^T g(t)\, db(t) = \sum_{i=0}^{m-1} E[f_i g_i (t_{i+1} - t_i)] = E \int_0^T f(t) g(t)\, dt.$$

Note that this calculation owes its simplicity to the fact that by definition the increment $b(t_{i+1}) - b(t_i)$ is independent of $f(t_i)$, so that all cross-terms disappear when taking the expectation. This independence also yields $E \int_0^T f(t)\, db(t) = 0$. We see that the stochastic integral defines an isometry from A_0^2 into $L^2(\Omega)$, which can be extended to all elements $f \in A^2$. This extension is called the stochastic integral of $f \in A^2$ with respect to Brownian motion.

In applications the increment $db(t)$ is usually denoted by the name 'white noise'. Often we will simply write $b(t)$ instead of $\{b(t) : t \geq 0\}$, if it is clear from the context that t does not have some definite value.

3.2 Stochastic integrals in a separable Hilbert space

The content of this and the next section has been taken from work by Dawson [8], Funaki [9], Itô [15] and Yor [39]. As before let the set-up $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$ satisfy the usual conditions. H will denote a real separable Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|$.

Definition 3.3 *A linear random functional on H is a linear map from H to $L_0(\Omega,\mathcal{F},P)$, where $L_0(\Omega,\mathcal{F},P)$ is the space of all real random variables on $(\Omega,\mathcal{F},P)$. We call a family of linear random functionals $\{B_t : t \geq 0\}$ on H a cylindrical Brownian motion on H if it satisfies the following condition: for every $x \in H$ $(x \neq 0)$ $t \mapsto B_t(x)/\|x\|$ is a standard Brownian motion (adapted to the filtration $\{\mathcal{F}_t\}$).*

Again the existence is established by invoking Theorem 3.1: we set $m(t,x) = 0$ and $V(t,x,s,y) = (s \wedge t)\langle x,y\rangle$. The Gaussian family we obtain is denoted by $\{B_t(x) : t \geq 0, x \in H\}$. If we define, for all $t \geq 0$, $\mathcal{F}_t$ as the completion of $\sigma(B_t(x) : 0 \leq s \leq t,\ x \in H)$ with respect to the probability measure of the family, the properties mentioned in the definition easily follow. As before the filtration $\{\mathcal{F}_t\}$ may contain more information than merely the information generated by the Gaussian family.

We remark that Proposition 3.1 implies that for every set $\{(t_i,x_i) : i \geq 1,\ t_i \geq 0,\ x_i \in H,\ \langle x_i,x_j\rangle = 0 \text{ if } i \neq j\}$ the associated family of centred Gaussian random variables $\{B_{t_i}(x_i)\}$ is independent.

We will now use for the first time the separability of H; consider a complete orthonormal system in H: $\{e_i : i \geq 1\}$. Let x be an arbitrary element of H: $x = \sum_i \alpha_i e_i$. As a consequence of the orthonormality $\{B_t(e_i) : i \geq 1\}$ is a family of independent standard one-dimensional Brownian motions, so that we can rewrite $B_t(x) = \sum_{i=1}^{\infty} \alpha_i B_t(e_i)$ as $B_t(x) = \sum_{i=1}^{\infty}\langle x,e_i\rangle b_i(t)$, where $\{b_i(t) : i \geq 1\}$ is a family of independent standard one-dimensional Brownian motions. We have thus obtained a general representation for a cylindrical Brownian motion, which we will use below in several calculations.

Definition 3.4 *Let $A^2([0,T] \times \Omega, H)$ be the Hilbert space of all H-valued, $\mathcal{F}_t$-adapted and measurable functions $f(t,\omega)$ satisfying $E\int_0^T \|f(t)\|^2\,dt < \infty$. For every $f \in A^2([0,T] \times \Omega, H)$ we define*

$$\int_0^T \langle f(t), dB_t\rangle \equiv \sum_{i=1}^{\infty} \int_0^T \langle f(t), e_i\rangle\, dB_t(e_i),$$

where $e_i, i \geq 1$ is a complete orthonormal system in H.

The definition is straightforward, if f is concentrated, uniformly in ω, on a finite number of e_i. For such f the integral defines an isometric mapping:

$$E\left|\int_0^T \langle f(t), dB_t\rangle\right|^2 = \sum_{i=1}^{n} E\Big(\int_0^T \langle f(t), e_i\rangle\, dB_t(e_i)\Big)^2$$

$$= \sum_{i=1}^{n} E \int_0^T \langle f(t), e_i \rangle^2 \, dt = E \int_0^T \|f(t)\|^2 \, dt,$$

whence it can be extended to all elements of $A^2([0,T] \times \Omega, H)$.

Definition 3.5 *Let $\mathcal{L}_2(H)$ denote the Hilbert space of Hilbert-Schmidt operators on H with norm $\|\cdot\|_{HS}$ and let $A^2([0,T] \times \Omega, \mathcal{L}_2(H))$ be the Hilbert space of all $\mathcal{L}_2(H)$-valued, $\mathcal{F}_t$-adapted and measurable functions $F(t,\omega)$ satisfying $E\int_0^T \|F(t)\|_{HS}^2 \, dt < \infty$.*
For every $F \in A^2([0,T] \times \Omega, \mathcal{L}_2(H))$ an H-valued stochastic integral $\int_0^T F(t) \, dB_t$ is defined by the following equality:

$$\left\langle \int_0^T F(t) \, dB_t, x \right\rangle = \int_0^T \langle F^*(t)x, dB_t \rangle, \qquad \forall x \in H,$$

where $F^*(t)$ is the adjoint of $F(t)$. By linearity it is enough to let x run through the sequence $\{e_i : i \geq 1\}$.

This definition too is straightforward, if F is, in a sense, finite: $Fe_i = 0$ for all $i > n$. For such F we again obtain an isometry:

$$\begin{aligned} E\|\int_0^T F(t) \, dB_t\|^2 &= E\sum_{i=1}^{n}\left\langle \int_0^T F(t) \, dB_t, e_i \right\rangle^2 \\ &= \sum_{i=1}^{n} E \int_0^T \langle F^*(t)e_i, dB_t \rangle^2 = \sum_{i=1}^{n} E \int_0^T \|F^*(t)e_i\|^2 \, dt \\ &= E \int_0^T \sum_{i=1}^{n} \|F^*(t)e_i\|^2 \, dt = E \int_0^T \|F(t)\|_{HS}^2 \, dt. \end{aligned}$$

By extension the integral is defined for all $F \in A^2([0,T] \times \Omega, \mathcal{L}_2(H))$.

One might ask whether it is possible to view B_t as an infinite-dimensional random variable. It is obvious that B_t is not H-valued: set in Definition 3.5 F equal to the projection onto the subspace of H spanned by the first n elements of the orthonormal basis. Then F approaches the identity operator as n tends to infinity. But, as $E\|\int_0^T F(t) \, dB_t\|^2 = E\int_0^T \|F(t)\|_{HS}^2 \, dt = nT$, the integral clearly does not converge to an element of H, which we could identify with B_t. However, we can construct a Hilbert space V, into which H can be densely embedded, such that B_t is V-valued. Let us be specific [39, p. 61]; V is defined as the space of all real sequences $(h_1, h_2, \ldots)$, that satisfy $\sum_{i=1}^{\infty} a_i h_i^2 < \infty$, where $\{a_i : i \geq 1\}$ is some sequence of real numbers,

$a_i > 0$ for all i and $\sum_{i=1}^{\infty} a_i < \infty$. Denote $(h_1, h_2, \ldots)$ by h and let $g = (g_1, g_2, \ldots)$. Then an inner product is defined by $\langle h, g\rangle_V \equiv \sum_{i=1}^{\infty} a_i h_i g_i$. With this inner product V is a real separable Hilbert space. H is densely embedded into V by means of a map $u : H \longrightarrow V$; if $x = \sum_{i=1}^{\infty} \alpha_i e_i$, then $u(x) = (\alpha_1, \alpha_2, \ldots)$.

Now we can identify B_t as an element of V in the following form: $B_t = (b_1(t), b_2(t), \ldots)$, where $b_i(t) \equiv B_t(e_i)$. Indeed, we have that $E\|B_t\|_V^2 = E\sum_i b_i^2(t) a_i = t\sum_i a_i < \infty$. This construction shows that we can replace the representation given earlier, $B_t(x) = \sum_{i=1}^{\infty}\langle x, e_i\rangle b_i(t)$, by

$$B_t = \sum_{i=1}^{\infty} e_i b_i(t), \tag{3.1}$$

provided we keep in mind that B_t is V-valued instead of H-valued. (Nevertheless, the space V will seldom be explicitly mentioned, as it its construction is to some extent arbitrary.)

This representation of B_t also yields an equivalent definition of the stochastic integral $\int_0^T F(t)\, dB_t$ (see for example [6]). Define

$$\int_0^T F(t)\, dB_t = \sum_i \int_0^T F(t) e_i \, db_i(t).$$

$\int_0^T F(t) e_i \, db_i(t)$ is in fact a one-dimensional (Hilbert space valued) stochastic integral. Its interpretation is completely analogous to the real valued one-dimensional stochastic integral. The Hilbert-Schmidt property makes the sum convergent and substitution into definition 3.5 shows the equivalence. We will often use this alternative definition. It is especially convenient when $F(t)e_i$ takes a simple form.

Another remark concerns the definitions of this section. We have defined the H-valued stochastic integral $\int_0^T F(t)\, dB_t$ in a rather indirect way. Of course it is also possible to define the stochastic integral directly using stepfunctions as in Section 3.1. (For an example of the last approach see [15], where Itô defines stochastic integrals in rather general infinite-dimensional spaces.) The definitions put forward in this section have been chosen because they require less background in functional analysis.

3.3 Some properties of the stochastic integral $\int_0^T F(t)\,dB_t$

We list some properties of the stochastic integral defined in Definition 3.5:

(i) $\int_0^T F(t)\,dB_t$ is a martingale, i.e. $\forall i$

$$E\Big(\Big\langle\int_0^t F(s)\,dB_s, e_i\Big\rangle|\mathcal{F}_{t_0}\Big) = \Big\langle\int_0^{t_0} F(s)\,dB_s, e_i\Big\rangle \quad t \geq t_0,$$
$$\Big\langle\int_0^t F(s)\,dB_s, e_i\Big\rangle \text{ is integrable and}$$
$$\Big\langle\int_0^t F(s)\,dB_s, e_i\Big\rangle \text{ is } \{\mathcal{F}_t\}\text{-adapted}$$

or, in other words, $\langle\int_0^t F(s)\,dB_s, e_i\rangle$ is a one-dimensional martingale for each i;

(ii) an infinite-dimensional Itô-formula holds [8,39];

(iii) if $\int_0^T (E\|F(t)\|_{HS}^{2p})^{1/p}\,dt < \infty$, then for $p = 1, 2, \ldots$ there exists a positive constant $C = C(p)$ such that $E\|\int_0^T F(t)\,dB_t\|^{2p} \leq C(\int_0^T (E\|F(t)\|_{HS}^{2p})^{1/p}\,dt)^p$;

(iv) a well-known martingale inequality (Doob's inequality, see for example [14, p. 110] and [23, p. 95]) yields for this special case

$$\begin{aligned} E\big(\sup_{0\leq t\leq T} \|\int_0^t F(s)\,dB_s\|^{2p}\big) &\leq \Big(\frac{2p}{2p-1}\Big)^p E\,\|\int_0^T F(s)\,dB_s\|^{2p} \\ &\leq C\,E\int_0^T \|F(s)\|_{HS}^{2p}\,ds, \end{aligned}$$

where C depends only on $p = 1, 2, \ldots$;

(v) the following Fubini theorem holds: if $F(t, s, \omega)$ is a $\mathcal{L}_2(H)$-valued measurable function ($F\colon [0,T] \times [0,T] \times \Omega \longrightarrow \mathcal{L}_2(H)$) satisfying $E\int_0^T\int_0^T \|F(t,s,\cdot)\|_{HS}^2\,dt\,ds < \infty$ and if, for every s, $F(t,s)$ is $\{\mathcal{F}_t\}$-adapted, then, almost surely,

$$\int_0^T\int_0^T F(t,s)\,dB_t\,ds = \int_0^T\int_0^T F(t,s)\,ds\,dB_t.$$

3.4 White noise

In Chapter 4 we will apply the theory presented here to the so-called stochastic heat equation and also to comparable stochastic partial differential equations. Throughout these applications we set $H = L^2[0, M]$, where M is an arbitrary constant. In this context the derivative of the cylindrical Brownian motion, dB_t, is also called (two-dimensional) white noise. This can be explained in the following way. Consider the rectangle $[0, T] \times [0, M]$. If the two-dimensional white noise is to be a generalization of the one-dimensional white noise, we should require that integration of the noise over small, disjoint subsets of $[0, T] \times [0, M]$ yields independent Gaussian random variables with mean zero and variance equal to the Lebesgue measure of the subsets. (cf. the definition of the one-dimensional Brownian motion and its derivative in Section 3.1). We can verify these properties as follows: first we identify the noise integrated over a small rectangle $(t_1, t_2] \times (x_1, x_2]$ with the random variable $B_{t_2}(1_{(x_1,x_2]}) - B_{t_1}(1_{(x_1,x_2]})$ and, equivalently, the noise integrated over $(s_1, s_2] \times (y_1, y_2]$ with $B_{s_2}(1_{(y_1,y_2]}) - B_{s_1}(1_{(y_1,y_2]})$. Note that these random variables are centred Gaussian random variables. Secondly we calculate the covariance of the random variables (using that $EB_t(x)B_s(y) = (s \wedge t)\langle x, y\rangle$).

$$\begin{aligned} &E(B_{t_2}(1_{(x_1,x_2]}) - B_{t_1}(1_{(x_1,x_2]}))(B_{s_2}(1_{(y_1,y_2]}) - B_{s_1}(1_{(y_1,y_2]})) \\ &= \quad (t_2 \wedge s_2 - t_2 \wedge s_1 - t_1 \wedge s_2 + t_1 \wedge s_1)\langle 1_{(x_1,x_2]}, 1_{(y_1,y_2]}\rangle. \end{aligned}$$

This expression equals 0 if $(t_1, t_2]$ and $(s_1, s_2]$ or $(x_1, x_2]$ and $(y_1, y_2]$ are disjoint, which implies the independence of the two random variables. Also it is clear from equating the two random variables that the variance equals $(t_2 - t_1)(x_2 - x_1)$.

Chapter 4

The stochastic heat equation

4.1 Evolution equations and semigroup theory

The (deterministic) heat equation is an example of evolution equations, that can be solved using semigroup theory. Therefore, in the first section of this chapter, some generally known facts about evolution equations and semigroup theory are presented. We follow the exposition that Goldstein gives in his book on semigroups of linear operators [10, pp. 3–25 and pp. 83–91].

An evolution equation in continuous time usually consists of an ordinary or partial differential equation supplemented with an initial value (and sometimes boundary values as well). The general equation in its simplest form reads

$$\frac{df(t)}{dt} = A(f(t)) \qquad t \geq 0\,; \tag{4.1}$$

$f(t)$ describes the state of some (physical) system and $Af(t)$ is the rate of change of the system; the initial value is assumed to be given: $f(0) = f_0$. If this equation is to make sense, we must specify the space in which f takes values as well as the mapping A. We will henceforth assume that f takes values in a real or complex Banach space, S, having norm $\|\cdot\|$ and that A is a closed linear (generally unbounded) operator from its domain $\mathcal{D}(A)$ in S to S (often this is abbreviated to: A is an operator on S). Recall that the operator A is closed if its graph $\mathcal{G}(A) = \{(f, Af): f \in \mathcal{D}(A)\}$ is a closed subspace of $S \times S$, or equivalently, if $f_n \to f$,

$f_n \in \mathcal{D}(A)$, and $Af_n \to g$ together imply that $f \in \mathcal{D}(A)$ and $Af = g$ [17, pp. 292, 293]. Usually, A is unbounded and $\mathcal{D}(A)$ is dense in S.

Definition 4.1 *A semigroup U on S is a family of bounded linear operators, $\{U_t : t \geq 0\}$, satisfying*

(i) $U_t U_s = U_{t+s}$*;*

(ii) $U_0 = I$ *(I is the identity operator);*

(iii) the map $t \mapsto U_t f$ is continuous for each $f \in S$.

U is called a contraction semigroup if for all $t \geq 0$ U_t is a contraction, i.e. $\|U_t\| \leq 1$.

The next definition and the subsequent theorem partially describe the relation between closed linear operators on S and semigroups on S.

Definition 4.2 *Let $U = \{U_t : t \geq 0\}$ be a semigroup on S. The (infinitesimal) generator A of U is defined by the formula*

$$Af = \lim_{t \downarrow 0} \frac{U_t f - f}{t} = \frac{d}{dt} U_t f \Big|_{t=0}$$

where the domain $\mathcal{D}(A)$ of A is the set of all f for which the above limit exists in the norm $\|\cdot\|$ of the space S.

Theorem 4.1 (Hille-Yosida[10]) *A linear operator A on S is the generator of a contraction semigroup if and only if A is closed, $D(A)$ is dense in S, $(0, \infty) \subset \rho(A)$ and $\|\lambda(\lambda I - A)^{-1}\| \leq 1$ for all $\lambda > 0$. $\rho(A)$ denotes the resolvent set of A: $\rho(A) = \{\lambda \in \mathbf{C}$ or $\mathbf{R} : (A - \lambda I)^{-1}$ exists and is densily defined and bounded$\}$ (see [17, p. 371] and [10, p. 13]).*

Goldstein also states a theorem that gives the more general conditions A has to satisfy if it is to generate an arbitrary semigroup that does not consist of contractions [10, p. 20]. As one might guess by now, the condition that A generates a (contraction) semigroup is sufficient for obtaining a solution of the equation (4.1). In fact, if A generates a semigroup, we can draw the conclusion, that for each $f_0 \in \mathcal{D}(A)$ there is a unique solution $f : [0, \infty) \longrightarrow \mathcal{D}(A)$ in $C^1([0, \infty), S)$ (cf. [10, p. 83]), given by $f(t) = U_t f_0$. Often the semigroup is such that $U_t S \subset \mathcal{D}(A)$ for $t > 0$. If this is true, then the conclusion holds that for each $f_0 \in S$ (4.1) has a unique solution $f \in C([0, \infty), S) \cap C^1((0, \infty), S)$, which is (again) given by $f(t) = U_t f_0$.

Example 4.1 (One-dimensional heat equation) As a first application consider the following simple equation (a one-dimensional heat equation)

$$\frac{\partial f}{\partial t} = \frac{\partial^2 f}{\partial x^2}, \tag{4.2}$$

where f is a function of two variables, t and x. Because we intend to use semigroup theory, we restrict $f(t)$ (which, for fixed t, is a function of x only) to take values in the function space $L^2[0, M]$. Instead of (4.2) we write

$$\frac{df(t)}{dt} = \frac{d^2}{dx^2} f(t), \tag{4.3}$$

which formula more clearly resembles (4.1). To solve the equation initial and boundary values have to be supplied. Let us choose $f(0) = f_0 \in L^2[0, M]$ and $f(t, 0) = f(t, M) = 0$. We set $A \equiv d^2/dx^2$ with $\mathcal{D}(A) = \{f \in L^2[0, M] : f'' \in L^2[0, M] \text{ and } f(0) = f(M) = 0\}$. Now A is a closed, densely defined, linear operator; A has a pure point spectrum, its eigenfunctions are $\{e_i : e_i(x) = \sqrt{2/M} \sin(\pi i x/M), i \geq 1\}$ with eigenvalues $\{-\pi^2 i^2/M^2 : i \geq 1\}$, so that $(0, \infty) \subset \rho(A)$, and, as can be verified using the eigenfunctions, also $\|\lambda(\lambda - A)^{-1}\| \leq 1$ for all $\lambda > 0$. All conditions of Theorem 4.1 are satisfied; let U denote the semigroup generated by A. As A and U_t commute on $\mathcal{D}(A)$ for all t and the eigenfunctions are in $\mathcal{D}(A)$, we can conclude that $U_t e_i = \exp(-\lambda_i t) e_i$, where $\lambda_i = \pi^2 i^2/M^2$ for $i \geq 1$. Thus we obtain the following representation of U_t:

$$U_t f_0 = U_t \Big(\sum_{i=1}^{\infty} \langle f_0, e_i \rangle e_i\Big) = \sum_{i=1}^{\infty} \langle f_0, e_i \rangle \exp(-\lambda_i t)\, e_i.$$

Also $U_t S \subset \mathcal{D}(A)$ for $t > 0$ and so we have that $f(t) = U_t f_0$ is the unique solution of the equation (4.3).

Next, consider the following type of evolution equations:

$$\frac{df(t)}{dt} = Af(t) + h(t), \tag{4.4}$$

where $h \in C([0, \infty), S)$, $f_0 \in S$ and A generates, as before, a semigroup for which $U_t S \subset \mathcal{D}(A)$ for $t > 0$. Now the unique solution, $f \in C([0, \infty), S) \cap C^1((0, \infty), \mathcal{D}(A))$, is, as can easily be verified (cf. linebreak [10, pp. 84, 85]), given by

$$f(t) = U_t f_0 + \int_0^t U_{t-s} h(s)\, ds.$$

Finally, one might consider the rather general evolution equation

$$\frac{df(t)}{dt} = Af(t) + h(t, f(t)). \tag{4.5}$$

Solving this equation requires that we impose rather severe conditions on A, f_0 and h. In order to avoid this it is convenient to slightly change our perspective. We will weaken the notion of solution. Instead of the differential equation (4.5) we will consider the integral equation

$$f(t) = U_t f_0 + \int_0^t U_{t-s} h(s, f(s))\, ds. \tag{4.6}$$

Indeed, any solution of the equation (4.5) also solves the integral equation, whereas the converse need not be true. Often both equations are combined in writing

$$df(t) = Af(t)\, dt + h(t, f(t))\, dt. \tag{4.7}$$

We say that a solution of the equation (4.5) is a *strong* solution of (4.7). A solution of the equation (4.6) is called a *mild* solution of (4.7). We now cite a theorem [10, pp. 89, 90].

Theorem 4.2 *Suppose that A generates a semigroup on the Banach space S, $f_0 \in S$ and $h \in C([0,\infty) \times S \longrightarrow S)$ satisfies a Lipschitz-condition: for each $\tau > 0$ there is a constant $K = K(\tau)$ such that*

$$\|h(t, f) - h(t, g)\| \leq K(\tau)\|f - g\|,$$

whenever $f, g \in S$ and $0 \leq t \leq \tau$. Then the equation (4.7) has a unique continuous (i.e. $f \in C([0,\infty), S)$) mild solution.

Let us conclude this section with some remarks.

(i) Because in Theorem 4.2 we only required a mild solution, the condition that $U_t S \subset \mathcal{D}(A)$ for $t > 0$ could be omitted. Without this condition (4.7) has no strong solution.

(ii) In the last theorem the function h need not be (jointly) continuous; measurability is enough.

(iii) The evolution equations considered here are by no means the most general ones. For example, we have assumed that A is time independent, whereas under certain conditions time dependent operators can generate generalized 'semigroups', the so-called families of evolution operators. Furthermore, perturbations P might be added to the operator A in such a way that $A + P$ also generates a semigroup (or a family of evolution operators, in case P is time dependent). See for these and other extensions for example references [6], [7] and [10].

4.2 Stochastic evolution equations

In this section we will discuss a rather general class of stochastic evolution equations, of which the stochastic heat equation (see (4.11) and (4.12)) is, for us, the most interesting example. The general equation is of the form

$$dX(t) = AX(t)dt + h(t, X(t))\,dt + \sigma(t, X(t))\,dB_t. \tag{4.8}$$

Here $X(t)$ is an H-valued stochastic process (H is a separable real Hilbert space), A is a closed linear operator on H generating some semigroup U_t, $h\colon [0,T] \times H \longrightarrow H$ as well as $\sigma\colon [0,T] \times H \longrightarrow \mathcal{L}(H)$ satisfy certain Lipschitz conditions ($\mathcal{L}(H)$ is the set of all bounded operators on H), and B_t is a cylindrical Brownian motion.

Let us note right away that the introduction of the stochastic term has limited the generality of the evolution equation. The Banach space S has been replaced by the separable real Hilbert space H. This restriction is, of course, due to the fact that we define (4.8) to be equivalent to the integral equation (4.9), where the stochastic integral is interpreted along the lines indicated in Chapter 3. As before a solution of the integral equation is called a mild solution. (However, contrary to the definition of a strong solution given in Section 4.1, in this context a strong solution is not the solution of some corresponding differential equation, but a mild solution satisfying the extra condition that $X(t) \in \mathcal{D}(A)$ for all t.) There is a second restriction caused by the use of the stochastic integral. Recall that the stochastic integral $\int_0^T F(t)\,dB_t$ was defined for Hilbert-Schmidt-operator valued functions F. This implies that either σ or, as becomes clear from (4.9), U_t should be Hilbert-Schmidt-operator valued.

Before we state a few theorems we recall some notation. An operator A on a Hilbert space is said to be Hilbert-Schmidt if $\sum_i \|Ae_i\|^2 < \infty$ for every orthonormal basis of H. As this sum does not depend on the choice of the basis, a corresponding Hilbert-Schmidt norm of A is defined as $\|A\|_2 = (\sum_i \|Ae_i\|^2)^{1/2}$. A is said to be a trace class or nuclear operator if $\|A\|_1^2 \equiv \sum_i |\langle e_i, Ae_i\rangle| < \infty$. The Hilbert space of all Hilbert-Schmidt operators on H is denoted by $\mathcal{L}_2(H)$ (or by $\mathcal{L}_{HS}(H)$) and $\mathcal{L}_1(H)$ denotes the Banach space of all trace-class or nuclear operators on H. The first theorem is a (somewhat adapted) version of Theorem 5.1 in [8].

Theorem 4.3 *Consider the (nonlinear) stochastic evolution equation on a separable, real Hilbert space H:*

$$X(t) = U_t X_0 + \int_0^t U_{t-s} h(s, X(s))\, ds + \int_0^t U_{t-s} \sigma(s, X(s))\, dB_s\,, \quad (4.9)$$
$$X(0) = X_0 \in H, \qquad 0 \le t \le T.$$

Assume the following:

(i) *U is a semigroup, generated by a closed linear operator A;*

(ii) *$h : [0,T] \times H \longrightarrow H$ is continuous and satisfies, uniformly in t, $\|h(t,x) - h(t,y)\| \le C_1 \|x - y\|$ for all $x, y \in H$;*

(iii) *$\sigma : [0,T] \times H \longrightarrow \mathcal{L}(H)$ is continuous and satisfies, uniformly in t, $\|\sigma(t,x) - \sigma(t,y)\| \le C_2 \|x - y\|$ for all $x, y \in H$; $\|\cdot\|$ denotes the norm of the Hilbert space as well as the operator norm of $\mathcal{L}(H)$;*

(iv) *condition (i) or condition (iii) is strengthened: suppose that $-A$ is a positive, self-adjoint operator such that $(-A)^{-1}$ exists and is a nuclear operator; furthermore, assume that $0 < \liminf_i \lambda_i / i^{1+\delta} \le \limsup_i \lambda_i / i^{1+\delta} < \infty$ for some $\delta > 0$, where $\{\lambda_i\}$ is the set of eigenvectors of $-A$; $\{e_i\}$ is the corresponding orthonormal family of eigenvectors;*
or, alternatively, suppose that condition (iii) holds with the operator norm $\|\cdot\|$ replaced by the Hilbert-Schmidt norm $\|\cdot\|_2$.

C_1 and C_2 are arbitrary positive constants.

Then the equation (4.9) has a unique (mild) solution whose sample paths are almost surely continuous from $[0,T]$ into H. The solution also satisfies $\sup_{0 \le t \le T} E\|X(t)\|^{2p} < \infty$ for all integer $p \ge 1$.

Assumption (iv) ensures that the stochastic integral exists as an H-valued random variable. The next theorem elaborates on the nonlinearity. For simplicity we only consider the case where the operator A satisfies the necessary conditions.

Theorem 4.4 *Assume that*

(i) *U is the semigroup generated by a closed linear operator A; $-A$ is a positive, self-adjoint operator such that $(-A)^{-1}$ exists and is a nuclear operator; furthermore, for some $\delta > 0$, $0 < \liminf_i \lambda_i / i^{1+\delta} \le \limsup_i \lambda_i / i^{1+\delta} < \infty$, where $\{\lambda_i\}$ is the set of eigenvectors of $-A$; $\{e_i\}$ is the corresponding orthonormal family of eigenvectors;*

(ii) P is a linear operator with $\mathcal{D}(P^*) \supset \mathcal{D}(A)$ such that $\|P^*e_i\| \leq \gamma\lambda_i^\alpha$, where γ is some constant and $\alpha < 1 - 1/(2(1+\delta))$ or, if P and A commute on $\mathcal{D}(A)$, we require that $\mathcal{D}(P) \supset \mathcal{D}(A)$ and $\|Pe_i\| \leq \gamma\lambda_i^\alpha$;

(iii) $h : [0,T] \times H \longrightarrow H$ is continuous and satisfies, uniformly in t, $\|h(t,x) - h(t,y)\| \leq C_1\|x-y\|$ for all $x, y \in H$;

(iv) $\sigma : [0,T] \times H \longrightarrow \mathcal{L}(H)$ is continuous and satisfies, uniformly in t, $\|\sigma(t,x) - \sigma(t,y)\| \leq C_2\|x-y\|$ for all $x, y \in H$.

C_1 and C_2 are arbitrary positive constants.

Then the stochastic evolution equation

$$dX(t) = AX(t)dt + Ph(t,X(t))\,dt + \sigma(t,X(t))\,dB_t\,, \qquad (4.10)$$
$$X(0) = X_0 \in H, \qquad 0 \leq t \leq T$$

has a unique mild solution whose sample paths are almost surely continuous from $[0,T]$ into H. Furthermore, the sample paths are Hölder-continuous from $[\epsilon,T]$ into H for all $\epsilon > 0$. The solution also satisfies $\sup_{0\leq t\leq T} E\|X(t)\|^{2p} < \infty$ for all integer $p \geq 1$.

PROOF. The proof is included in Appendix A. □

Example 4.2 (Stochastic heat equation) We take up the example discussed in Section 4.1. If we add a noise term consisting in a cylindrical Brownian motion to the equation, we obtain

$$dX(t) = \frac{d^2}{dx^2}X(t)\,dt + \sigma(t,X(t))\,dB_t\,, \qquad (4.11)$$
$$X(0) = X_0 \in L^2[0,M], \qquad 0 \leq t \leq T.$$

This equation is called a (one-dimensional) stochastic heat equation. As before set $H \equiv L^2[0,M]$ and $A \equiv d^2/dx^2$ with $\mathcal{D}(A) = \{f \in L^2[0,M]: f'' \in L^2[0,M]$ and $f(0) = f(M) = 0\}$. Again U denotes the semigroup A generates. $(-A)^{-1}$ exists and is a nuclear operator; the eigenfunctions of $-A$ are $\{e_i : e_i(x) = \sqrt{2/M}\sin(\pi i x/M),\ i \geq 1\}$ with eigenvalues $\{\lambda_i = \pi^2 i^2/M^2 : i \geq 1\}$. It is obvious that $\lim_i \lambda_i/i^2 = \pi^2 < \infty$. We conclude that, if σ satisfies condition (iii) of Theorem 4.3, (4.11) has a unique mild solution satisfying $\sup_{0\leq t\leq T} E\|X(t)\|^{2p} < \infty$ for all integer $p \geq 1$, whose sample paths are almost surely continuous from $[0,T]$ into H.

Example 4.3 (Stochastic heat equation with convection) Consider the following extension of (4.11):

$$dX(t) = \frac{d^2}{dx^2}X(t)dt - \frac{d}{dx}h(t, X(t))\,dt + \sigma(t, X(t))\,dB_t\,, \quad (4.12)$$
$$X(0) = X_0 \in L^2[0, M], \qquad 0 \le t \le T,$$

where h and σ satisfy condition (iii) and (iv) of Theorem 4.4, respectively. The term $d/dx\, h(t, X(t))\, dt$ is sometimes called 'convection' term. To apply Theorem 4.4, choose A as in the previous example. Verification of condition (ii) is easy: the operators d/dx and d^2/dx^2 commute and $\|d/dx\, e_i\| = \|\lambda_i^\alpha e_i\|$ for $\alpha = 0.5$; further $\alpha < 1 - 1/2((1 + \delta))$ for $\alpha = 0.5$ and $\delta = 1$. Again we draw the conclusion that the equation has a unique mild solution, whose sample paths are almost surely continuous from $[0, T]$ into H and that satisfies $\sup_{0 \le t \le T} E\|X(t)\|^{2p} < \infty$ for all integer $p \ge 1$.

4.3 The stochastic heat equation and stationary traffic flow

The last section of this chapter has two objectives. The first aim to motivate why a (particular) stochastic heat equation might be a good model of stationary traffic flow. The second one is to present some additional properties of this stochastic heat equation.

4.3.1 On boundary values and noise term

Instead of discussing the (one-dimensional) stochastic heat equation as presented in the last example of the previous section we consider a modification that is suitable for the application we have in mind. After presenting the modification we motivate the choice.

In the first place we set $h(t, X(t)) = c_0 X(t)$, where c_0 is a constant, and $\sigma(t, X(t)) = \sigma$, where σ is a fixed bounded linear operator on $H = L^2[0, M]$. Secondly, we redefine the operator A: $A \equiv K d^2/dx^2$ with $\mathcal{D}(A) = \{f \in L^2[0, M] : f'' \in L^2[0, M],\ f(0) = f(M),\ f'(0) = f'(M)\}$, where K is some positive constant. These changes lead to the equation

$$dX(t) = K\frac{d^2}{dx^2}X(t)dt - c_0\frac{d}{dx}X(t)\,dt + \sigma\,dB_t\,, \quad (4.13)$$
$$X(0) = X_0 \in L^2[0, M], \qquad 0 \le t \le T.$$

The redefined operator A does not quite satisfy the conditions of Theorems 4.3 or 4.4. The set of eigenvectors of A constitutes an orthonormal basis, which we denote by $\{e_i : i \geq 0\}$. It consists of three subsets, $\{e_0\}$, $\{\phi_i : i \geq 1\}$ and $\{\psi_i : 1 \geq 1\}$, where $e_0 = 1/\sqrt{M}$, $\phi_i(x) = \sqrt{2/M}\sin(2\pi ix/M)$ and $\psi_i(x) = \sqrt{2/M}\cos(2\pi ix/M)$. But e_0 has eigenvalue 0. This implies that A^{-1} cannot be defined on the whole space $L^2[0,M]$. It can be seen, however, that this is not a serious restriction. It is enough if A^{-1} exists, and is nuclear, on $\overline{\mathcal{R}(A)}$.

There is an obvious reason for partitioning the sequence $\{e_i : i \geq 1\}$ into two parallel sequences: for each $i \geq 1$ there corresponds to ϕ_i and to ψ_i the same eigenvalue, viz. $\lambda_i = 4\pi^2 i^2/M^2$. Working with ϕ_i and ψ_i makes calculations easier.

Recall from Chapter 2, page 12 that a linear stochastic continuum model of traffic flow might look like

$$\frac{\partial R}{\partial t} = K\frac{\partial^2 R}{\partial x^2} - c_0\frac{\partial R}{\partial x} + \text{noise term},$$

where $R(t,x)$ denotes the fluctuation of the density of the traffic around some mean value. We claim that (4.13) is a suitable first attempt to give a precise reformulation of this intuitive model equation, provided that σ is properly chosen. Three arguments will give support to this claim.

First of all the choice of h makes the heat equation linear. Secondly, as will be shown in the next subsection, the choice of the domain of A ensures stationarity with respect to the space variable. Intuitively, this property can be understood by noting that each choice of $\mathcal{D}(A)$ corresponds to a particular set of boundary conditions. In this case

$$X(t,0) = X(t,M) \text{ and } \frac{\partial X}{\partial x}(t,0) = \frac{\partial X}{\partial x}(t,M) \quad \forall t.$$

(Note, however, that the second boundary condition needs careful interpretation as, in general, $X(t,x)$ is not differentiable with respect to x.) We have identified the endpoints of the interval $[0,M]$. Speaking in traffic flow terms we might say that the vehicles are supposed to drive on a (large) circular road of circumference M. This 'picture' of traffic flow may at first seem unrealistic, yet as M grows larger the picture becomes more realistic.

The third argument concerns the noise term. If there is no specific knowledge about the noise structure, white noise seems to be the appropriate 'default' choice; σ then would equal a constant (times the identity operator). Such a choice leads to relatively easy equations.

There is, however, one major objection against truly white noise: it entails a violation of the principle of 'conservation of vehicles'. The simplest way to meet this objection is the following: write $B_t = \sum_i b_i(t)e_i$, where $\{e_i\}$ is the orthonormal basis of eigenvectors of A and remove the term corresponding to e_0. Recall that e_0 is the constant eigenfunction ($e_0 = 1/\sqrt{M}$). We then have $B_t = \sum_{i\geq 1} b_i^s(t)\phi + b_i^c(t)\psi$, where $\{b_i^s(t), b_i^c(t)\}$ is a collection of independent standard Brownian motions; $\phi_i(x) = \sqrt{2/M}\sin(2\pi ix/M)$ and $\psi_i(x) = \sqrt{2/M}\cos(2\pi ix/M)$. Of course, the 's' and the 'c' refer to 'sine' and 'cosine', respectively. We could also say that σ equals a constant times a projection operator (that projects onto the subspace of H spanned by all ϕ_i and ψ_i, i.e. $\sigma f = \sigma_0 \sum_{i\geq 1}\langle f, \phi_i\rangle\phi_i + \langle f, \psi_i\rangle\psi_i$ for $f \in H$). This choice of the noise term ensures that at every instant the noise integrated over the interval $[0, M]$ is zero. Thus the noise constantly redistributes the traffic without creating new vehicles or destroying existing ones.

Probably the best way of capturing the modification of the noise term is to consider, instead of $L^2[0, M]$, the Hilbert space $L_0^2[0, M] = \{f \in L^2[0, M] : \langle f, e_0\rangle = 0\}$. Obviously, the set $\{e_i : i \geq 1\}$ consisting of the subsets $\{\phi_i : i \geq 1\}$ and $\{\psi_i : 1 \geq 1\}$ constitutes the orthonormal family of eigenvectors of A, if $\mathcal{D}(A) = \{f \in L_0^2[0, M] : f'' \in L^2[0, M],\ f(0) = f(M),\ f'(0) = f'(M)\}$. The noise introduced above is exactly the cylindrical Brownian motion on $L_0^2[0, M]$. In this setting σ is simply a constant. Furthermore, A now satisfies entirely the conditions of Theorem 4.4, as A^{-1} is defined on all of $L_0^2[0, M]$.

4.3.2 Continuity and stationarity

We now have a closer look at (4.13). We write hereafter R instead of X because of the specific application we have in mind. Again we set $A = Kd^2/dx^2$ and $\mathcal{D}(A) = \{f \in L^2[0, M]: f'' \in L^2[0, M],\ f(0) = f(M),\ f'(0) = f'(M)\}$. But now σf equals $\sigma_0 \sum_{i\neq 0}\langle f, e_i\rangle e_i$ for all $f \in H$ and R_0 satifies $\langle R_0, e_0\rangle = 0$, i.e. $R_0 \in L_0^2[0, M]$ ($L_0^2[0, M] = \{f \in L^2[0, M]: \langle f, e_0\rangle = 0\}$). If no confusion is to be expected we will often simply write σ instead of σ_0. All this yields the equation

$$dR(t) = K\frac{d^2}{dx^2}R(t)dt - c_0\frac{d}{dx}R(t)\,dt + \sigma\,dB_t\,, \qquad (4.14)$$
$$R(0) = R_0, \qquad 0 \leq t \leq T.$$

According to Theorem 4.4 the equation (4.14) has a unique mild solution whose sample paths are almost surely continuous from $[0, T]$ into $H =$

$L^2[0, M]$. Furthermore, the sample paths are Hölder-continuous from $[\epsilon, T]$ into H for all $\epsilon > 0$. The solution also satisfies $\sup_{0 \le t \le T} E\|X(t)\|^{2p} < \infty$ for all integer $p \ge 1$.

It can be shown quite easily that the operator $Kd^2/dx^2 - c_0\, d/dx$ also generates a semigroup. We expect that the solution of (4.14) can also be written as

$$R(t) = V_t R_0 + \int_0^t V_{t-s} \sigma dB_s, \tag{4.15}$$

where V is the semigroup generated by $A_c = Kd^2/dx^2 - c_0\, d/dx$ (with $\mathcal{D}(A_c) = \mathcal{D}(A)$). The proof of the next proposition shows that this is true; it uses the equivalence in order to provide a more direct proof of the joint Hölder-continuity of $R(t, x)$. V has a rather simple characterization.

$$V_t e_i(x) = e^{-\lambda_i K t} e_i(x - c_0 t),$$

where $\{e_i\}$ is the same orthonormal basis as above; note, however, that the $\{e_i\}$ are not eigenvectors of A_c. For general $f \in L^2[0, M]$ we have

$$\begin{aligned} V_t f(x) &= \int_0^M q(t, x, y) f(y)\, dy \qquad \text{with} \\ q(t, x, y) &= \frac{1}{M} + \sum_{i=1}^{\infty} [\phi_i(x - c_0 t)\phi_i(y) + \psi_i(x - c_0 t)\psi_i(y)] e^{-\lambda_i K t}. \end{aligned}$$

Proposition 4.1 *Write $R(t, x)$ instead of $R(t)$ to indicate that we view each realization of R as a real valued function of two parameters, t and x. Then $R(t, x)$ is almost surely jointly Hölder-continuous on $[\epsilon, T] \times [0, M]$ for every $\epsilon > 0$.*

PROOF. See Appendix A. □

Definition 4.3 *Let $X(t)$ be an H-valued stochastic process with n-dimensional time parameter. H is a separable Hilbert space endowed with its Borel σ-algebra. X is called stationary if the finite joint distributions of X are invariant under time shift. X is called weakly stationary if its covariance functional $r(t, s; f, g)$ defined by $r(t, s; f, g) = E\langle X(t), f\rangle\langle X(s), g\rangle$ for $f, g \in H$, is invariant under time shift. By linearity it is enough if the property holds for all members of an orthonormal basis. If X is a centred Gaussian process these two properties coincide.*

Proposition 4.2 *Suppose R_0 is not a fixed element of $L_0^2[0,M]$, but an $L_0^2[0,M]$-valued random variable that is $\mathcal{F}_0$-measurable. Then there is a unique measure μ for R_0 such that $R(t)$ is a stationary $L_0^2[0,M]$-valued process. Moreover, if R_0 is distributed according to μ, $R(t,x)$ is a stationary (under shifts in $\mathbf{R}^2$) real-valued process.*

PROOF. See Appendix A. □

The last proposition says that, whether or not R_0 is stochastic, asymptotically $R(t,x)$ is a stationary process. As we want to apply the theory brought forward in this chapter to stationary traffic flow, we will henceforth only consider this stationary process.

In the next proposition we give an explicit expression of the stationary process $R(t,x)$ in terms of its parameters K, σ, c_0 and M.

Proposition 4.3 *The stationary solution of (4.14) can be represented as*

$$R(t,x) = \sum_{i=1}^{\infty} a_i^s(t)\phi_i(x - c_0t) + a_i^c(t)\psi_i(x - c_0t),$$

where $\phi_i(x) = \sqrt{2/M}\sin(2\pi ix/M)$ and $\psi_i(x) = \sqrt{2/M}\cos(2\pi ix/M)$; $\{\phi_i : i \geq 1\}$ and $\{\psi_i : 1 \geq 1\}$ together form the set $\{e_i : 1 \geq 1\}$ of eigenvectors of A. Note that the eigenfunction $e_0 = 1/\sqrt{M}$ is excluded. Further, $\{a_i^s(t), a_i^c(t) \colon i \geq 1\}$ is a family of mutually independent Ornstein-Uhlenbeck processes, i.e. each $a_i^s(t)$ $a_i^c(t)$ is a centred stationary Gaussian process having covariance function

$$Ea_i^s(t)a_i^s(s) = Ea_i^c(t)a_i^c(s) = \frac{\sigma^2}{2\lambda_i K}\exp(-\lambda_i K|\Delta|),$$

where $\lambda_i = 4\pi i^2/M^2$ and $\Delta = (t-s)$. This implies that $R(t,x)$ is also a centred stationary Gaussian process. Its covariance function, $r(\Delta, z)$, reads

$$r(\Delta,z) = \frac{\sigma^2 M}{4\pi^2 K}\sum_{i=1}^{\infty}\frac{1}{i^2}\exp(-\lambda_i K|\Delta|)\cos(2\pi i\frac{z - c_0\Delta}{M}), \qquad (4.16)$$

where $\Delta = (t-s)$ and $z = x - y$; (t,x) and (s,y) are two points in the plane.

PROOF. Again the proof is included in Appendix A. □

4.3.3 The noise term revisited

Close inspection of the covariance function, $r(\Delta, z)$, reveals the crucial role of the length M. M (co)determines the amplitude of the function and the appearance of M in the argument of the cosine makes it also the typical length associated with the process. A consequence of the last fact is for example that at any fixed time the values of the process are strongly correlated at points that are at a maximal distance from each other (in formula: $r(0, M/2) = -r(0,0)/2$. This is not exactly what we want. M should be some large parameter, that is of little importance. The cause of the undesirable role of M stems from a confusion of three distinct lengths. To be able to solve the equation describing the evolution of the density-fluctuations we had to specify boundary conditions. We identified the endpoints of the interval $[0, M]$. In traffic flow terms the boundary conditions say that vehicles are supposed to drive on a (large) circular road. We reserve the symbol M for the length of the circumference. It seems natural to assume that this circumference is much longer than the stretch that we are observing. Therefore, we introduce the observation length, the length of the stretch of the freeway along which we observe the traffic stream. Call this length L. The third length is the most important one. Let us write down again the decomposition of the noise term introduced above: $B_t = \sum_{i\geq 1} b_i^s(t)\phi + b_i^c(t)\psi$, where $\{b_i^s(t), b_i^c(t)\}$ is a collection of independent standard Brownian motions; $\phi_i(x) = \sqrt{2/M}\sin(2\pi i x/M)$ and $\psi_i(x) = \sqrt{2/M}\cos(2\pi i x/M)$. The 's' and the 'c' refer to 'sine' and 'cosine', respectively. It is clear from this representation that the noise-components with low 'i-value' have a large range. For $i = 1$ the range is M, for $i = 2$ $M/2$ and so on. Let us call the maximum range of the noise the disturbance length. We see that for the noise term we used up till now the disturbance length equals the length of the circumference, M. Hereafter we will assume that these two lengths are distinct. The disturbance length will be denoted by S. Furthermore, we will assume that $S \ll L \ll M$ and also that $mS = M$ for some large integer m. The last assumption is included, because it makes calculations easier, while not being very restrictive. Figure 5.1 illustrates the differences between the three lengths.

In order to get a maximum disturbance length S, we decide to delete all noise components, for which the range is larger than S (i.e. for which $i < m$). Thus, the appropriate choice of the noise term becomes $B_t = \sum_{i\geq m} b_i^s(t)\phi + b_i^c(t)\psi$. Calculations analogous to those of the previous subsection show that again $R(t, x)$ is a centred stationary Gaussian process, but now its decomposition into Ornstein-Uhlenbeck processes is given by

$$R(t,x) = \sum_{i=m}^{\infty} a_i^s(t)\phi_i(x - c_0 t) + a_i^c(t)\psi_i(x - c_0 t), \tag{4.17}$$

and its covariance function, $r(\Delta, z)$, reads

$$r(\Delta, z) = \frac{\sigma^2 M}{4\pi^2 K} \sum_{i=m}^{\infty} \frac{1}{i^2} \exp(-\lambda_i K|\Delta|) \cos(2\pi i \frac{z - c_0\Delta}{M}), \tag{4.18}$$

where $\Delta = (t - s)$, $z = x - y$ and (t,x) and (s,y) are two points in the plane. In this covariance function M not really plays a prominent role. In fact its influence is largely compensated by the appearance of m, at least if we assume that S is constant, so that M and m are always proportional. It will be shown below that, when m tends to infinity for constant S, the covariance function as well as the process $R(t,x)$ converge in a well defined way to a limit process.

First, we investigate what has happened to the 'whiteness' of the white noise as a consequence of the deletion of all 'low frequency' components. Obviously, whiteness in time direction is preserved. Therefore, it is sufficient to calculate the covariance of $B_t(1_{(x_1,x_2]})$ and $B_t(1_{(y_1,y_2]})$.

$$\begin{aligned} &EB_t(1_{(x_1,x_2]})B_t(1_{(y_1,y_2]}) \\ &= t\Big(\langle 1_{(x_1,x_2]}, 1_{(y_1,y_2]}\rangle - \langle 1_{(x_1,x_2]}, e_0\rangle\langle 1_{(y_1,y_2]}, e_0\rangle \\ &\quad - \sum_{1\le i\le m-1} (\langle 1_{(x_1,x_2]}, \phi_i\rangle\langle 1_{(y_1,y_2]}, \phi_i\rangle + \langle 1_{(x_1,x_2]}, \psi_i\rangle\langle 1_{(y_1,y_2]}, \psi_i\rangle)\Big) \\ &\approx t\Big(\langle 1_{(x_1,x_2]}, 1_{(y_1,y_2]}\rangle - 2\frac{(x_2 - x_1)(y_2 - y_1)}{S}\Big), \end{aligned}$$

where the last approximate equality is valid if $(x_2 - x_1)$, $(y_2 - y_1)$, $|y_2 - x_2|$ and $|y_1 - x_1|$ are all much smaller than S. This means that on a scale much smaller than the disturbance length the noise is approximately white in space direction too. This is, of course, hardly surprising; the deletion of the noise components having large 'wave lengths' should not strongly influence the small scale noise structure. Let us remark, in conclusion, that the approximate equality is exact if $S = M$, if we only delete the noise component corresponding to e_0.

4.3.4 Convergence of R for m tending to infinity

As m will be variable from now on, let us denote henceforth the stationary stochastic process by $R_m(t,x)$ or R_m and reserve the symbol

$R(t,x)$ or R for the limit process to which R_m converges as m tends to infinity. The same remark holds for the covariance functions r_m and r. We rewrite r_m as follows (with $j = i/m$):

$$r_m(\Delta, z) = \frac{\sigma^2 S}{4\pi^2 K} \sum_j \frac{1}{j^2} \exp(-\frac{4\pi^2 j^2}{S^2} K|\Delta|) \cos(2\pi j \frac{z - c_0\Delta}{S}) \frac{1}{m},$$

where j takes on the values 1, $1 + 1/m$, $1 + 2/m$ etcetera.

Lemma 4.1 *The covariance function r_m concerges pointwise to the covariance function r, which is defined as*

$$r(\Delta, z) = \frac{\sigma^2 S}{4\pi^2 K} \int_1^\infty \frac{1}{l^2} \exp(-\frac{4\pi^2 l^2}{S^2} K|\Delta|) \cos(2\pi l \frac{z - c_0\Delta}{S})\, dl. \quad (4.19)$$

PROOF. $r_m(\Delta, z)$ clearly is a Riemann-sum converging to the integral as $m \to \infty$. □

Theorem 4.5 *Let R be the stationary, centred, Gaussian process characterized by the covariance function r. Then R_m converges in distribution to R as m tends to infinity, if R_m as well as R are viewed as $C([0,T] \times [0,L])$-valued random variables, where $C([0,T] \times [0,L])$ is endowed with the Borel σ-algebra.*

PROOF. See Appendix A. □

Note that the convergence in distribution is related to the function space $C([0,T] \times [0,L])$. This means that for large m R_m and R are almost indiscernible, if we can only partially observe the processes. If for example the observation length would grow proportionally to m, the convergence would be destroyed.

Chapter 5

Presentation of the model

Let us now summarize the results of the Chapters 2, 3 and 4 presenting once again the ideas put forward in these chapters, without entering into the mathematical details.

In Chapter 2 we motivated our choice of a linear stochastic continuum model to describe freeway traffic flow. The Chapters 3 and 4 dwelt on stochastic integrals and on a specific example of the stochastic heat equation. These mathematical expositions were meant to give a precise and adequate meaning to the noise term appearing in the proposed stochastic continuum model (see (1.1) in Chapter 1, page 3 and (2.7) in Chapter 2, page 12).

The proposed linear stochastic continuum model is, formally, written as

$$\frac{\partial R}{\partial t} = K\frac{\partial^2 R}{\partial x^2} - c_0\frac{\partial R}{\partial x} + \sigma\frac{\partial B}{\partial t}, \tag{5.1}$$

although we must keep in mind that the solution is not differentiable, neither with respect to time nor with respect to space. $R(t,x)$ denotes the deviation of the density (or the density-fluctuation) at time t and location x around some fixed reference-value. The evolution equation describes how the density evolves in time in response to three causes (the three terms of the right hand side): a smoothing term corresponding to the fact that in dense traffic vehicles normally tend to drive at approximately equidistant spacings and at comparable speeds, the convection term caused by the displacement of the traffic stream (with velocity c_0) and a noise term. The noise term is derived from two-dimensional white noise, but is constrained, at each instant, to sum to zero over a 'disturbance-length' S. This constraint might be written, again using a formal notation, as $\int_{x_0}^{x_0+S} \partial B/\partial t\, dx \approx 0$ for every x_0. K and σ are

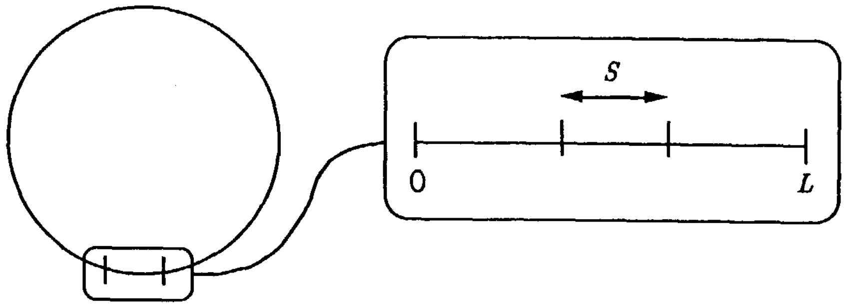


Figure 5.1: The vehicles are driving on a circular road with circumference M. We observe the traffic on a small portion of the circle of length L, $L \ll M$. The maximum length of the disturbances S is in turn much smaller than L.

positive parameters, that determine the strength of the smoothing term and the noise term, respectively.

In order to be able to solve the evolution equation initial and boundary conditions had to be imposed. They were chosen in such a way that a stationary solution could be obtained. We have supposed, mainly for technical reasons, that the space variable should be confined to a finite interval $[0, M]$. To ensure stationarity in space we have identified the endpoints of the interval. The boundary conditions are such that the boundaries are actually nonexistent. Translated into traffic flow terms the boundary conditions say that vehicles are supposed to drive on a (large) circular road of circumference M (see Figure 5.1). We assumed that $M = mS$ for some large integer m. The initial value is chosen to be a stochastic variable, distributed according to some probability measure μ, so that the process $R(t, x)$ is stationary in time as well.

As a consequence of the linearity of the equation, the properties of the noise term and the initial and boundary values the process $R(t, x)$ is a real-valued centred Gaussian process with a two dimensional 'time-parameter' (t, x). It can be represented as

$$R(t,x) = \sum_{i=m}^{\infty} a_i^s(t)\phi_i(x - c_0 t) + a_i^c(t)\psi_i(x - c_0 t),$$

where $\{a_i^s(t), a_i^c(t) : i \geq m\}$ is a family of mutually independent Ornstein-Uhlenbeck processes, i.e. each $a_i(t)$ or $b_i(t)$ is a centred stationary Gaussian process having covariance function ($\Delta = t - s$)

$$Ea_i^s(t)a_i^s(s) = Ea_i^c(t)a_i^c(s) = \frac{\sigma^2 M}{8\pi^2 K i^2}\exp(-\frac{4\pi^2 K i^2}{M^2}|\Delta|),$$

and $\phi_i(x) = \sqrt{2/M}\sin(2\pi i x/M)$ and $\psi_i(x) = \sqrt{2/M}\cos(2\pi i x/M)$. The covariance function of the process itself is given by

$$r(\Delta, z) = \frac{\sigma^2 M}{4\pi^2 K}\sum_{i=m}^{\infty}\frac{1}{i^2}\exp(-\lambda_i K|\Delta|)\cos(2\pi i\frac{z - c_0\Delta}{M}), \qquad (5.2)$$

where $\Delta = t - s$, $z = x - y$ and (t, x) and (s, y) are two points in the plane.

The role played by M and m is in fact rather unimportant, given that they are large. It turns out that the process $R_m(t, x)$ (the suffix indicates the dependence of the process on m) converges to a limit process, denoted by $R(t, x)$, which is completely determined by its covariance function

$$r(\Delta, z) = \frac{\sigma^2 S}{4\pi^2 K}\int_1^{\infty}\frac{1}{l^2}\exp(-\frac{4\pi^2 l^2}{S^2}K|\Delta|)\cos(2\pi l\frac{z - c_0\Delta}{S})\,dl. \qquad (5.3)$$

This limit process will be the starting point for the chapters to follow. Its unknown parameters are K, σ, S and, possibly, c_0. Estimation of these parameters can be performed on the basis of measurements during a time interval $[0, T]$, over a space interval $[0, L]$. We assume that L is chosen such that $S \ll L$. There is one complication. We do not observe the process R directly, but we observe the sum of some 'mean' value R_0 (not to be confused with an initial value) and the process R, which is to be interpreted as the fluctuation of the density of the traffic around the mean value R_0.

Although not entirely in agreement with the linearization, it seems sensible with a view to the application we have in mind to assume that R_0 may be (slowly) varying in time. This dependence on time may compensate for a small drift in the mean traffic stream density. If we denote the observed density by $\tilde{R}(t, x)$, then

$$\tilde{R}(t, x) = R_0(t) + R(t, x).$$

Chapter 6

Statistical analysis

In this chapter we consider the stationary, centred, Gaussian process, $R(t, x)$, which is completely determined by the covariance function

$$r(\Delta, z) = \frac{\sigma^2 S}{4\pi^2 K} \int_1^\infty \frac{1}{l^2} \exp\left(-\frac{4\pi^2 l^2}{S^2} K|\Delta|\right) \cos\left(2\pi l \frac{z - c_0\Delta}{S}\right) dl. \quad (6.1)$$

The unknown parameters are K, σ, S and, possibly c_0. They are explained in for example Chapter 5.

As indicated at the end of Chapter 5 we observe in practice the sum, $\tilde{R}(t, x)$, of some slowly varying 'mean' value $R_0(t)$ and the process $R(t, x)$, which is to be interpreted as the fluctuation of the density of the traffic around this mean value. We assume, however, that we can extract with great precision the process $R(t, x)$ from $\tilde{R}(t, x)$. Analogously we assume that the parameter c_0 is known. It can be determined in advance with the same precision as $R_0(t)$.

6.1 Observing the process on a time-space rectangle

We introduce two new constants: $A = \sigma^2 S/(4\pi^2 K)$ and $a = 4\pi^2 K/S^2$. In fact, these constants replace K and σ as parameters that are to be estimated.

Estimation of these alternative parameters is performed on the basis of one measurement during a time interval $[0, T]$, over a space interval $[0, L]$. We assume that the fixed observation length L is chosen such that $S \ll L$. T will be 'asymptotically' large. The idea is to estimate first the covariances of the process $R(t, x)$ for a finite number of (Δ, z)-values;

in a second step we will apply a (non-linear) regression procedure to these estimates. It is sufficient if we only consider nonnegative values of Δ. Furthermore, we will restrict ourselves to values of z that are nonnegative also. This is reasonable as the relatively large values of the covariances occur when Δ and z have the same sign (at least, if we assume that c_0 is positive).

The mean value $R_0(t)$ is estimated by $\bar{R}(t) = 1/L\int_0^L \tilde{R}(t,x)\,dx$. Thus $R(t,x)$ is extracted from $\tilde{R}(t,x)$ writing

$$\tilde{R}(t,x) - \bar{R}(t) = R(t,x) - \frac{1}{L}\int_0^L R(t,x)\,dx$$

and noting that $1/L\int_0^L R(t,x)\,dx$ is negligible if $S \ll L$.

An obvious estimator of $r(\Delta, z)$ is

$$\hat{r}(\Delta,z) \equiv \frac{1}{T-\Delta}\frac{1}{L-z}\int_0^{T-\Delta}\int_0^{L-z} R(t,x)R(t+\Delta,x+z)\,dx\,dt. \quad (6.2)$$

This estimator possesses some 'standard' properties. It is unbiased, and furthermore it is asymptotically normally distributed. The last statement needs justification.

Proposition 6.1 *The estimator $\hat{r}$ is asymptotically normally distributed:*

$$\frac{1}{\tau\sqrt{T-\Delta}}\int_0^{T-\Delta}(Q(t,\Delta,z) - r(\Delta,z))\,dt \xrightarrow{\text{d}} N(0,1) \quad \text{as } T\to\infty,$$

where $Q(t,\Delta,z) \equiv 1/(L-z)\int_0^{L-z} R(t,x)R(t+\Delta,x+z)\,dx$ *and*

$$\tau^2 = 2\int_0^\infty E(Q(0,\Delta,z) - r(\Delta,z))(Q(t,\Delta,z) - r(\Delta,z))\,dt.$$

PROOF. According to Billingsley ([3, p. 376]) we only have to show that the process $Q(t)$ is α-mixing. A process $X(t)$ is said to be α-mixing if for all t and all $A_1 \in \sigma(X_s : s \le t)$ and $A_2 \in \sigma(X_s : s \ge t+u)$ we have the following

$$|P(A_1 \cap A_2) - P(A_1)P(A_2)| \le \alpha(u),$$

where $\alpha(u) = O(u^{-5})$. It is enough to verify that the underlying process $R(t)$ is α-mixing, because every set in for example $\sigma(Q_s : s \le t)$ also belongs to the σ-algebra $\sigma(R_s : s \le t)$.

Let us define the process $Z(t)$: $Z(t,x) = T_{-t}R(t,x) \equiv R(t,x+c_0t)$. (See for the definition of T_t also the proof of Lemma A.4.) It is fairly easy to prove that $Z(t)$ is α-mixing. Because for example $\{T_s : s \leq t\}$ corresponds to a (bi)measurable mapping of $\sigma(Z_s : s \leq t)$ onto $\sigma(R_s : s \leq t)$, $R(t)$ has the same property.

The α-mixing-property will be shown in three of steps. First we note that Z can be approximated by a process Z_m that takes values in $L^2[0, M]$. Secondly, the process Z_m is written as a countable sum of real valued and independent processes. Finally, we prove that each of these processes is α-mixing; if we call the corresponding functions α_i, we find that $\alpha(t) \equiv \sum_i \alpha_i(t)$ satisfies the required condition.

As was shown before, the process R was obtained as the limit of a sequence of processes denoted by R_m. In the same way Z can be viewed as the limit process of the sequence Z_m, where $Z_m(t) \equiv T_{-t}R_m(t)$. $Z_m(t)$ can be represented as $\sum_{i \geq m} a_i(t)e_i$. Here $\{a_i(t) : i \geq m\}$ is a family of independent Ornstein-Uhlenbeck processes with covariance function $r^{(i)} = \sigma^2 M^2/(8\pi^2 K i^2) \exp(-4\pi^2 K i^2 t/M^2)$; the meaning of the various parameters is the same as before. For each i the sum contains in fact two independent terms, but for simplicity of notation this feature is not made explicit.

We associate probability measures P on $\mathcal{B}(C([0,T] \times [0,L] \rightarrow \mathbf{R}))$ and P_m on $\mathcal{B}(C([0,T] \rightarrow L^2[0,M]))$ with the processes Z and Z_m, respectively. (For ease of notation we identify the probability space and the sample path space.) Note that the second Borel σ-algebra contains the first one. Let $\mathcal{A} = \mathcal{A}(C([0,T] \times [0,L] \rightarrow \mathbf{R}))$ denote the algebra that is generated by the ϵ-balls $B_{f,\epsilon} = \{g \in C([0,T] \times [0,L] \rightarrow \mathbf{R}) : \sup_{t,x} |g(t,x) - f(t,x)| \leq \epsilon\}$. Further $\mathcal{A}_1 = \mathcal{A}(C([0,t] \times [0,L] \rightarrow \mathbf{R}))$, $\mathcal{A}_2 = \mathcal{A}(C([t+u,T] \times [0,L] \rightarrow \mathbf{R}))$, $\mathcal{B}_1 = \mathcal{B}(C([0,t] \times [0,L] \rightarrow \mathbf{R}))$ and $\mathcal{B}_2 = \mathcal{B}(C([t+u,T] \times [0,L] \rightarrow \mathbf{R}))$. For $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$ we have that

$$\begin{aligned} &|P(A_1 \cap A_2) - P(A_1)P(A_2)| \\ \leq\ & |P(A_1 \cap A_2) - P_m(A_1 \cap A_2)| + |P_m(A_1 \cap A_2) - P_m(A_1)P_m(A_2)| \\ +\ & |P_m(A_1)P_m(A_2) - P(A_1)P(A_2)|; \end{aligned}$$

t is arbitrary, but fixed; u as well as T will eventually grow to infinity; the fact that T will be variable along with u implies that m will depend on u, i.e. $m = m(u)$.

As $Z_m \xrightarrow{d} Z$ the first and third term tend to zero. This is true, because for all sets A in the algebra $\mathcal{A}$ it can be shown that $P(\partial A) = 0$. We will prove below that for all $A_1 \in \mathcal{B}_1$ and $A_2 \in \mathcal{B}_2$ it holds that $|P_m(A_1 \cap A_2) - P_m(A_1)P_m(A_2)| \leq \alpha(u)$ for some function α, i.e. Z_m

is α-mixing, so that for $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$ we obtain the desired inequality.

The observation that $\{A_1 \in \mathcal{B}: |P(A_1 \cap A_2) - P(A_1)P(A_2)| \leq \epsilon\}$ for fixed $A_2 \in \mathcal{B}$ and $\{A_2 \in \mathcal{B}: |P(A_1 \cap A_2) - P(A_1)P(A_2)| \leq \epsilon\}$ for fixed $A_1 \in \mathcal{B}$ are monotone classes, implies that the α-mixing property can be extended to all Borel sets. Thus, the problem is reduced to proving the α-mixing property for Z_m. Of course, the function α that will bound the second term on the right hand side of the inequality must be chosen independently of m.

Let π_i denote the projection of $C([0,T] \to L^2[0,M])$ onto $C([0,T] \to \mathbf{R})$, that picks out the i^{th} Fourier coefficient of $f \in C([0,T] \to L^2[0,M])$ (i.e. $\pi_i(f) = \langle f, e_i \rangle$). We decompose $A \in \mathcal{B}(C([0,T] \to L^2[0,M]))$ as

$$A = A^0 \cap A^1 \cap \ldots \cap A^i \cap \ldots,$$

where $A^i \equiv \pi_i^{-1}(\pi_i(A)) \equiv \{f \in C([0,T] \times [0,L] \to \mathbf{R}): \pi_i(f) \in \pi_i(A)\}$. The independence of the processes $\{a_i(t)\}$ implies that $P_m(A) = P_m(A^0)$ $P_m(A^1)P_m(A^2)\ldots.$

Using this factorization we obtain the following bound:

$$\begin{aligned}
&|P_m(A_1 \cap A_2) - P_m(A_1)P_m(A_2)| \\
&\quad \leq \; |P_m(A_1^m \cap A_2^m) - P_m(A_1^m)P_m(A_2^m)| \\
&\quad + \; |P_m(A_1^{m+1} \cap A_2^{m+1}) - P_m(A_1^{m+1})P_m(A_2^{m+1})| \\
&\quad + \; \ldots
\end{aligned}$$

We see that we only have to obtain an appropriate bound for each separate term, which means in fact proving the α-mixing property for all Ornstein-Uhlenbeck processes $\{a_i : i \geq m\}$ separately. Following Rosenblatt ([33, p. 74]) we argue that a_i is α-mixing if

$$\sup_{\xi \in \sigma(a_i(s):s \leq t), \eta \in \sigma(a_i(s):s \geq t+u)} |E\xi\eta| \leq \alpha_i(u),$$

where $E\xi = E\eta = 0$, $\sigma_\xi^2 = \sigma_\eta^2 = 1$ and α_i is such that $\alpha_i(u) = O(u^{-5})$. (Of course, here the expectation E is taken with respect to the probability measure P_m.) Indeed, if $\kappa = 1_{A_1^i} - P_m(A_1^i)$ and $\lambda = 1_{A_2^i} - P_m(A_2^i)$, the inequality implies that

$$\begin{aligned}
&|P_m(A_1^i \cap A_2^i) - P_m(A_1^i)P_m(A_2^i)| \\
&\quad = \; |E\kappa\lambda| \\
&\quad \leq \; \alpha_i(u)\sqrt{P_m(A_1^i) - P_m(A_1^i)^2}\sqrt{P_m(A_2^i) - P_m(A_2^i)^2} \\
&\quad \leq \; \alpha_i(u).
\end{aligned}$$

Because of the fact that the process a_i is normally distributed, it is enough to check the property for all finite linear combinations of $a_i(s_k)$ (see Rosenblatt [33, p. 74–76]). Put $\xi = \sum_k \gamma_k a_i(s_k)$ and $\eta = \sum_l \delta_l a_i(t_l)$. We obtain that

$$\begin{aligned}
|E\xi\eta| &= |E\sum_{k,l} \gamma_k \delta_l a_i(s_k) a_i(t_l)| \\
&= \Big|\sum_{k,l} \gamma_k \delta_l \frac{\sigma^2 M^2}{8\pi^2 K i^2} \exp(-\frac{4\pi^2 K i^2 (t_l - s_k)}{M^2})\Big| \\
&= \exp(-\frac{4\pi^2 K i^2 u}{M^2}) \Big|\sum_{k,l} \gamma_k \delta_l \frac{\sigma^2 M^2}{8\pi^2 K i^2} \exp(-\frac{4\pi^2 K i^2 (t_l - s_k - u)}{M^2})\Big| \\
&= \exp(-\frac{4\pi^2 K i^2 u}{M^2}) |E\xi\overleftarrow{\eta}\,| \le \exp(-\frac{4\pi^2 K i^2 u}{M^2}),
\end{aligned}$$

where $\overleftarrow{\eta}$ is defined as $\overleftarrow{\eta} = \sum_l \delta_l a_i(t_l - u)$.

We define $\alpha_i(u) \equiv \exp(-4\pi^2 K i^2 u / M^2)$; this choice of α_i is sufficient as

$$\begin{aligned}
\sum_{i \ge m} \alpha_i(u) &= \sum_{i \ge m} \exp(-\frac{4\pi^2 K i^2 u}{M^2}) \\
&\le \exp(-\frac{4\pi^2 K u}{S^2}) \sum_{i \ge m} \exp(-\frac{4\pi^2 K (i^2 - m^2) u}{M^2}) \\
&\le C \exp(-\frac{4\pi^2 K u}{S^2}).
\end{aligned}$$

Obviously, $\alpha(u) \equiv \sum_{i \ge m} \alpha_i(u)$ satisfies the condition $\alpha(u) = O(u^{-5})$. □

Suppose we want to estimate the covariances at a number of points $\{(\Delta_i, z_i)\}$, where $0 \le i \le n$. Assume that the points are numbered such that $\Delta_i \le \Delta_j$ if $i < j$ and denote $\max\{z_i\}$ by z_0. In order to be able to compare the estimates we redefine $Q(t, \Delta, z)$: $Q(t, \Delta, z) \equiv 1/(L - z_0) \int_0^{L - z_0} R(t, x) R(t + \Delta, x + z)\, dx$. Thus the estimator reads

$$\hat{r}(\Delta_i, z_i) \equiv \frac{1}{T - \Delta_n} \frac{1}{L - z_0} \int_0^{T - \Delta_n} \int_0^{L - z_0} R(t, x) R(t + \Delta_i, x + z_i)\, dx\, dt. \tag{6.3}$$

The estimates are (asymptotically) normally distributed with mean $r(\Delta_i, z_i)$. But the error terms are not independent nor identically distributed. Let us write τ_{ij} for the covariance of the error terms.

If we apply the central limit theorem of the proposition to $\hat{r}(\Delta_i, z_i)$, $\hat{r}(\Delta_j, z_j)$ and $\hat{r}(\Delta_i, z_i) + \hat{r}(\Delta_j, z_j)$, respectively, we obtain three asymptotic variances, viz. τ_{ii}, τ_{jj} and, say, ς^2, where

$$\begin{aligned}\varsigma^2 &= 2\int_0^\infty E\left[Q(0,\Delta_i,z_i) + Q(0,\Delta_j,z_j) - r(\Delta_i,z_i)r(\Delta_j,z_j)\right] \\ &\qquad\qquad \left[Q(t,\Delta_i,z_i) + Q(t,\Delta_j,z_j) - r(\Delta_i,z_i)r(\Delta_j,z_j)\right] dt.\end{aligned}$$

At the same time we have that $\varsigma^2 = \tau_{ii} + \tau_{jj} + 2\tau_{ij}$, so that

$$\begin{aligned}\tau_{ij} &= \int_0^\infty E\left[Q(0,\Delta_i,z_i) - r(\Delta_i,z_i))(Q(t,\Delta_j,z_j) - r(\Delta_j,z_j)\right] dt \\ &+ \int_0^\infty E\left[Q(0,\Delta_j,z_j) - r(\Delta_j,z_j))(Q(t,\Delta_i,z_i) - r(\Delta_i,z_i)\right] dt.\end{aligned}$$

We will now give an approximate evaluation of this general (asymptotic) covariance τ_{ij} with $i \leq j$. An approximation will be given, because exact evaluation is far too complicated. Fortunately, an approximation seems to be enough for our purposes.

First we can rewrite the formula for τ_{ij} as follows:

$$\begin{aligned}\tau_{ij} &= \int_0^\infty \left[EQ(0,\Delta_i,z_i)Q(t,\Delta_j,z_j) - r(\Delta_i,z_i)r(\Delta_j,z_j)\right] dt \\ &+ \int_0^\infty \left[EQ(0,\Delta_j,z_j)Q(t,\Delta_i,z_i) - r(\Delta_j,z_j)r(\Delta_i,z_i)\right] dt.\end{aligned}$$

For $EQ(0,\Delta_i,z_i)Q(t,\Delta_j,z_j)$ we obtain the following expression

$$\begin{aligned}&EQ(0,\Delta_i,z_i)Q(t,\Delta_j,z_j) \\ &= \frac{1}{L-z_0}\frac{1}{L-z_0}\int_0^{L-z_0}\int_0^{L-z_0} \\ &\qquad ER(0,x)R(\Delta_i,x+z_i)R(t,y)R(t+\Delta_j,y+z_j)\,dx\,dy,\end{aligned}$$

and an equivalent one for $EQ(0,\Delta_j,z_j)Q(t,\Delta_i,z_i)$, so that

$$\begin{aligned}\tau_{ij} &= \int_0^\infty \frac{1}{L-z_0}\frac{1}{L-z_0}\int_0^{L-z_0}\int_0^{L-z_0} \\ &\qquad \Big[ER(0,x)R(t,y)\; ER(\Delta_i,x+z_i)R(t+\Delta_j,y+z_j) \\ &+ \quad ER(0,x)R(t+\Delta_j,y+z_j)\; ER(t,y)R(\Delta_i,x+z_i) \\ &+ \quad ER(0,x)R(t,y)\; ER(\Delta_j,x+z_j)R(t+\Delta_i,y+z_i) \\ &+ \quad ER(0,x)R(t+\Delta_i,y+z_i)\; ER(t,y)R(\Delta_j,x+z_j)\Big]\,dx\,dy\,dt,\end{aligned}$$

using the fact that for jointly normally distributed random variables X_1, X_2, X_3 and X_4 it holds that $E(X_1X_2X_3X_4) = E(X_1X_2)E(X_3X_4) + E(X_1X_3)E(X_2X_4) + E(X_1X_4)E(X_2X_3)$.

The integrand contains four terms. Let us consider the first one and approximate the integral of this term only, using the expression of the covariance function given above. We obtain the following

$$
\begin{aligned}
& A^2 \int_0^\infty \frac{1}{(L-z_0)^2} \int_0^{L-z_0} \int_0^{L-z_0} \\
& \int_1^\infty \int_1^\infty \frac{1}{l^2}\frac{1}{m^2} \exp(-atl^2 - a(t+\Delta_j - \Delta_i)m^2) \\
& \Big(\cos(2\pi l \frac{y-x-c_0t}{S}) \\
& \cos(2\pi m \frac{y+z_j-x-z_i-c_0t-c_0(\Delta_j-\Delta_i)}{S})\Big)\, dl\, dm\, dx\, dy\, dt.
\end{aligned}
\tag{6.4}
$$

We exchange the order of integration: for fixed t, l and m we first integrate with respect to x and y. Writing

$$
\begin{aligned}
& \cos(2\pi l \frac{y-x-c_0t}{S}) \cos(2\pi m \frac{y+z_j-x-z_i-c_0t-c_0(\Delta_j-\Delta_i)}{S}) \\
& = \frac{1}{2}\Big(\cos(2\pi(l+m)\frac{y-x-c_0t}{S} + 2\pi m \frac{z_j-z_i-c_0(\Delta_j-\Delta_i)}{S}) + \\
& \qquad \cos(2\pi(l-m)\frac{y-x-c_0t}{S} - 2\pi m \frac{z_j-z_i-c_0(\Delta_j-\Delta_i)}{S})\Big)
\end{aligned}
$$

makes it clear that the integral will practically vanish except for those values of l and m such that the second cosine on the right hand side has an argument close to zero for all x and y. Indeed, integration of the right hand side of the above expression with respect to x and y yields

$$
\begin{aligned}
& \frac{S^2}{8\pi^2(l+m)^2} \cos(2\pi(l+m)\frac{c_0t}{S} - 2\pi m \frac{z_j-z_i-c_0(\Delta_j-\Delta_i)}{S}) \\
& [1-\cos(2\pi(l+m)\frac{L-z_0}{S})] \\
+ \quad & \frac{S^2}{8\pi^2(l-m)^2} \cos(2\pi(l-m)\frac{c_0t}{S} + 2\pi m \frac{z_j-z_i-c_0(\Delta_j-\Delta_i)}{S}) \\
& [1-\cos(2\pi(l-m)\frac{L-z_0}{S})].
\end{aligned}
$$

If we multiply the absolute value of the first term by the factor $1/m^2$ $\exp(-a(t+\Delta_j-\Delta_i)m^2)$ and integrate it with respect to m, we get a contribution that, after multiplication by $1/(L-z_0)^2$, is of order $S^2/(L-z_0)^2$. The second term is also multiplied by the factor $1/m^2 \exp(-a(t+\Delta_j-\Delta_i)m^2)$, and integrated with respect to m, but now m is restricted to $(1,\infty)\setminus(l-S/(2\pi(L-z_0)), l+S/(2\pi(L-z_0))$. This results in a contribution, that, after multiplication by $1/(L-z_0)^2$, is roughly equal to

$$\frac{S}{8\pi(L-z_0)}\frac{1}{l^2}\exp(-a(t+\Delta_j-\Delta_i)l^2)\cos(2\pi l\frac{z_j-z_i-c_0(\Delta_j-\Delta_i}{S}).$$

We assumed that $c_0t \ll L-z_0$, so that $2\pi(l-m)c_0t/S$ is virtually zero for all relevant m-values. The assumption is reasonable, as large values of t do not substantially contribute to the integral (6.4). The third contribution is due to the second term, when

$$|l-m| \lessapprox \frac{S}{2\pi(L-z_0)},$$

and approximately equals (using the same assumption on c_0t)

$$\frac{2S}{2\pi(L-z_0)}\frac{1}{l^2}\exp(-a(t+\Delta_j-\Delta_i)l^2)\cos(2\pi l\frac{z_j-z_i-c_0(\Delta_j-\Delta_i}{S});$$

$2S/(2\pi(L-z_0))$ is the 'width' of the peak of the integrand integrated with respect to x and y. We conclude that integration with respect to x, y and m and subsequent multiplication by $1/(L-z_0)^2$ give that the integral (6.4) roughly equals

$$C_1\int_0^\infty\int_1^\infty \frac{1}{l^4}\exp(-a(2t+\Delta_{ij})l^2)\cos(2\pi l\frac{z_j-z_i-c_0\Delta_{ij}}{S})\,dl\,dt$$

with $C_1 = A^2S/(\pi(L-z_0))$ and $\Delta_{ij} = \Delta_j+\Delta_i$. Performing the same approximation for the remaining three terms of the original integrand, we obtain

$$\begin{aligned}
\tau_{ij} \approx\ & C_1\int_0^\infty\int_1^\infty \Big[\frac{1}{l^4}\exp(-a[2t+\delta_{ij}]l^2)\cos(2\pi l\frac{z_j-z_i-c_0\delta_{ij}}{S}) \\
+\ & \frac{1}{l^4}\exp(-a[2(t\vee\Delta_i)+\delta_{ij}]l^2)\cos(2\pi l\frac{z_j+z_i-c_0\Delta_{ij}}{S}) \\
+\ & \frac{1}{l^4}\exp(-a[2(t\vee\delta_{ij})-\delta_{ij}]l^2)\cos(2\pi l\frac{z_j-z_i-c_0\delta_{ij}}{S}) \\
+\ & \frac{1}{l^4}\exp(-a[2(t\vee\Delta_j)-\delta_{ij}]l^2)\cos(2\pi l\frac{z_j+z_i-c_0\Delta_{ij}}{S})\Big]\,dl\,dt,
\end{aligned}$$

where $\delta_{ij} = \Delta_j - \Delta_i$ and $\Delta_{ij} = \Delta_j + \Delta_i$. Integration with respect to t is straightforward, so that

$$\begin{aligned} \tau_{ij} &\approx C_2 \int_1^\infty \Big[\frac{1}{l^4}\exp(-a\delta_{ij}l^2)[a\delta_{ij} + \frac{1}{l^2}]\cos(2\pi l\frac{z_j - z_i - c_0\delta_{ij}}{S}) + \\ & \frac{1}{l^4}\exp(-a\Delta_{ij}l^2)[a\Delta_{ij} + \frac{1}{l^2}]\cos(2\pi l\frac{z_j + z_i - c_0\Delta_{ij}}{S})\Big]\, dl. \qquad (6.5) \end{aligned}$$

6.2 Observing the process in time at some fixed values of the space variable

In this section we assume that we are able to observe the process, $R(t,x)$, during some time interval $[0,T]$ at some fixed values of the space variable. Denote these values by $\{x_p : 1 \le p \le k\}$. The covariance function along the direction $x = x_p$ reduces to

$$r(\Delta, 0) = A\int_1^\infty \frac{1}{l^2}\exp(-a\Delta l^2)\cos(2\pi l\frac{c_0\Delta}{S})\, dl.$$

We will estimate the covariance at a number of (possibly equidistant) points along the line $z = 0$. Set $\Delta_i < \Delta_j$ for $1 \le i < j \le n$. We now proceed along the same lines as in the previous section. Thus the estimators of $\{r(\Delta_i, 0)\}$ will be

$$\hat{r}(\Delta_i, 0) \equiv \frac{1}{T - \Delta_n}\frac{1}{k}\int_0^{T-\Delta_n}\sum_{p=1}^{k} R(t, x_p)R(t + \Delta_i, x_p)\, dt. \qquad (6.6)$$

These estimators are unbiased and asymptotically normally distributed. The analog of the proposition of the previous section reads

Proposition 6.2 *The estimators $\{\hat{r}(\Delta_i, 0)\}$ are asymptotically normally distributed:*

$$\frac{1}{\sqrt{T - \Delta_n}}\int_0^{T-\Delta_n}(Q(t,\Delta) - r(\Delta_i, 0))\, dt \xrightarrow{\mathrm{d}} N(0, \tau) \quad \text{as } T \to \infty,$$

where $Q(t,\Delta) = 1/k\sum_{p=1}^{k} R(t, x_p)R(t + \Delta_i, x_p) - r(\Delta_i, 0))$ *and*

$$\begin{aligned} \tau_{ij} &= \int_0^\infty E(Q(0,\Delta_i) - r(\Delta_i, 0))(Q(t,\Delta_j - r(\Delta_j, 0))\, dt \\ &+ \int_0^\infty E(Q(0,\Delta_j) - r(\Delta_j, 0))(Q(t,\Delta_i) - r(\Delta_i, 0))\, dt. \end{aligned}$$

Again we want to find more explicit expressions for τ_{ij}. In the same manner as before we calculate that

$$\begin{aligned}
\tau_{ij} &= \int_0^\infty \frac{1}{k^2}\sum_{p=1}^{k}\sum_{q=1}^{k}\Big[ER(0,x_p)R(t,x_q)\, ER(\Delta_i,x_p)R(t+\Delta_j,x_q) \\
&+ ER(0,x_p)R(t+\Delta_j,x_q)\, ER(t,x_q)R(\Delta_i,x_p) \\
&+ ER(0,x_p)R(t,x_q)\, ER(\Delta_j,x_p)R(t+\Delta_i,x_q) \\
&+ ER(0,x_p)R(t+\Delta_i,x_q)\, ER(t,x_q)R(\Delta_j,x_p)\Big]\, dt.
\end{aligned}$$

At this point we will use in fact the same approximation as in the previous section. Let us consider the first term of the last integral.

$$A^2\int_0^\infty \frac{1}{k^2}\sum_{p=1}^{k}\sum_{q=1}^{k}\int_1^\infty\int_1^\infty \frac{1}{l^2}\frac{1}{m^2}\exp(-atl^2 - a(t+\Delta_j-\Delta_i)m^2)$$
$$\cos(2\pi l\frac{x_q - x_p - c_0 t}{S})\cos(2\pi m\frac{x_q - x_p - c_0(t+\Delta_j-\Delta_i)}{S})\, dl\, dm\, dt.$$

Now summing over p and q will have about the same effect as the integration over x and y before, provided that the range of the x-values is large compared to S and also that $(x_{p+1} - x_p \ll S$ for all p. This statement can be made more precise if we assume that the x-values are equidistant: $x_p = pd$. Then

$$\begin{aligned}
&\sum_{p=1}^{k}\sum_{q=1}^{k}\cos(2\pi l\frac{x_q - x_p - c_0 t}{S})\cos(2\pi m\frac{x_q - x_p - c_0(t+\Delta_j-\Delta_i)}{S}) \\
&= \frac{1}{2}\cos(2\pi m\frac{c_0(\Delta_j-\Delta_i)}{S} + 2\pi(l+m)\frac{c_0 t}{S}) \\
&\quad \Big\{[\frac{1}{2}\frac{\sin((k+1)2\pi(l+m)d/S)}{\tan(\pi(l+m)d/S)} - \cos^2((k+1)\pi(l+m)d)/S]^2 \\
&\quad +[\frac{1}{2}\sin(2k\pi(l+m)d/S) + \frac{\sin^2(k\pi(l+m)d/S)}{\tan(\pi(l+m)d/S)}]^2\Big\} \\
&+ \frac{1}{2}\cos(2\pi m\frac{c_0(\Delta_j-\Delta_i)}{S} + 2\pi(l-m)\frac{c_0 t}{S}) \\
&\quad \Big\{[\frac{1}{2}\frac{\sin((k+1)2\pi(l-m)d/S)}{\tan(\pi(l-m)d/S)} - \cos^2((k+1)\pi(l-m)d/S)]^2 \\
&\quad +[\frac{1}{2}\sin(2k\pi(l-m)d/S) + \frac{\sin^2(k\pi(l-m)d/S)}{\tan(\pi(l-m)d/S)}]^2\Big\}.
\end{aligned}$$

All the terms are small compared to $1/k^2$, unless we have that $|l \pm m| \approx nS/d$, for integer $n \geq 0$. Because of the assumption on the ratio of d and S and because of the fact that only relatively small values of l and m contribute significantly to the integral, we conclude that the relevant contribution arises, when $|l - m| \lesssim S/(2\pi kd)$. Thus we get the same result as in the previous section, the only difference being that we must substitute in (6.5) $z_i = 0$ as well as $z_j = 0$.

6.3 Details of the analysis

In the previous two sections we have discussed how to obtain estimates for $r(\Delta_i, z_i)$ at a number of points $\{(\Delta_i, z_i): 1 \leq i \leq n\}$. We assumed that $\Delta_i \geq 0$, $z_i \geq 0$ and also that the points are numbered such that $\Delta_i \leq \Delta_j$ if $i < j$. Further we have obtained approximate expressions for the covariance structure of these asymptotically normally distributed estimates. Let us summarize the results in the following proposition.

Proposition 6.3 *The estimators (6.3) and (6.6) lead to two regression problems. The first is*

$$Y_i = A \int_1^\infty \frac{1}{l^2} \exp(-a\Delta_i l^2) \cos(2\pi l \frac{z_i - c_0\Delta_i}{S})\, dl + \epsilon_i,$$

where ϵ_i is normally distributed with mean zero and covariance matrix τ_{ij} given by (6.5). The second regression problem is obtained from the first one by substituting $z_i = 0$.

For practical purposes, however, some additional calculations and remarks are appropiate:

- an explicit description of the procedure of validation will be given;
- the regression functions as well as the covariance functions of the error terms have to be rewritten in a form suitable for input into computer programs that perform the regression;
- some remarks on non-linear regression will be made as well as
- remarks on the discretization of the process.

6.3.1 The procedure of estimation and validation

The procedure for testing the validity of the model and for the estimation of the parameters will be as follows:

1. Simulate artificial data sets using the covariance structure given by (6.5).

2. Using a least squares regression procedure, estimate the parameters A, a and S.

3. Calculate the covariance structure from the estimates of the parameters of the error terms.

4. Repeat step 2, now using the estimated covariance matrix instead of the identity matrix.

5. Repeat steps 3 and 4, until convergence is reached.

6. Repeat steps 2 to 5 using artificial data sets obtained from simulating the original process $R(t, x)$.

7. If the steps 2 to 5 are found to be appropiate, apply them to the real data.

8. Repeat step 6, with the parameters A, a and S set to the values found in step 7, in order to get an idea of the precision of the estimates.

9. Examine the (transformed) residuals coming from the regression on the real data. Test them for independence, normality, trends etcetera.

10. Test the (original) process, $R(t, x)$, for normality.

In the next subsections we will comment on these steps.

6.3.2 Series expansions of the regression functions and the covariance functions of the error terms

First we will give some series expansions of the regression functions. In order to simplify the notation we define

$$
\begin{aligned}
I_0(p, b) &= \int_1^\infty \exp(-pl^2)\cos(pbl)\,dl \\
I_1(p, b) &= \int_1^\infty \frac{1}{l}\exp(-pl^2)\sin(pbl)\,dl \\
I_2(p, b) &= \int_1^\infty \frac{1}{l^2}\exp(-pl^2)\cos(pbl)\,dl \\
\ldots \quad & \quad \ldots \\
\ldots \quad & \quad \ldots \\
I_{2n}(p, b) &= \int_1^\infty \frac{1}{l^{2n}}\exp(-pl^2)\cos(pbl)\,dl \quad \text{and}
\end{aligned}
$$

$$I_{2n+1}(p,b) = \int_1^\infty \frac{1}{l^{2n+1}} \exp(-pl^2)\sin(pbl)\,dl,$$

where b is some fixed real constant and p is positive . According to Proposition B.1 in Appendix B these integrals can be expanded in series as follows

$$\begin{aligned}
I_0(p,b) &= \sum_{j=0}^{\infty} \frac{(-1)^j (pb)^{2j}}{(2j)!} K_j \\
I_1(p,b) &= \sum_{j=0}^{\infty} \frac{(-1)^j (pb)^{2j+1}}{(2j+1)!} K_j \\
I_2(p,b) &= \sum_{j=0}^{\infty} \frac{(-1)^j (pb)^{2j}}{(2j+1)!} K_{j-1} \\
\ldots & \quad \ldots \\
\ldots & \quad \ldots \\
I_{2n}(p,b) &= \frac{1}{2n-1}[\exp(-p)\cos(pb) - 2pI_{2n-2} - pbI_{2n-1}] \\
I_{2n+1}(p,b) &= \frac{1}{2n}[\exp(-p)\sin(pb) - 2pI_{2n-1} + pbI_{2n}].
\end{aligned}$$

The numbers $\{K_j : j \geq 0\}$ are given by

$$\begin{aligned}
K_0(p) &= \int_1^\infty \exp(-pl^2)dl \\
K_j(p) &= \frac{1}{2p}[\exp(-p) + (2j-1)K_{j-1}].
\end{aligned}$$

With $p = a\Delta_i$ and $b = 2\pi(z_i/\Delta_i - c_0)/(aS)$ we find that the first regression function can be rewritten as

$$r(\Delta_i, z_i) = A\, I_2(a\Delta_i, 2\pi \frac{z_i/\Delta_i - c_0}{aS}).$$

The second regression function equals (with $p = a\Delta_i$ and $b = 2\pi c_0/(aS)$)

$$r(\Delta_i, 0) = A\, I_2(a\Delta_i, 2\pi \frac{c_0}{aS}).$$

Figure 6.1 plots the second regression function.

If $\Delta_i = 0$ and $z_i \neq 0$, these series expansions cannot be used. We can, however, use the following alternative:

$$\begin{aligned}\int_1^\infty \frac{1}{l^2}\cos(pl)\,dl \\ &= \frac{-1}{l}\cos(pl)\Big|_1^\infty - p\int_1^\infty \frac{1}{l}\sin(pl)\,dl \\ &= \cos(p) + p\Big[\frac{\pi}{2} - \sum_{j=0}^{\infty}\frac{(-1)^j p^{2j+1}}{(2j+1)(2j+1)!}\Big].\end{aligned}$$

Substituting $p = 2\pi z/S$ and multiplying by A yields $r(0,z)$. Proposition B.2 of Appendix B gives some details as well as references with regard to this result.

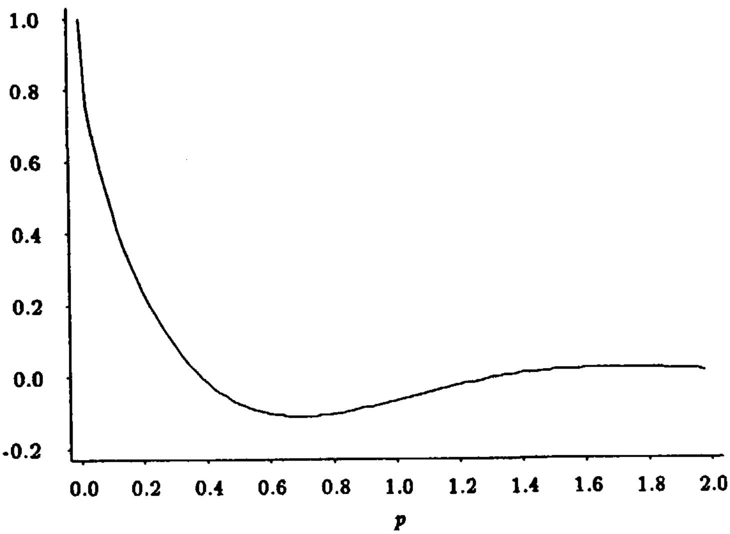


Figure 6.1: Plot of the integral $\int_1^\infty l^{-2}\exp(-pl^2)\cos(pbl)\,dl$, with $b = 3$.

By making again use of the series expansions given above (see also Proposition B.1 of Appendix B), we can easily obtain numerical approximations of the covariances of the error terms. For the case that $z_i = z_j = 0$ we have

$$\mathrm{cov}(\epsilon_i,\epsilon_j) = C\Big(I_6(p_{ij},b)+p_{ij}I_4(p_{ij},b)+I_6(P_{ij},b)+P_{ij}I_4(P_{ij},b)\Big), \quad (6.7)$$

where $p_{ij} = a\delta_{ij} = a(\Delta_j - \Delta_i)$, $P_{ij} = a\Delta_{ij} = a(\Delta_j + \Delta_i)$, $b = 2\pi c_0/(aS)$

and C is a constant. For the general case we obtain the same expression, but now $b = 2\pi[(z_j - z_j)/(\Delta_j - \Delta_i) - c_0]/(aS)$ in the first two terms and $b = 2\pi[(z_j + z_i)/(\Delta_j + \Delta_i) - c_0]/(aS)$ in the last two terms. Below plots are shown of the integral $\int_1^\infty l^{-4} \exp(-pl^2)[p + l^{-2}] \cos(pbl)\, dl$ for $b = 0$ and $b = 3$, respectively.

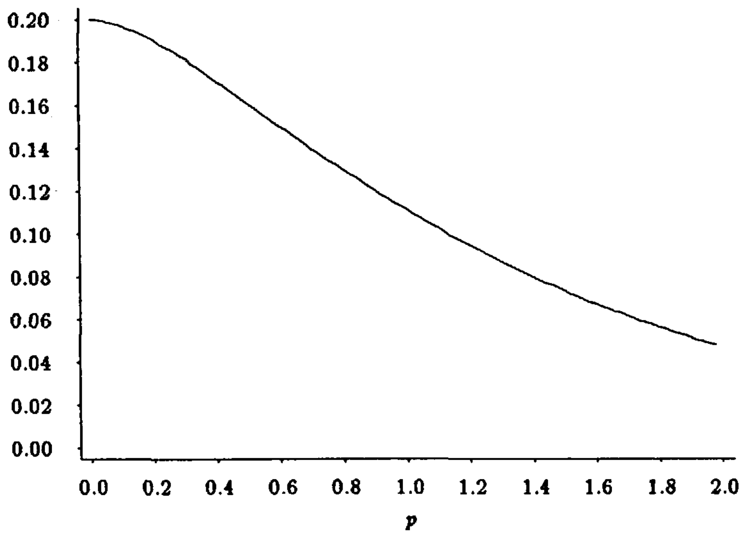


Figure 6.2: Plot of the integral $\int_1^\infty l^{-4} \exp(-pl^2)[p+l^{-2}]\, dl$.

We note, finally, that the series I_0, I_1 and I_2 converge (too) slowly, when the parameter a is very small, i.e. when b is large, even though the product pb does not depend on a. For small a, however, Riemann-sum like approximations of the regression function and its derivatives can be used.

6.3.3 Some remarks on non-linear regression

We only consider the type of observation discussed in Section 6.2, i.e. the process $R(t, x)$ is observed at a number of fixed values of the space variable. We are, therefore, dealing with the second regression problem

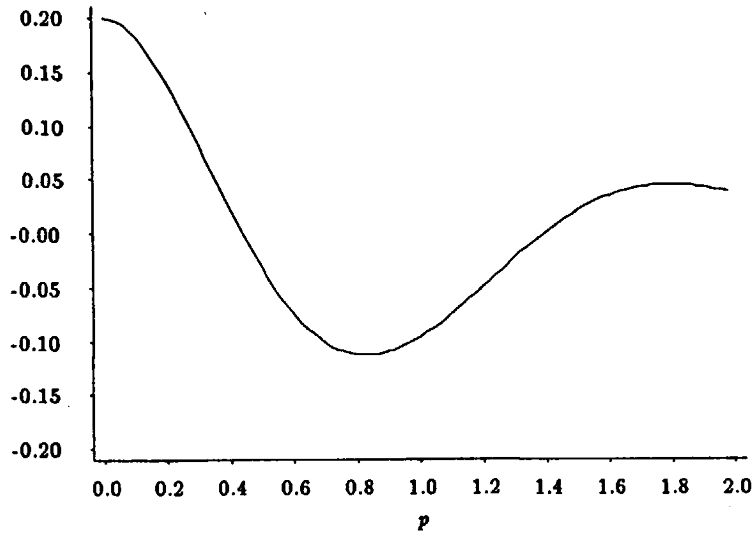


Figure 6.3: Plot of the integral
$\int_1^\infty l^{-4}\exp(-pl^2)[p+l^{-2}]\cos(pbl)\,dl$, with $b=3$.

of Proposition 6.3.

$$Y_i = A\int_1^\infty \frac{1}{l^2}\exp(-a\Delta_i l^2)\cos(2\pi l\frac{c_0\Delta_i}{S})\,dl + \epsilon_i,$$

where ϵ_i is normally distributed with mean zero and covariance matrix (6.7). Let us now introduce some standard notation. First, β denotes the vector of parameters (i.e. A, a and S). Further, we write $Y = F(\beta,\Delta)+\epsilon$, where F is shorthand for the regression function. Y is the vector of observations, Δ the vector of Δ_i-values and ϵ the vector of errors. Note that with this notation F is also a vector. F is shorthand for the regression function. Let τ denote the covariance matrix of the errors. Step 4 of the procedure sketched in the previous subsection is performed as indicated below. We transform the non-linear regression model as follows:

$$Z = \tau^{-1/2}Y = \tau^{-1/2}F(\beta,\Delta)+\eta, \qquad (6.8)$$

where $\tau^{-1/2}$ is the (Cholesky-)root of τ^{-1} ($\tau^{-1} = (\tau^{-1/2})^t\tau^{-1/2}$) and $\eta = \tau^{-1/2}\epsilon$ the vector of transformed and, thus, independent and identically distributed errors. Then the so-called normal equation reads

$$X^t \tau^{-1/2} F(\beta) = X^t Z,$$

where

$$X = \frac{\partial \tau^{-1/2} F(\beta, \Delta)}{\partial \beta}.$$

Applying least squares regression to the equation (6.8) is the same as solving the normal equation (cf. [35, Section 2.1]). Suppose we have a starting value β_0 for the parameter vector. Linerarization around β_0 yields that

$$X^t \tau^{-1/2} (F(\beta_0) + \frac{\partial F}{\partial \beta}(\beta - \beta_0)) = X^t Z,$$

so that

$$\frac{\partial F^t}{\partial \beta} \tau^{-1} \frac{\partial F}{\partial \beta}(\beta - \beta_0) = \frac{\partial F^t}{\partial \beta} \tau^{-1} (Y - F(\beta_0)),$$

where the derivative of F is evaluated in β_0. Solving for $d\beta = \beta - \beta_0$ we find that

$$d\beta = (\frac{\partial F^t}{\partial \beta} \tau^{-1} \frac{\partial F}{\partial \beta})^{-1} \frac{\partial F^t}{\partial \beta} \tau^{-1} e,$$

writing $e = Y - F(\beta_0)$. Next, we update β and repeat the steps again and again until convergence is reached. This procedure, which essentially consists of a number of least squares regressions applied to a linear equation, bears the name Gauss-Newton method.

$\partial F / \partial \beta$ consists of three columns, containing $\partial F / \partial A$, $\partial F / \partial a$ and $\partial F / \partial S$, repectively. We have the following expressions

$$\begin{aligned} \frac{\partial F(\beta)}{\partial A} &= & & I_2(p, b) \\ \frac{\partial F(\beta)}{\partial a} &= & -A\,\Delta & \quad I_0(p, b) \\ \frac{\partial F(\beta)}{\partial S} &= & A \frac{2\pi c_0 \Delta}{S^2} & \quad I_1(p, b), \end{aligned}$$

where I_i refers to the integrals discussed in Subsection 6.3.2; $p = a\Delta$ and $b = 2\pi c_0 \Delta / S$. Although the second and third integral diverge for $\Delta \to 0$, the derivatives converge to zero as the divergence is of the order $1/\sqrt{p}$.

The asymptotic distribution of $\hat{\beta}$ is given by

$$\hat{\beta} \sim \mathrm{N}(\beta, v),$$

where v can be estimated by $(e^tV^{-1}e)(X^tV^{-1}X)^{-1}/(n-r)$ (n is the number of observations and r the rank of X, which is, in this case, 3). This result is extracted from the theory on linear regression (cf. [31, p. 230] and [16, p. 213]). Because of the non-linearity the number of observations has to be quite large if the result is to be useful.

6.3.4 Discretization of the process

The next issue to be discussed is the computer simulation of the process $R(t,x)$. Of course, the process has to be discretized with respect to the space as well as with respect to the time coordinate. The discretization is performed in three steps. First, we return to the process $R_m(t,x)$ (see Subsection 4.3.4). Secondly, the process $R_m(t,x)$ is discretized with respect to the space coordinate, approximating it by the N-dimensional process $R_m^N(t)$, that is determined by the N-dimensional stochastic differential equation

$$\begin{aligned} R_m^N(t) &= K\Delta R_m^N(t)\,dt - c_0\nabla R_m^N(t)\,dt + \sqrt{\frac{N}{M}}\,\sigma\,dw(t), \\ & R_m^N(0) \sim \mathrm{N}(0,V), \end{aligned}$$

where Δ and ∇ are the difference operators defined (here) as

$$\begin{aligned} \Delta f(k) &= \left(\frac{N}{M}\right)^2 [f(k+1) - 2f(k) + f(k-1)] \\ \nabla f(k) &= \frac{N}{2M}[f(k+1) - f(k-1)] \end{aligned}$$

for any vector-valued function f. Recall that M is the total length of the space interval on which the process $R_m(t,x)$ is defined. The covariance matrix V is given by

$$V_{kl} = \sum_{i=m}^{N/2-1} \frac{\sigma^2}{(M\lambda_i K)} \cos(2\pi i\frac{k-l}{N}) + \frac{\sigma^2}{(2M\lambda_i K)}(-1)^{k-l}$$

and $w(t)$ is an N-dimensional Brownian motion, having a suitable covariance function.

And, thirdly, this N-dimensional stochastic differential equation is discretized with respect to the time coordinate, using the standard Euler-scheme. Thus we obtain the following:

$$\begin{aligned} R_d(i+1,j) &= K\Delta R_d(i,j)\,\delta - c_0\nabla R_d(i,j)\,\delta + \sigma\sqrt{\frac{N}{M}}\sqrt{\delta}\,w(i,j), \\ & R_d(0) \sim \mathrm{N}(0,V) \end{aligned} \tag{6.9}$$

where $\delta = T/N_t$, N_t being the number of time steps; $w(i,j)$ is a small Gaussian error term with zero mean and covariance

$$Ew(h,k)w(i,l) = \delta_{h,i}\left(\delta_{k,l} - \frac{1}{N} - \frac{2}{N}\sum_{i=1}^{m-1}\cos(2\pi i\frac{(k-l)}{N})\right).$$

Of course, here $\delta_{h,i}$ is the Kronecker symbol. In (6.9) d stands for 'discretized'. In fact R_d depends on three discretization parameters, viz. m, N and N_t. In order to obtain a good approximation of the process R all these parameters have to be large. Following the three discretization steps in reversed order, we claim that

$$\lim_{m\to\infty}\lim_{N\to\infty}\lim_{N_t\to\infty} R_d \stackrel{d}{=} R.$$

having embedded R_d into the space $C([0,T]\times[0,M])$ by identifying $R_d(i,j)$ and $R_d(t,x)$ for $t = iT/N_t$ and $x = jL/N$ and by subsequent linear interpolation. The proof of the claim consists in proving convergence in distribution for each step separately. Convergence of R_m to R was already proven in Subsection 4.3.4. Convergence of R_m^N to R_m is a consequence of Theorem B.1 in Appendix B and, finally, convergence of R_d to R_m^N follows from for example Theorem 1.10 in [26, p. 32].

Chapter 7

Results of the analysis of the data

The results of the analysis of three types of data are discussed below. First, we apply the non-linear regression procedure, sketched in the previous chapter, to artificial data sets, obtained directly from adding correlated errors to the regression function. This step is primarily directed at testing (the implementation of) the regression procedure. Next, we apply the same regression to simulated data sets. These data sets were obtained using the discretization described in Subsection 6.3.4. Finally, we consider the real traffic data observed at a freeway near the city of Delft.

The regression procedure has been implemented in SAS-IML. It consists of two steps. First, the Gauss-Newton algorithm is applied under the assumption that the errors are uncorrelated. Using the resulting estimates of the parameters we determine (a first estimate of) the covariance matrix of the errors. The second step iterates the Gauss-Newton algorithm, minimizing each time the sum of squares of the residuals weighted by the inverse of the covariance matrix, that was obtained in the previous iteration.

We need two convergence criteria. The iteration of the Gauss-Newton step (within a single instance of the Gauss-Newton algorithm) is stopped when the vector of the remaining errors is nearly perpendicular to the vector consisting of the last changes in the estimated values of the dependent variable. Specifically, the iteration halts as soon as

$$\frac{e^t V^{-1} X d\beta}{e^t V^{-1} e}$$

drops below some specified value. The notation was already explained in Subsection 6.3.3. The second convergence criterium determines when the iteration of the Gauss-Newton algorithm stops. We have chosen the relative change of the estimated values of the parameters, i.e. this iteration halts when

$$\max_i \frac{|d\beta(i)|}{|\beta(i)| + c}$$

is small enough; c is some small stabilizing constant (we have chosen $c = 1 \cdot 10^{-6}$).

7.1 Artificial data sets

The artificial data sets were generated and analyzed using the following values for the constants and the parameters.

A	1 (km^{-2})	c_0	0.030 (km/s)	nobs	21
a	0.1 (s^{-1})	Δ_0	1 (s)	ampl	0.1 (km^{-2})
S	0.5 (km)	seed	200	crit1, crit2	$1 \cdot e^{-6}$

Δ_0 is the time interval between successive points. We assume that the covariances of the process R are determined at equidistant points Δ_i on the time axis. 'Seed' denotes the seed-value with which the random number generator 'Rannor' of SAS has started and 'nobs', of course, stands for the number of observations. 'Ampl' is the amplitude of the covariance matrix (i.e. it equals the constant C_2 in (6.5)) and, finally, crit1 and crit2 are the 'critical' values that define what convergence means here.

Table 7.1 lists the results for 10 data sets. The first three columns show the estimates under the assumption of uncorrelated errors.

From the result on the asymptotic distribution of the parameters (see Subsection 6.3.3) we deduce roughly the following standard deviations: for A 0.03 and 0.05, for a 0.01 and 0.02 and for S 0.02 and 0.03; the first number corresponds to the regression under the assumption of uncorrelated errors, the second one to the regression assuming correlated errors.

Both regression procedures perform rather well. Convergence is almost invariably reached within 10 iterations. As one would expect the weighted regression yields a little better results.

	uncorrelated errors			correlated errors		
	A	a	S	A	a	S
1	0.997	0.078	0.520	0.985	0.107	0.498
2	1.039	0.097	0.545	1.035	0.095	0.524
3	0.997	0.043	0.571	1.015	0.082	0.490
4	0.889	0.105	0.481	0.912	0.114	0.475
5	0.996	0.093	0.440	0.984	0.111	0.510
6	1.039	0.067	0.537	1.089	0.070	0.500
7	0.883	0.165	0.619	0.930	0.124	0.539
8	0.902	0.104	0.547	0.968	0.100	0.487
9	0.934	0.173	0.417	0.935	0.123	0.465
10	0.991	0.076	0.539	0.970	0.110	0.514

Table 7.1: Regression results from 10 artificial data sets.

7.2 Simulated data sets

A simulation based on the discretized equations was implemented in Think Pascal. The following parameters and constants were used.

K	0.005 ($\mathrm{km}^2\mathrm{s}{-1}$)	c_0	0.030 (km/s)	N_t	40 000
σ	0.2 ($\mathrm{km}^{-1/2}\mathrm{s}^{-1/2}$)	Δ_0	3 (s)	m	10
M	20 (km)	dt	0.025 (s)	nobs	21
L	6 (km)	N	400		

The meaning of most symbols should be clear. Nobs denotes the number of points (in time) for which the covariance of the process R is sampled. Δ_0 is the time interval between successive points. L is the observation length. The values of the process R are recorded at 30 equidistant sites within L. These sites correspond to the detector stations where the passing traffic is detected. Furthermore, they are recorded only once in every 120 time steps. The deletion of the intermediate values made programming easier (note that $3 = 120 \cdot 0.025$); it does not constitute a serious loss of information, as the process R is continous.

The random number generator has been taken from a book on simulation by Bratley, Fox and Schrage [5, pp. 319]. The uniformly distributed random numbers have been converted into normally distributed random numbers by means of a procedure found in the same reference ([p. 327]).

The results of the analysis of 40 simulated data sets are shown in the six histograms of Figure 7.1.

The subscript '0' indicates the assumption of uncorrelated errors. From the histograms we see that all estimates are somewhat biased. The values A and A_0 are larger than the theoretical value $\sigma^2 S/(4\pi^2 K) \approx$ 0.405. The same is true for S and S_0 (theoretical value is 2). The values of a and a_0 are smaller than the expected value, $4\pi^2 K/S^2 \approx 0.0493$.

The bias of A and A_0 is obviously due to the fact that m is relatively small. The sommation $\sum_{i=10} 1/i^2$ yields approximately 0.105 instead of $1/m = 0.100$.

The source of the other biases is probably also the small value of m. For small m the covariance of the process R, which is a summation, is only roughly approximated by the (5.3). (Cf. Subsection 4.3.4). The first term of the summation is to some extent the dominant term. This leads to underestimating a.

Again we see that the regression using uncorrelated errors performs just somewhat better than the one using correlated errors. Especially the estimates of a and S have smaller standard deviation.

7.3 Real traffic data

7.3.1 Description of the data

The raw data consist of 104 936 records, each record containing the time a vehicle passes a detector station, its speed and the number of the detector. Four detectors make up one detector station. There is one detector for each lane; three lanes are for regular use, the fourth is the emergency lane.

The detector stations are situated on the western carriageway of the freeway A13 (at the locations A13W9.0 up to A13W16.5). They are numbered 1 up to and including 16. The observations were recorded on September 27, 1989 in connection with on-ramp meterings experiments. The Transportation and Traffic Engineering Division (DVK) of the institute 'Rijkswaterstaat' of the Dutch Ministry of Transport has granted the use of the data in behalf of this research.

The observations used here were made between 15.30 h. and 16.40 h. (The clocktime ran from 8559.493 to 12758.709 seconds.) During this period of the day the traffic on the freeway can usually be described as high density (and more or less) stationary freeway traffic. In order to eliminate as much as possible the influence of the on- and off-ramps only

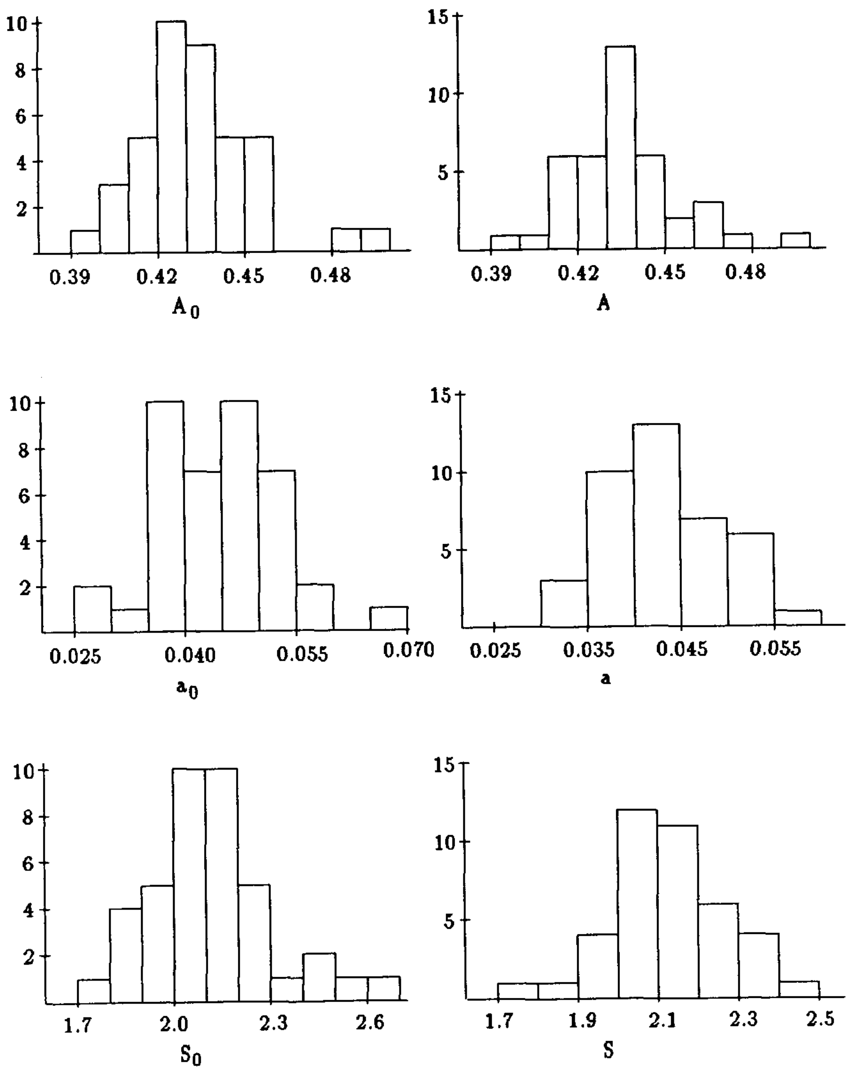


Figure 7.1: Histograms of the estimates of the parameters of 40 simulated data sets.

observations coming from the detector stations 8 up to and including 15 (which cover the freeway between Delft-Zuid and Zestienhoven – locations A13W12.5 up to A13W16.0) are considered. Thus about half of the records are discarded.

7.3.2 Preparation of the data for the analysis

Once we have decided on these preliminary questions, some more delicate decisions have to be made, viz.

- how is the density of the traffic defined in terms of the observations,
- how do we extract the mean density and the fluctuations from the observed sum
- what must be the length of the time interval on which the estimates of the parameters are based and
- how do we determine the constant c_0.

One might propose to define for each lane the density at the time that a vehicle is passing a detector as two times the inverse of the distance between the vehicle in front and the one following. Let us denote the positions of the vehicles on a lane by $x_i(t)$, such that the index i increases upstream (so that the larger the index, the later the vehicle is observed at the detector station). When vehicle i passes a particular detector at time t, the density at that time and at that detector is

$$\frac{2}{x_{i-1}(t) - x_{i+1}(t)}.$$

For fixed x (the position of the detector) this defines the density at a large number of t-values. Linear interpolation completes the definition for one lane. The density for the whole carriageway is obtained by taking the mean over the lanes.

There is an obvious disadvantage of this procedure. Two vehicle driving very close after each other (possibly within 10 metres) give rise to a very unrealistic peak of the density. Furthermore, it takes into account the coupling of the lanes only afterwards.

We have, therefore, chosen an alternative definition. Again we determine the density at a fixed x-value (the position of a detector station). At each instant we count the number of vehicles within a 100 metres distance, upstream or downstream. This number can be non-integer as each vehicle is thought to be spread out, backwards and forwards, over half of the following distances. Denote at fixed t the position of some vehicle of interest by $x_{k,i}(t)$ and its velocity by $v_{k,i}(t)$. The index k

stands for the lane of the vehicle; the index i now is 1 for a vehicle that is about to enter the interval, it is n_k for the vehicle that last left the interval (on lane k). Thus the density (vehicles per km per lane) equals

$$\frac{5}{3}\Big(\sum_{k=1}^{3}(n_k - 3 + \frac{x_{k,2}+100}{x_{k,2}-x_{k,1}} + \frac{100-x_{k,n_k-1}}{x_{k,n_k}-x_{k,n_k-1}})\Big).$$

Of course, we do not know exactly the positions of the vehicles. They must be estimated assuming that the vehicles do not change their velocities during the time that they are in scope, so that the position of a particular vehicle relative to the detector station equals the time interval between time t and the passing time multiplied by its velocity. Clearly, the constant velocity assumption limits the length of the interval. The length of 100 metres is chosen as it corresponds to a driving time of about 4 seconds. In the absence of incidents a period of 4 seconds seems short enough to guarantee constant velocities. Further, a length of 200 meters seems a reasonable choice of an increment of the space variable.

The next question concerns the extraction of the mean density and the fluctuations from the observed sum. To get an idea of how the density process behaves, the observed process was plotted for detector station nr. 8. (The plots are shown in Chapter 1.) Apart from the short range fluctuations a medium range fluctuation of the mean density can be discerned with a typical time of, say, 80 seconds. We have chosen, therefore, to calculate the mean density at a particular detector station as the moving avarage of the observed process using a window of 80 seconds. This choice immediately implied another decision, viz. not to combine the data of adjacent detector stations as far as the mean density is concerned. The mean density is changing so fast that we cannot assume it to be constant over the length of the freeway covered by the detector stations. But then the question arises, whether or not we also should try to estimate the parameters, that determine the fluctuation process, for each detector station separately. As the answer to this question is not clear a priori, both possibilities are pursued in the next subsections. Considering the data from the detector stations as separate data sets has two advantages. First, the resulting analysis of the data is more readily suitable for applications (cf. Chapter 8) and, secondly, the choices made on the basis of the observations at detector station nr. 8 can be evaluated, when we turn to one of the other detector stations.

Further, we have to decide on the length of the time interval on which the estimates of the parameters are based. Some experimenting with the data coming from detector station nr. 8 shows that a time interval of 15

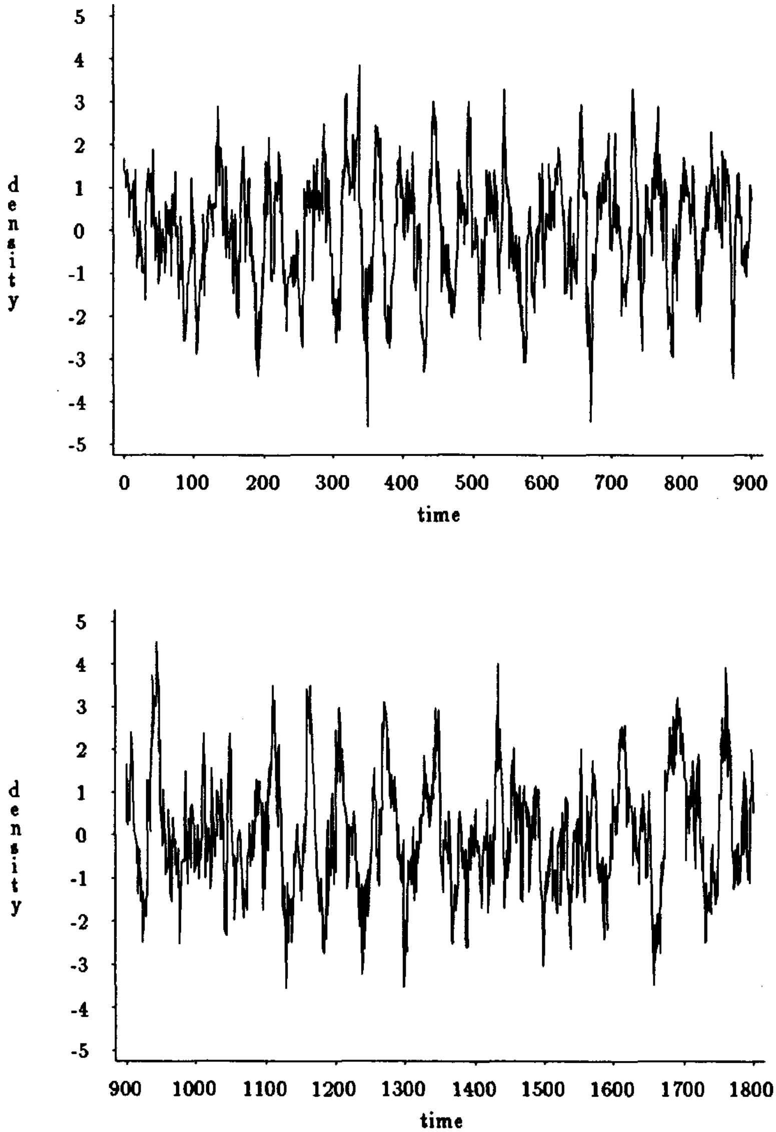
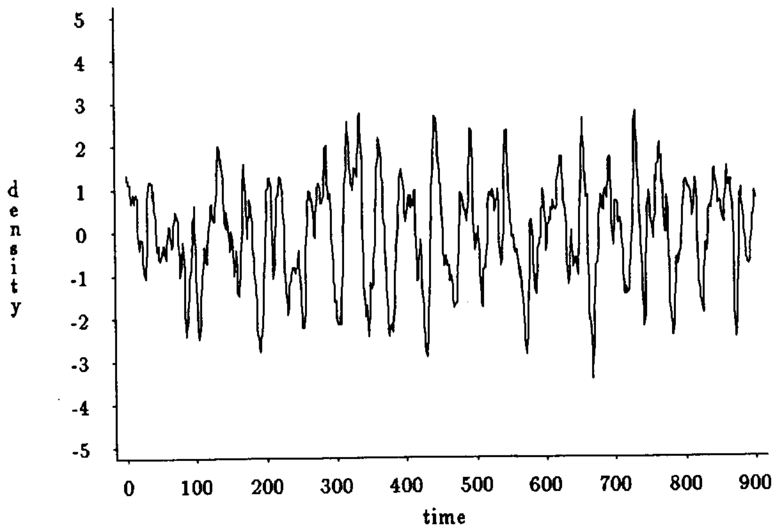


Figure 7.2: The simulated process $R(t, x)$ at a fixed value of the space parameter, x. The time is given in seconds and the density-fluctuations in km^{-2}.



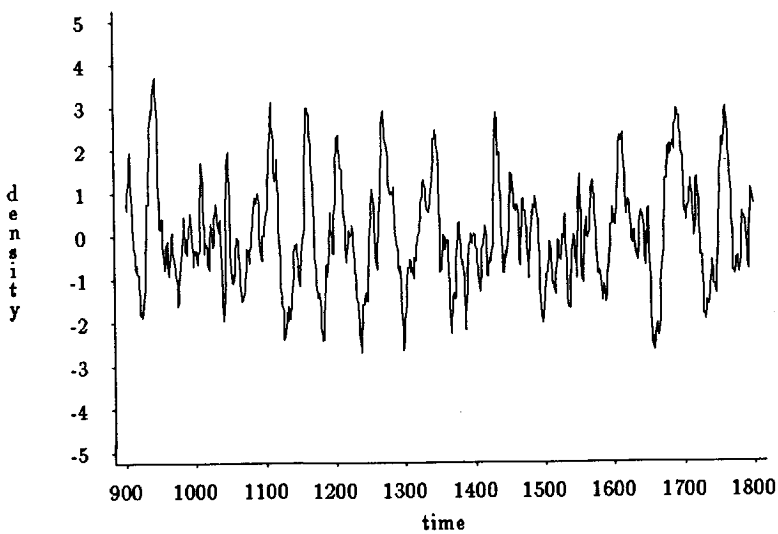


Figure 7.3: A smoothed version of the simulated process $R(t, x)$ at a fixed value of the space parameter, x. The time is given in seconds and the density-fluctuations in km^{-2}.

minutes is appropriate if we consider only one detector station at a time. This choice yields 4 successive estimates for each detector station. In on-line applications it might be advisable to use overlapping intervals, so that estimates are updated more frequently than once every quarter of an hour. For the case of combined data a time interval of 10 minutes seems to contain sufficient information (see also Subsection 7.3.5).

Finally, c_0 is determined by taking every second of the 'estimation interval' the mean of the velocities of all vehicles present in the 200 metres interval. Now we do not consider the vehicles as spread out, instead they are now treated as 'point vehicles'. The constant is set equal to the mean of these mean velocities over the appropiate time interval (and, possibly, the relevant set of detector stations). Note that the slower vehicles are counted more often than the faster ones. This feature compensates for the fact that simply taking the mean of all passing velocities overestimates the mean velocity c_0, as a fast car is more likely to be detected than a slower one.

Before we proceed to the statistical analysis of the data, we show some pictures of a realization of the simulated process (cf. Section 7.2, but now $K = 0.00125\,\mathrm{km^2s^{-1}}$, $dt = 0.1\,\mathrm{s^{-1}}$ and $N_t = 20\,000$) at a fixed value of the space parameter (Figure 7.2). When we compare these pictures to for example the observed density at detector station nr. 8 (see Figure 1.2), we notice two major differences. One is, of course, the variability of the mean density in the case of the observed process. The second difference concerns the amount of smoothness. The observed process is clearly smoother than the simulated process, at least on a time scale of a few seconds. An obvious explanation of this discrepancy lies in the way in which we calculate the density on the basis of passage times and velocities. The assumption that the vehicles pass through the 200 metres interval with velocity equal to the passage velocity causes some smoothing of the values of the density on a time scale of about 4 seconds (100 metres divided by 25 metres a second). The influence of this small scale smoothing is illustrated by Figure 7.3. This Figure shows two pictures that are obtained from the pictures of Figure 7.2 by taking the 4 second moving average.

7.3.3 Modification of the model

The inverse of the covariance matrix computed on the basis of the estimated parameters is ill-conditioned. The covariance matrix itself is almost singular. This is due to the fact that adjacent errors are very highly correlated. The matrix elements have to be calculated with high precision to get a good approximation of the inverse. The necessary precision cannot be obtained when a is too small, i.e. when we have to use the Riemann-like sum in approximating the relevant integrals. If, however, $a = 0$, we can use the alternative series expansion (given in Proposition B.2). One might, therefore, suggest to replace the covariance matrix corresponding to a small value of a (say $a < 0.02$, if $S = 2$) with the one corresponding to the case $a = 0$. But even if we succeed in calculating the matrix elements with sufficient precision, the weighted regression procedures applied to the real traffic data fail to converge. Apparently the structure of the errors of the observed covariances is not adequately described by the theoretical covariance matrix. Also the weighted regression procedure is computer-time consuming and therefore expensive. For these reasons we have chosen to use the ordinary least squares criterion.

Thus the model has been modified, on the basis of the data observed at detector station nr. 8, in two ways. First, we have chosen to calculate the mean density for each detector station separately. And, secondly, we decided to use least squares regression. The modification is an operationalization as well as a simplification of the original model. We will see how the choices work out when we turn to two of the remaining detector stations (nrs. 11 and 14). After that we consider the combined data from all detector stations (nrs. 8–15).

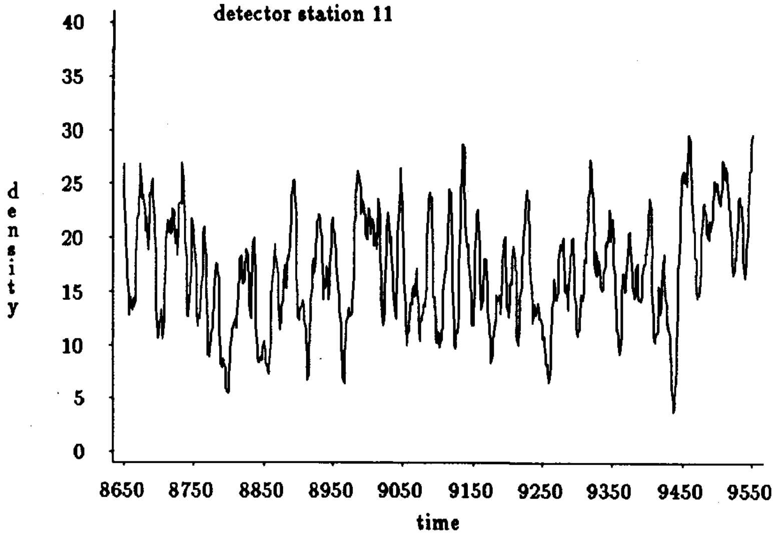
7.3.4 Estimation of the parameters for separate detector stations

Below Figures 7.4 and 7.5 show plots of the covariances, $\hat{r}$, together with the fitted covariance functions (using ordinary least squares). Table 7.2 shows the estimated values of the parameters.

We see that the estimates of the parameter a are (very) small. (In one case the parameter is even virtually zero.) As mentioned before this has some computational consequences. We have to take recourse to Riemann-sum like approximations of the regression function and its derivatives for the larger values of the time variable. If the value of a is too small (say below 0.001), we set it equal to zero and estimate the remaining parameters using the appropiate formulae given in Proposition B.2. This was done for data set nr. 5.

We mention some remarks concerning the results.

– Only part of the damping can be ascribed to the influence of a.



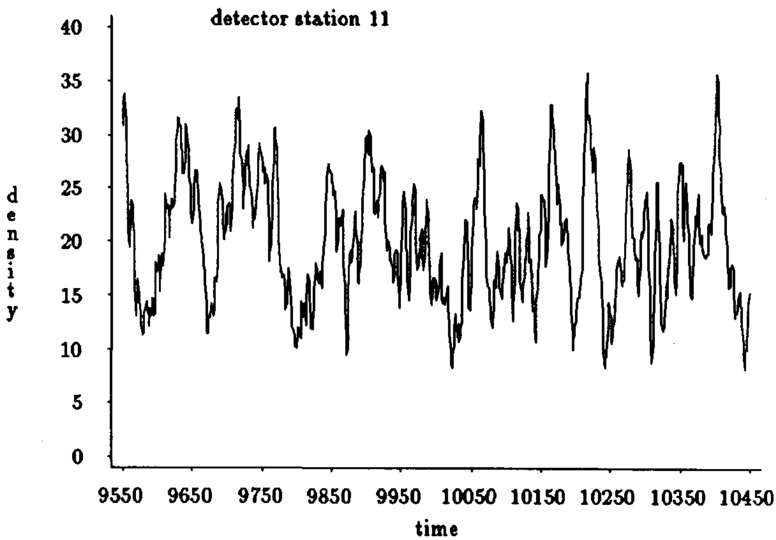
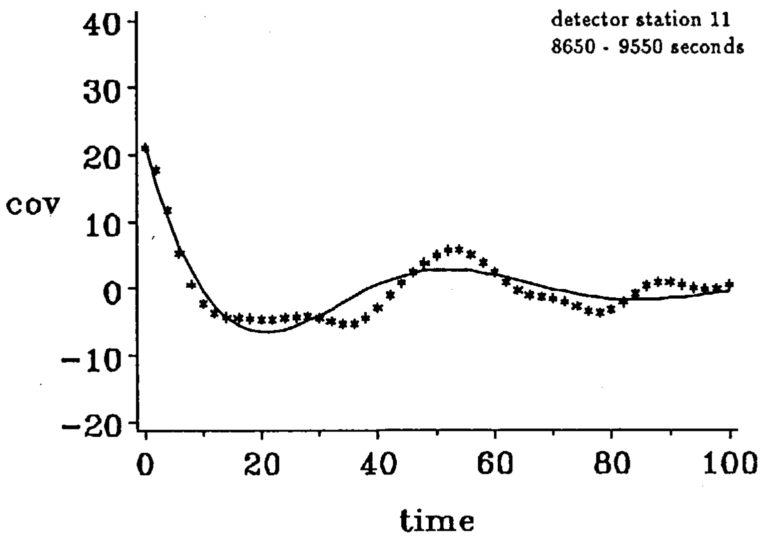


Figure 7.4: The density of the traffic (vehicles per km per lane) at detector station nr. 11 during the first two periods of 900 seconds.



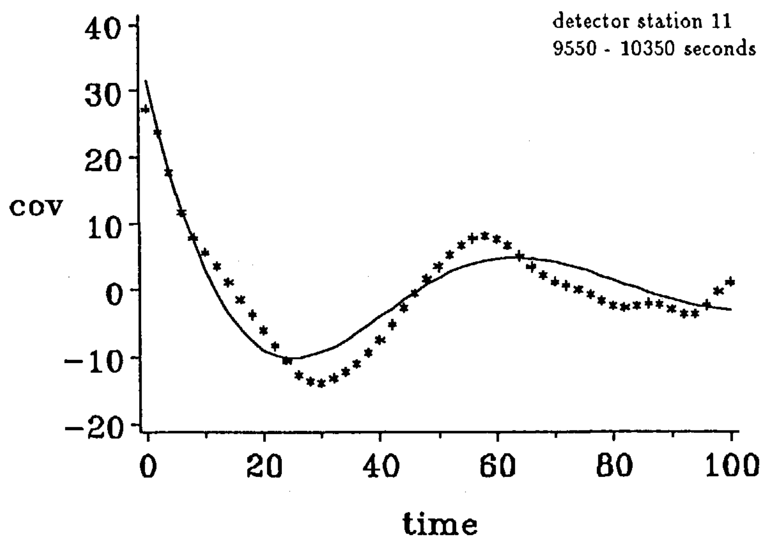
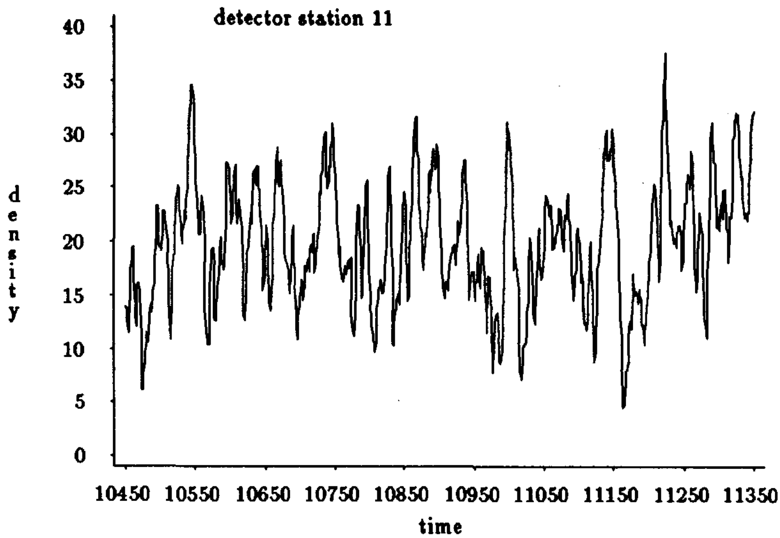


Figure 7.4: The estimated covariances of the density-fluctuations (*) at detector station nr. 11 with ordinary least squares fit (solid line) for the first two periods of 900 seconds. The time is given in seconds.



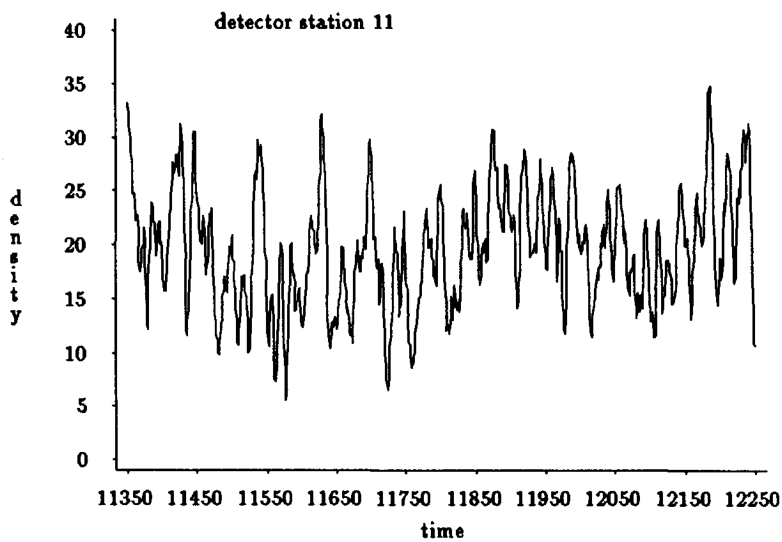
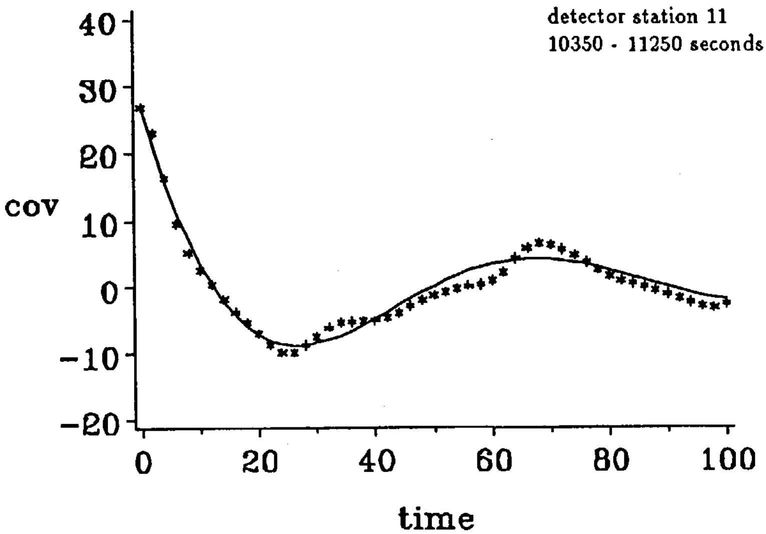


Figure 7.4: The density of the traffic (vehicles per km per lane) at detector station nr. 11 during the last two periods of 900 seconds.



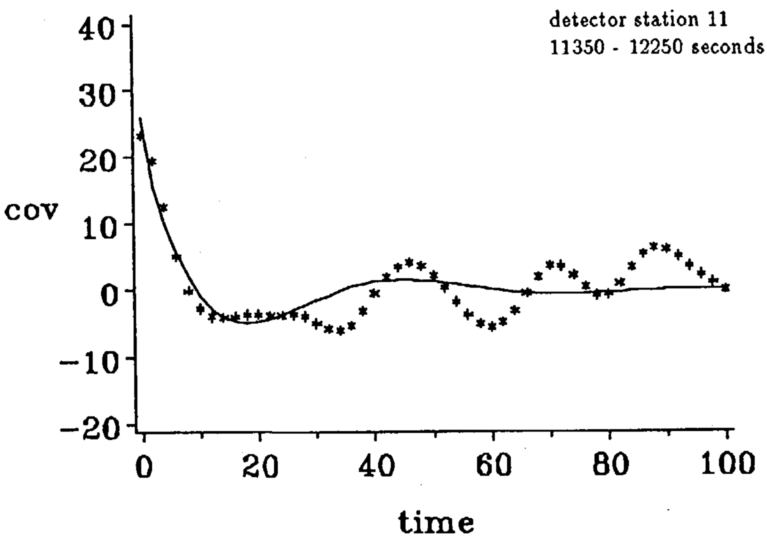
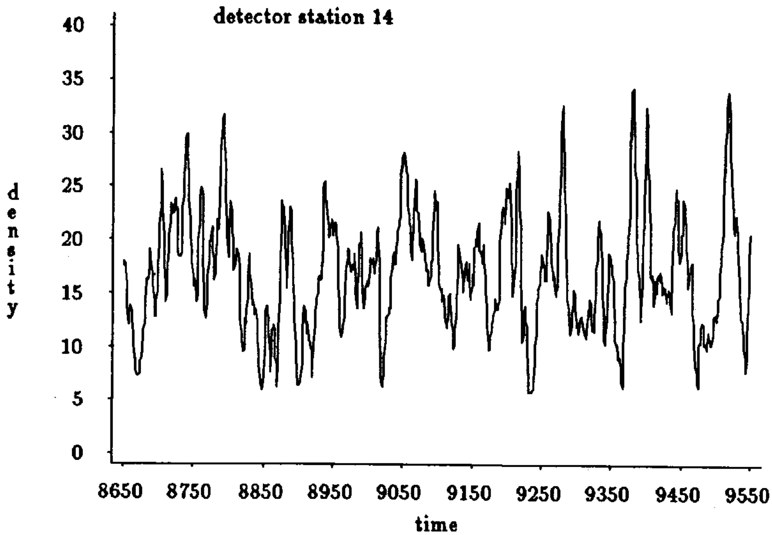


Figure 7.4: The estimated covariances of the density-fluctuations (∗) at detector station nr. 11 with ordinary least squares fit (solid line) for the last two periods of 900 seconds. The time is given in seconds.



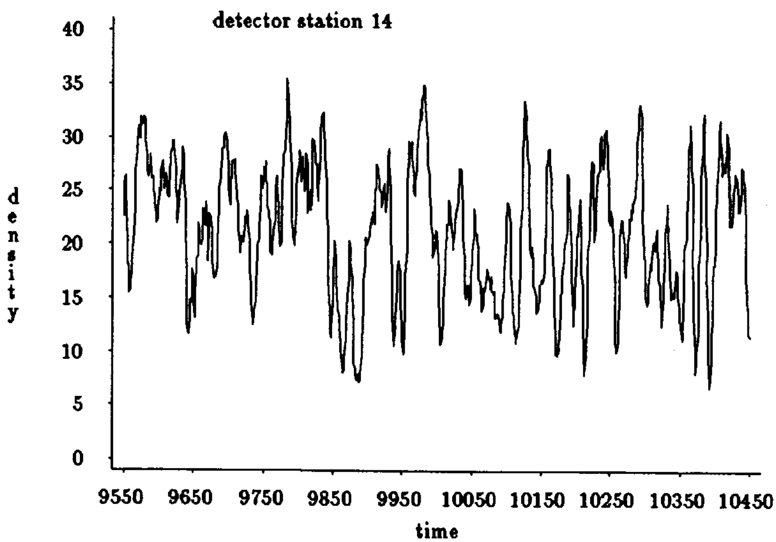
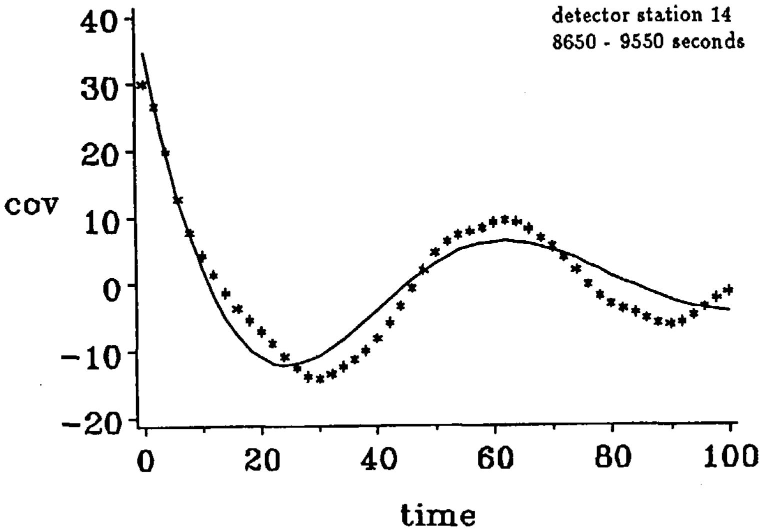


Figure 7.5: The density of the traffic (vehicles per km per lane) at detector station nr. 14 during the first two periods of 900 seconds.



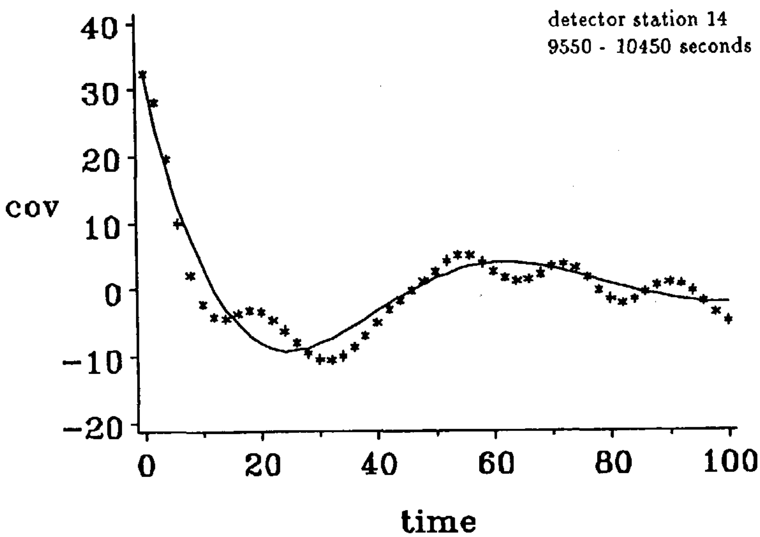


Figure 7.5: The estimated covariances of the density-fluctuations (*) at detector station nr. 14 with ordinary least squares fit (solid line) for the first two periods of 900 seconds. The time is given in seconds.

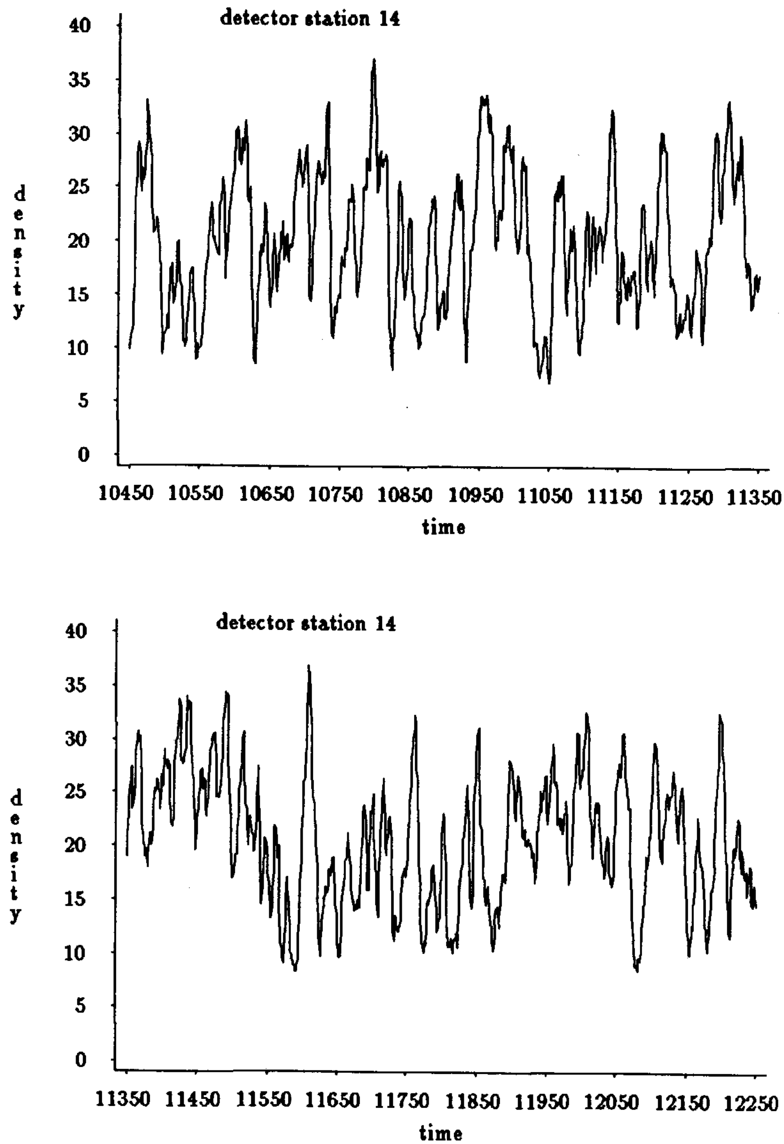


Figure 7.5: The density of the traffic (vehicles per km per lane) at detector station nr. 14 during the last two periods of 900 seconds.

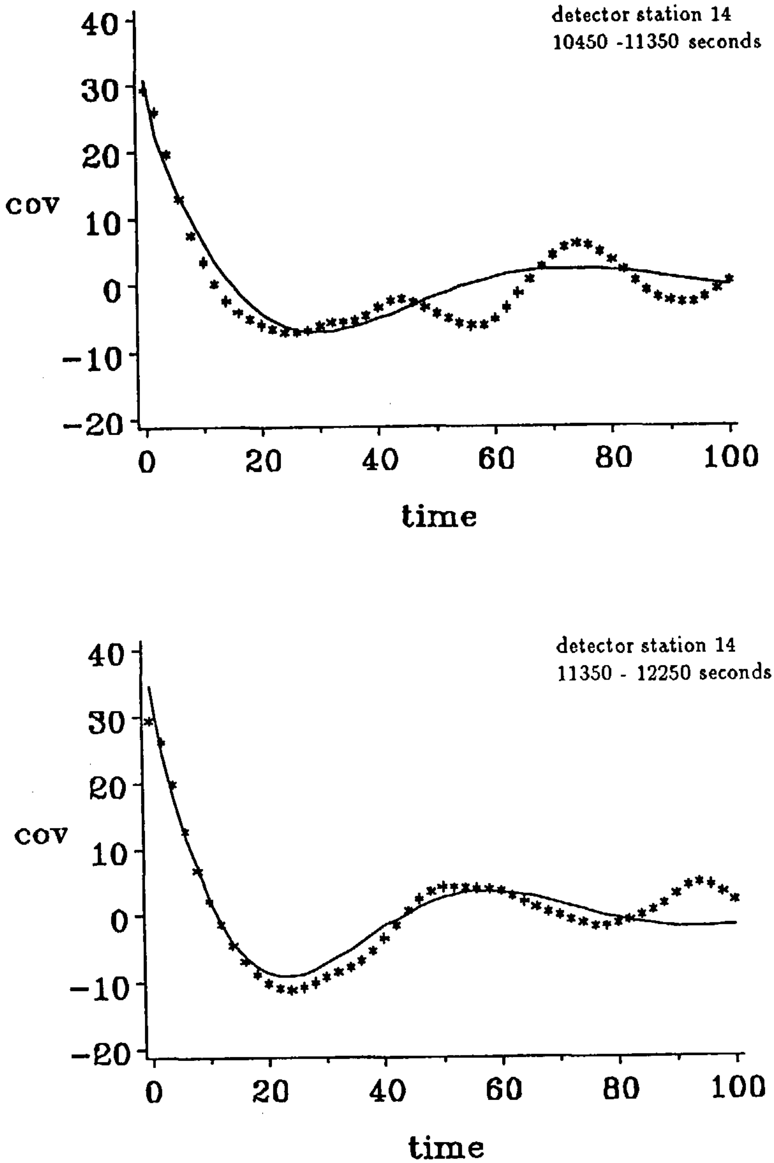


Figure 7.5: The estimated covariances of the density-fluctuations (∗) at detector station nr. 14 with ordinary least squares fit (solid line) for the last two periods of 900 seconds. The time is given in seconds.

	c_0	A	a	S	s.d. of A	s.d. of a	s.d. of S
1	0.0269	21.707	0.00519	1.773	1.465	0.00262	0.050
2	0.0285	31.461	0.00212	2.297	2.097	0.00215	0.064
3	0.0286	26.811	0.00181	2.463	0.932	0.00107	0.036
4	0.0283	25.954	0.02090	1.575	2.853	0.00773	0.118
5	0.0263	34.666	0	2.091	1.627		0.041
6	0.0272	32.135	0.00553	2.140	1.852	0.00203	0.056
7	0.0276	30.722	0.00970	2.526	2.242	0.00263	0.101
8	0.0270	34.647	0.00838	1.960	2.014	0.00231	0.056

Table 7.2: Regression results from 8 data sets from detector stations 11 and 14. Nrs. 1–4 are successive estimates at detector station nr. 11. Nrs. 5–8 are estimates at detector station nr. 14. The abbreviation s.d. stands for standard deviation.

- Only part of the damping can be ascribed to the influence of a. The other part of the damping is due the cosine in the formula of the regression function. Of course this mixing of these effects would be absent, if we had been able to observe the traffic flow, while moving along with the stream.
- The value of S is closely related to the value of c_0. If a wrong value of c_0 is taken, this only affects the estimation of S.
- In the neighbourhood of $t = 0$ the values of $\hat{r}$ tend to be reduced. The effect is certainly caused by the way the density is calculated. The assumption that the vehicles pass through the 200 metres interval with constant velocity causes some smoothing of the values of the density as was already illustrated by Figures 7.2 and 7.3.
- If we deduce the values of the original parameters K and σ from the estimated values of A, a and S, we find that $K \approx 10^{-5} - 10^{-4}$, which is very low; calculations on the basis of the parameters of the non-linear equations, that are discussed in Chapter 2, page 11, would lead us to expect that $K \approx 10^{-3} - 10^{-2}$.
- Further, as we can write $A = \sigma^2/(aS)$ and the parameters A and S are roughly constant, we see that σ^2 is by and large proportional to a.
- The pictures of the estimated covariances of the process at detector station no. 8, 11 and 14 suggest that there are two dips for $t \approx 30$ seconds and two peaks for $t \approx 60$ seconds. One might suggest that

	c_0	A	a	S	s.d. of A	s.d. of a	s.d. of S
1	0.0257	23.801	0.0072	1.812	1.524	0.0025	0.054
2	0.0279	34.666	0.0045	2.357	1.453	0.0014	0.044
3	0.0281	36.792	0	2.054	1.438	0	0.034
4	0.0284	30.458	0.0126	2.077	1.262	0.0019	0.049
5	0.0274	23.091	0.0018	2.689	1.191	0.0014	0.062
6	0.0278	27.860	0.0081	1.745	1.712	0.0027	0.049

Table 7.3: Regression results from 6 data sets (combined data from all detector stations). The abbreviation s.d. stands for standard deviation.

this is due to different behaviour of the left and the middle lane and the right lane (which is characterized by much heavy traffic).

7.3.5 Estimation of the parameters for combined data

Table 7.3 and Figure 7.6 present the results of the estimation of the parameters on the basis of the combined data (from detector stations nrs. 8–15). The estimation time interval is 10 minutes. An interval of 5 minutes contains too little information to estimate the parameters. This implies that combining the data from several detector stations does not lead to much shorter estimation times. The results do not differ significantly from those presented in the previous subsection.

7.3.6 The accuracy of the estimates

A rather delicate question is the estimation of the standard deviations of the estimates of the parameters. Table 7.2 and Table 7.3 list the values obtained from the result on the asymptotic distribution of the parameters (see Subsection 6.3.3) under the assumption that the errors are uncorrelated. It seems worthwile to compare these values with values from a sample of artificial data sets, which are obtained by adding correlated errors to the regression function. For lack of the true covariance matrix we use the matrix V. Forty data sets were generated. The (typical) values of the parameters and constants are

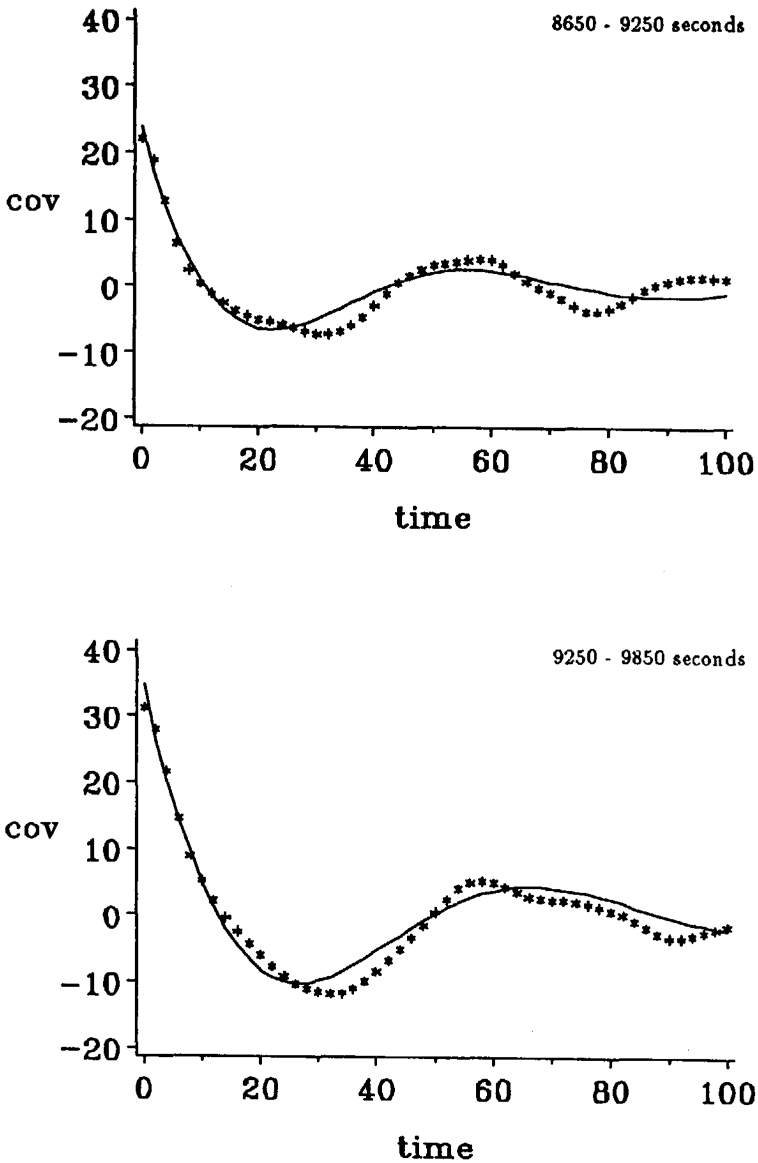
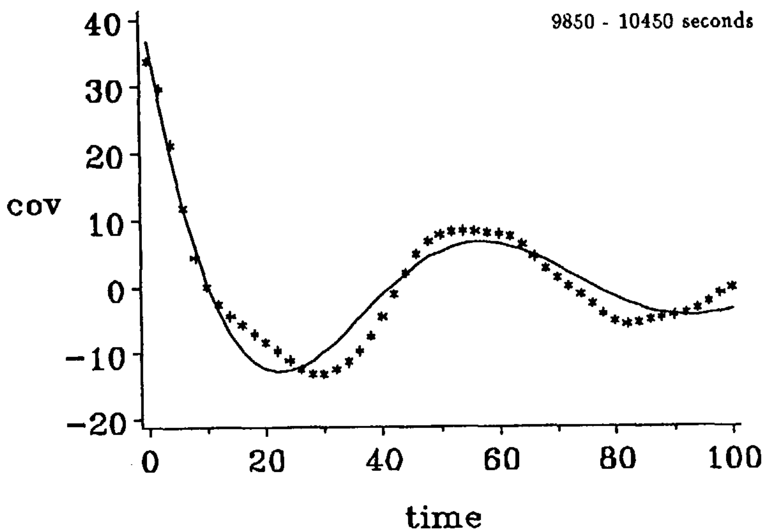


Figure 7.6: The estimated covariances of the density-fluctuations (*) for all detector station together with ordinary least squares fits (solid line). The time is given in seconds.



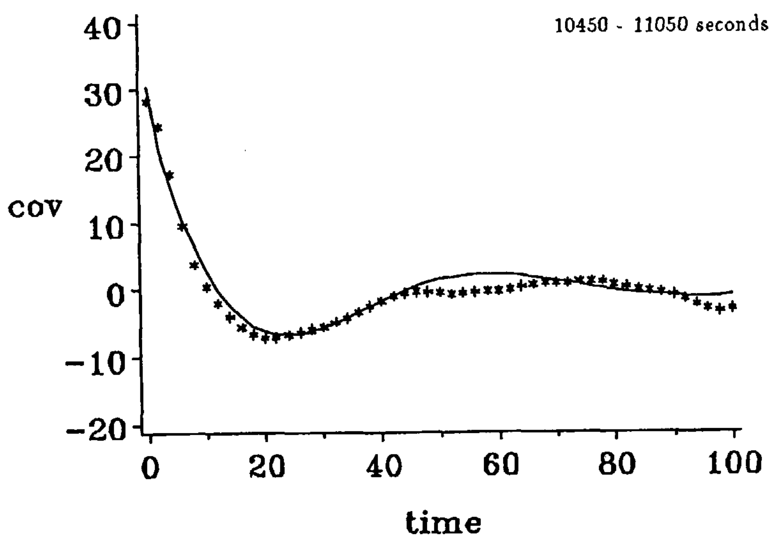


Figure 7.6: Continued.

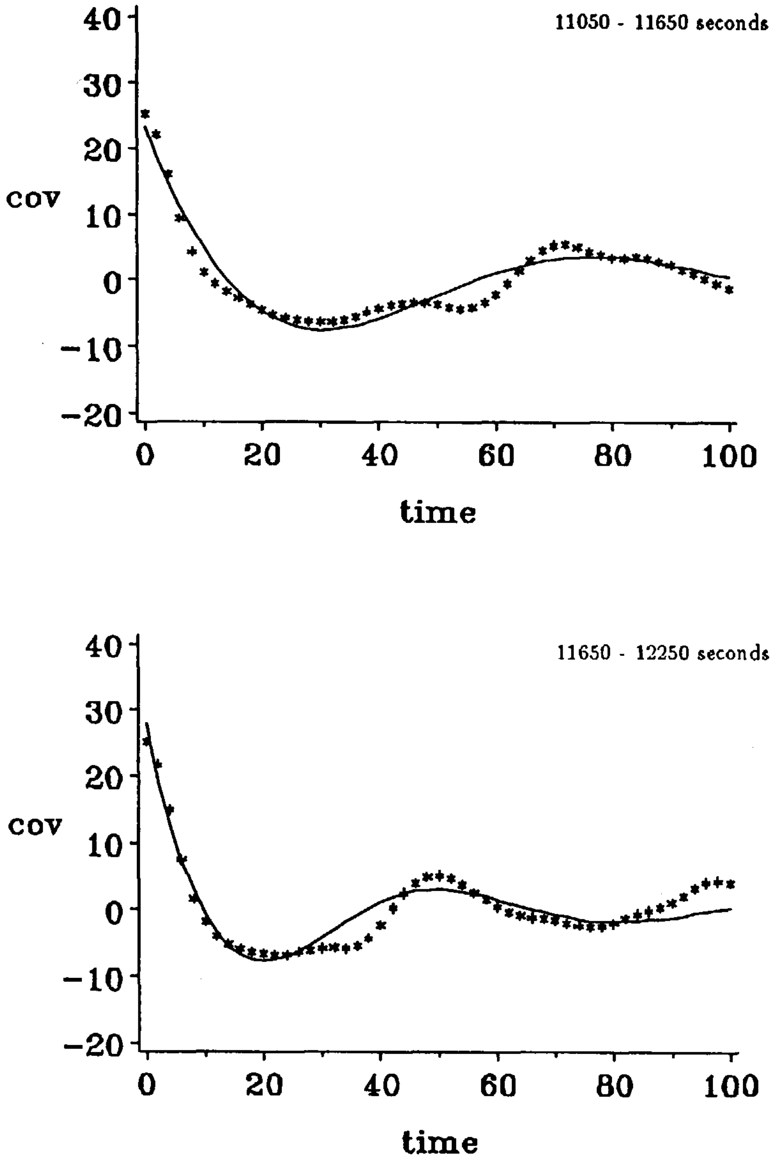


Figure 7.6: Continued.

The meaning of the symbols is the same as in Section 7.2. Note that the value of a is rather high for a 'typical' value. This is unavoidable as for lower values of a it is impossible to calculate the matrix V with sufficient precision. Doing the (ordinary least squares) regression yields the following results (s.d. stands for standard deviation):

A	28.9 (km^{-2})	s.d. of A	4.0
a	0.014 (s^{-1})	s.d. of a	0.007
S	2.5 (km)	s.d. of S	0.4

Clearly the estimated standard deviations from these data sets are much larger than the values given in Table 7.2 and Table 7.3. This discrepancy can be explained by noting that the covariance matrix V corresponds to errors that are correlated over the entire range of the time variable, whereas the values given in the tables assume uncorrelated errors. What a very large range can cause, is nicely illustrated by Figure 7.7. The figure shows how a large amount of the errors can be disappear into the estimated regression function. As a consequence the parameters are rather poorly estimated. Since the pictures of the observed data suggest that for these data the range of the correlation is somewhere between these two extremes a reasonable conclusion seems to be that the parameters A and S can be estimated with an accuracy of about 10%. Estimation of the parameter a is more difficult; If the parameter is very small, little more than that can be said. If it is somewhat larger, say about 0.01 (for $S = 2.5$), an accuracy of 30% is likely. The larger a is, the more accurately it can be determined.

7.3.7 Two additional analyses

Two additional analyses were carried out. The first one concerns the estimation of a. If we could observe the fluctuations of the density while moving along with the stream, this would yield rather direct information about the parameter a. In a sense the influence of the length S would be cancelled, as the regression function reduces to

$$r(\Delta, c_0\Delta) = A \int_1^\infty \frac{1}{l^2} \exp(-a|\Delta|l^2)\, dl.$$

We have not observed the traffic stream in this way. However, we do have a few observations at (successive) detector stations, which yield

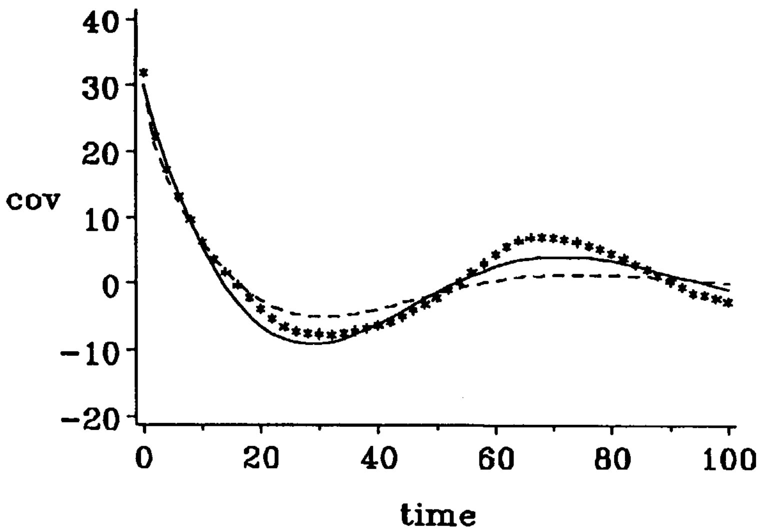


Figure 7.7: A typical artificial data set (*) with the true regression function (dashed line) and ordinary least squares fit (solid line). The time is given in seconds.

estimates

$$\hat{r}(\Delta_i, c_0\Delta_i) \equiv \frac{1}{T-\Delta_n}\frac{1}{L-c_0\Delta_n}\int_0^{T-\Delta_n}\int_0^{L-c_0\Delta_n} R(t,x)R(t+\Delta_i, x+c_0\Delta_i)\,dx\,dt$$

for a few values of Δ_i, viz. values of Δ_i, such that $c_0\Delta_i = (i-1)\cdot 0.5\,\text{km}$, $i = 0\ldots n$. These estimates of $\hat{r}(\Delta_i, c_0\Delta_i)$, for $n = 6$, have been calculated for the first ten minutes estimation interval (cf. Table 7.3, the first data set). Figure 7.8 shows the data and the ordinary least squares fit. The estimated values of the parameters are $A = 21.41\,(\pm\,0.94)$ and $a = 0.00315\,(\pm\,0.00044)$. The standard deviation are, as before, based on the asymptotic distribution formula. Of course, the fact that $n = 6$ throws some doubt on the reliability of the values. Nevertheless, it seems quite clear that, if the value of the parameter a is very small, and if one has observations from several successive detector stations at his disposal,

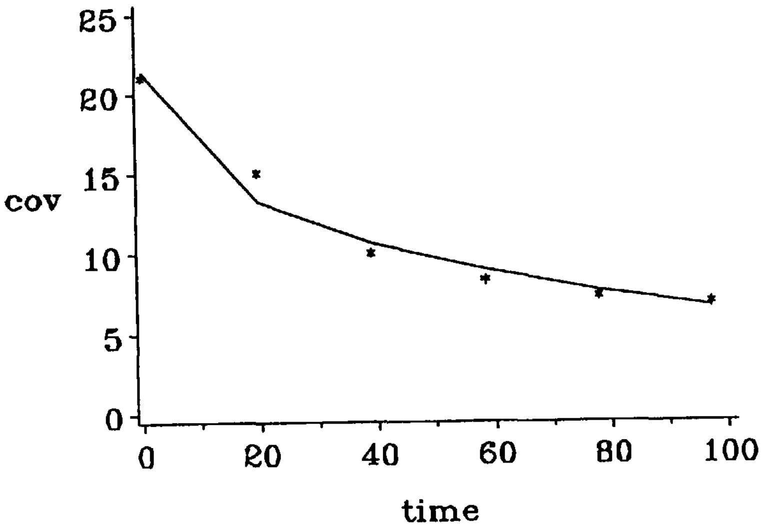


Figure 7.8: Estimates of $\hat{r}(\Delta_i, c_0\Delta_i)$ $(*)$ with ordinary least squares fit (solid line) for the first ten minutes estimation interval. The time is given in seconds.

this manner of determining a is very useful. If, on the other hand, a is large, the usefulness is restricted. Because the covariance drops to zero much faster, the information is to be extracted from only two or three points. At the same time the method based on the estimates of $\hat{r}(\Delta_i, 0)$ will perform significantly better for large a.

Another interesting feature of the density-fluctuation process is the way in which the observations at two neighbouring detector stations are correlated. Figure 7.9 shows the estimated covariances $\hat{r}(\Delta, 0.5\,\text{km})$ at detector stations 11 and 12 for the first 15 minutes of the observation time. The plot exhibits the expected behaviour. An attempt to fit a regression curve through these data was not undertaken.

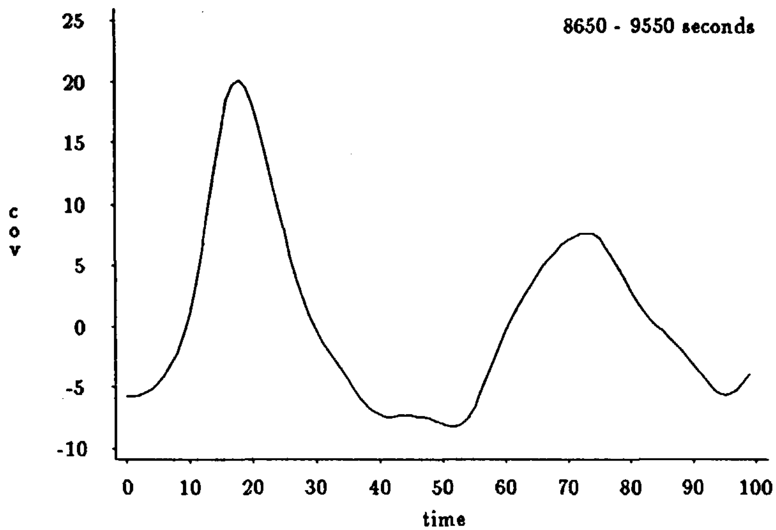


Figure 7.9: Estimate of $\hat{r}(\Delta, 0.5\,\mathrm{km})$ at detector stations 11 and 12. The time is given in seconds.

7.3.8 On the normality of the process R

The normality of the process R has been tested by sampling once every hundred seconds from the data at detector stations nrs. 11 and 14. It is assumed that these values are (almost) independent. Therefore normal probability plots can be used to assess the normality of the process (see Figure 7.10).

7.4 Discussion of the results

The determination of the mean density required an adaptation of the model. It changes much more rapidly than was hoped for. Partly this is caused by the irregular input of vehicles at the junctions, partly it is due to the steadiness of the irregularity. This feature is reflected by the low values of a. The variability of the mean density entailed the necessity to calculate it for each detector station separately.

Furthermore, we changed the model in that we used ordinary least

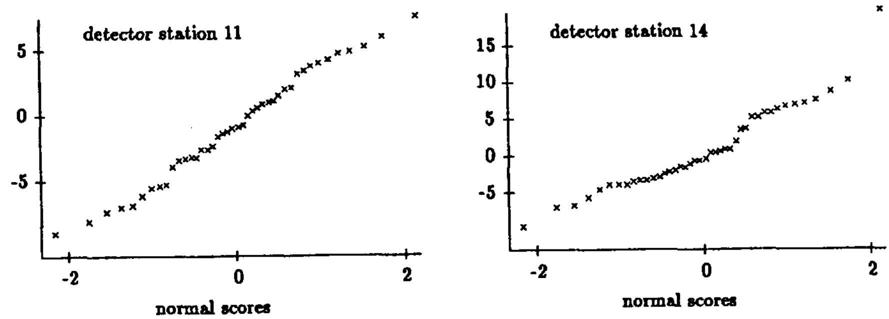


Figure 7.10: Normal probability plots of data sampled at detector stations nr. 11 and nr. 14 once every 100 seconds. The i^{th} normal score is defined here as the median of the i^{th} order statistic in a sample of size n drawn from a standard normal distribution.

squares regression instead of the weighted regression. From the results presented in the first sections it is clear that the weighted regression is to be prefered over the ordinary least squares regression, provided the errors are correlated according the theoretical covariance matrix. Apparently the errors of the observed covariances are not correlated in this way. This not only holds for the data coming from one detector station, but also for the combined data, for which the matrix was actually derived. The pictures suggest that the range over which the errors are correlated is smaller than is implied by the theoretical covariance matrix.

The adapted model describes reasonably well the observed covariances of the density. We note that the fact that the regression procedures almost always converged indicates that the amount of damping present in the observed covariances is covered by the model. In the two exceptional cases putting $a = 0$ yields acceptable results. We conclude that the observed covariances obey the constraint that $a \geq 0$. It is probably the most successful feature of the model that it reveals this property of the traffic flow so clearly. We will come back to this matter in Section 8.2.

Because the parameter a is very small, it is difficult to estimate it accurately.

The normal probability plots suggest that the fluctuations, as extracted from the observed density, can be assumed to be normally distributed.

In view of the difficulties we encountered when using weighted least squares regression the question arises whether the choice, made in Chapter 6, to estimate the parameters via the estimates of the covariances should be maintained. An alternative is to return to the original model equation (5.1) and find estimates on the basis of its discretized version (see Subsection 6.3.4 for the notation)

$$R_d(i+1,j) = K\Delta R_d(i,j)\,\delta - c_0\nabla R_d(i,j)\,\delta + \sqrt{\frac{N}{M}}\sqrt{\delta}\,\sigma\,w(i,j).$$

Writing

$$e(i,j+1) = \Big(R_d(i+1,j) - K\Delta R_d(i,j)\,\delta + c_0\nabla R_d(i,j)\,\delta\Big),$$

and minimizing $e^tW^{-1}e$, where W is the covariance matrix of the error term would yield estimates of K, S and σ. This approach does not rely so much on the exact stationarity of the process. Furthermore, the covariance matrix W is relatively simple (compared to the covariance matrix V). There are, however, three objections to this proposal. The first one is the huge size of the regression problem. Note that the route via the covariance estimates starts with a significant compression of the information. Secondly, the parameters S and σ have to be estimated from the structure of the residuals, whereas the procedure followed here has absorbed the parameters S and σ into the regression function. As to σ, the estimation from the matrix elements of $1/n(e^te)$, where e is the residuals matrix (n is the number of time steps, i.e. the range of the i-variable) should pose no problem, but the estimation of S seems to be a very difficult task. The last objection concerns the availability of data. One must have at one's disposal observations of the process along the entire interval $[0,L]$. These data are difficult to obtain.

Chapter 8

Summary, conclusions and recommendations

8.1 A short summary of the model

Let us recall the objective of this thesis. It 'aims at a mathematical description of traffic flow on a freeway, thereby identifying some useful flow characteristics. Estimating these characteristics and taking appropriate action may lead to enhancement of the homogeneity of the flow.' (See Chapter 1, page 1.) In Section 1.2 we restricted the scope of the research to stationary high density multilane freeway traffic flow. The reason for this was the impossibility to design one single model for all kinds of traffic flow.

Faced with a problem like the task of modelling traffic flow on a freeway one can in general choose between two approaches. The first is to gather data and derive conclusions. The second approach is to start with the mathematics. In this case intuition about the nature of the traffic flow is translated into mathematical equations. Consequences of the model are derived and are subjected to the criticism of the data only in a later stage of the research. The advantages and disadvantages of the latter (theoretical) approach can be illustrated by the results of the research presented here. We will come back to these aspects in the next section.

In Chapter 2 a kind of 'conservation of vehicles' equation suggested itself as the natural starting point of the analysis. As a consequence the density of the traffic (number of vehicles per km) became the central dependent variable (depending on time and place). This equation and other, more heuristic, reasoning led to the stochastic model presented in Chapter 5. The main model equation was obtained by linearizing a pair of non-linear equations. This step was taken to simplify the model. It is believed to be acceptable if we deal with stationary traffic flow. In the case of stationary flow it seems natural to discern between some (possibly slowly varying) mean density and fluctuations superimposed on this mean density. These fluctuations of course must sum up to zero in order not to loose or gain vehicles.

The model essentially consists in a description of the fluctuations of the density of a freeway traffic stream as a two-parameter Gaussian process, denoted by $R(t,x)$; t and x are the time and space parameter, respectively. The process is completely specified by its covariance function $r(\Delta, z)$.

$$r(\Delta, z) = A \int_1^\infty \frac{1}{l^2} \exp(-a|\Delta|) \cos(2\pi l \frac{z - c_0\Delta}{S})\, dl, \qquad (8.1)$$

where $A = \sigma^2 S/(4\pi^2 K)$ and $a = 4\pi^2 l^2/(S^2)K$. K and σ are parameters occurring in the stochastic differential equation having the Gaussian process as its stationary solution, K is a parameter that determines how strong the smoothing tendency of the process is that counteracts the influence of the stochastic disturbances (i.e. the noise). The noise is determined by the parameters σ and S; σ determines the amplitude of the disturbances and S their range. Further, c_0 is the mean velocity of the traffic velocity. K, σ and S may be called the characteristics, as they determine the degree of (non-)homogeneity of the flow.

Alternatively, A, a and S may be chosen as the relevant parameters. Indeed, they are the ones that are estimated. The interpretation of these parameters is as follows. A is the squared amplitude of the fluctuations, S is the typical length of a fluctuation and a is a damping factor that determines how fast or slowly a certain configuration of fluctuations changes; $a \approx 0$ would mean that a configuration would travel along the freeway (almost) unchanged. It is clear that large values of A and of a characterize a wildly changing process. In Appendix C we state an extreme value theorem, which makes this statement more precise.

8.2 Conclusions

As mentioned in the previous section the approach of developing a model by translating intuition into mathematical statements has advantages and disadvantages.

The prime disadvantage is that the data contradict some of the assumptions underlying the model. This necessitated us to adapt and simplify the model. Because the adaptation is ad hoc, the model is to some extent weakened.

The major advantage of the theoretical approach seems to be that it reveals the importance of the parameter a and, at the same time, provides a means to determine the value of this 'hidden' parameter. As can be seen from the extreme value theorems stated in Appendix C, the probability of the occurrence of extremes, relative to the amplitude, is determined by the value of a/S, which stresses the importance of finding the value of a. But, whereas the squared amplitude of the fluctuations A as well as the length S are readily deduced from the data, a is not easily determined.

The parameter a measures the amount of damping one sees if one would travel along with the traffic stream. Some efforts were made to obtain an observation of a traffic stream by means of aerial photographs. These fotographs would have enabled us estimate the parameter a more directly. Unfortunately, this part of the research could not be brought to a successful conclusion.

A large value of a implies that fluctuations occur rather spontaneously and also, more importantly, that the extreme values tend to be more extreme (at least if A and S are constant). The model's usefulness lies in the fact that it is able to derive from rather limited observations at detector stations the approximate values of the parameter a, as well as the values of the other parameters. Together these values can be used for taking measures to 'regularize' the stream (see Section 8.3).

Perhaps the most serious criticism of the lines of research followed here lies in the remark that the adequate formalism for describing the nature of traffic flow is by no means obvious. In this thesis the analogy between fluid flow and traffic flow is exploited, but already in the introductory chapter we pointed out that the analogy only holds to a limited extent. This limitation is illustrated by the somewhat arbitrary choice of the working definition of the notion density in Chapter 7.

There are two answers to this objection. The first is that the model presented here is not meant to cover all freeway traffic situations — the area of application is restricted (see also Chapter 1). The second answer concerns the (first) application of the model to data, for after we removed some apparently unrealistic elements of the model (in particular with respect to the mean density), the (remaining part of the) model performs reasonably well.

8.3 Recommendations

The comparison of the model with data suggests that the model is worth applying in everyday traffic situations, although additional validation is required. Perhaps the most important open question is the range of values of the parameter a. As the data used here were gathered in pre-peak-hour period, we conjecture that in the middle of the peak-hour much larger values of the parameter occur, such that the values found here can all be said to be (very) close to zero. The small values of K that are deduced from the values of a and S (see also Subsection 7.3.4) support this view. If this guess is confirmed, one of the consequences will be that the parameter a can be estimated with sufficient precision whenever it is important to know its value, i.e. when it is large.

If the model passes this additional test, the following application is proposed. We interpret the phenomenon of large $\tilde{R}(t,x)$ — which corresponds to a local concentration of vehicles — as a situation in which there is a large probability of congestion. Recall that $\tilde{R}(t,x)$ is defined as the sum of the mean density and the fluctuation. Clearly, this phenomenon depends on the value of the mean density and on the value of the parameters. Of these the parameters A (measuring the mean squared amplitude of the fluctuations) and a (measuring the damping present in the process) are the most important ones.

Large values of the mean density and the parameters A and a imply a large probability of an extreme value above some crucial level. A small value of S has the same effect, but it seems that this parameter is not likely to change very much. If the density exceeds the crucial level the flow is likely to break down. Of course the value of this crucial level will depend on the mean density.

The parameters can be determined along the lines followed in Chapter 7. It does not make much difference whether data from one or several detector stations are used for the estimation of the parameters.

This sketch does not lead to quantitative statements. Therefore, the critical values as well as the range of the parameters under various circumstances have to be determined from studying a large amount of data. Once the critical values of the parameters have been found, the state of the traffic flow can be assessed (in an on-line configuration) and appropriate measures be taken, for example with the help of a signalling system.

Appendix A

This appendix contains all the proofs omitted in Chapter 4 as well as results needed in the proofs. Before a proof is given the theorem or proposition is stated again.

A.1

The first part of Appendix A concerns Section 4.2. Recall the probabilistic setting of Chapter 3. $(\Omega, \mathcal{F}, P)$ is a complete probability space with a right-continuous increasing family of sub-σ-algebra's, $\{\mathcal{F}_t : t \geq 0\}$, each containing all P-null sets, i.e. the set-up $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$ satisfies the usual conditions. H always denotes a real separable Hilbert space.

Lemma A.1 *Let $f : [0,T] \times \Omega \longrightarrow H$ be a H-valued, measurable and $\{\mathcal{F}_t\}$-adapted function satisfying $E\int_0^T \|f(t)\|^{2p}\, dt < \infty$ and $F : [0,T] \times \Omega \longrightarrow \mathcal{L}_2(H)$ a $\mathcal{L}_2(H)$-valued, measurable and $\{\mathcal{F}_t\}$-adapted function that satisfies $E\int_0^T \|F(t)\|_{HS}^{2p}\, dt < \infty$. Then we have the estimates*

$$\begin{aligned} E\Big(\int_0^T \langle f(t), dB_t\rangle\Big)^{2p} &\leq C_1\Big\{\int_0^T (E\|f(t)\|^{2p})^{1/p}\, dt\Big\}^p \quad \text{and} \\ E\,\|\int_0^T F(t)\, dB_t\|^{2p} &\leq C_2\Big\{\int_0^T (E\|F(t)\|_2^{2p})^{1/p}\, dt\Big\}^p, \end{aligned}$$

where C_1 and C_2 are constants depending only on p.

PROOF. See [9, p. 134 and 135]. □

Lemma A.2 (Hölder inequality) *Let f and g be functions in $L^p[0,M]$, where M is some constant and $p > 1$. Then*

$$\Big(\int |fg|\, dx\Big)^p \leq \int |f|^p\, dx \,\Big(\int |g|^{p/(p-1)}\, dx\Big)^{p-1}.$$

Analogously, if $\{a_i\}$ *and* $\{b_i\}$ *are elements of* l^p *(the Hilbert space of sequences* $x = (x_1, x_2, \cdots)$ *such that* $\sum_1^\infty |x_i|^p < \infty$*), then (for* $p \geq 2$*)*

$$\left(\sum |a_i b_i|\right)^p \leq \sum |a_i|^p \left(\sum |b_i|^{p/(p-1)}\right)^{p-1}.$$

Theorem A.1 (= Theorem 4.4) *Assume that*

(i) U *is the semigroup generated by a closed linear operator* A; $-A$ *is a positive, self-adjoint operator such that* $(-A)^{-1}$ *exists and is a nuclear operator; furthermore, for some* $\delta > 0$, $0 < \liminf_i \lambda_i / i^{1+\delta} \leq \limsup_i \lambda_i / i^{1+\delta} < \infty$, *where* $\{\lambda_i\}$ *is the set of eigenvectors of* $-A$; $\{e_i\}$ *is the corresponding orthonormal family of eigenvectors;*

(ii) P *is a linear operator with* $\mathcal{D}(P^*) \supset \mathcal{D}(A)$ *such that* $\|P^* e_i\| \leq \gamma \lambda_i^\alpha$, *where* γ *is some constant and* $\alpha < 1 - 1/(2(1+\delta))$ *or, if* P *and* A *commute on* $\mathcal{D}(A)$, *we require that* $\mathcal{D}(P) \supset \mathcal{D}(A)$ *and* $\|Pe_i\| \leq \gamma \lambda_i^\alpha$;

(iii) $h : [0,T] \times H \longrightarrow H$ *is continuous and satisfies, uniformly in* t, $\|h(t,x) - h(t,y)\| \leq C_1 \|x - y\|$ *for all* $x, y \in H$;

(iv) $\sigma : [0,T] \times H \longrightarrow \mathcal{L}(H)$ *is continuous and satisfies, uniformly in* t, $\|\sigma(t,x) - \sigma(t,y)\| \leq C_2 \|x - y\|$ *for all* $x, y \in H$.

C_1 *and* C_2 *are arbitrary positive constants.*

Then the stochastic evolution equation

$$dX(t) = AX(t)\,dt + Ph(t, X(t))\,dt + \sigma(t, X(t))\,dB_t\,, \qquad \text{(A.1)}$$
$$X(0) = X_0 \in H, \qquad 0 \leq t \leq T$$

has a unique mild solution whose sample paths are almost surely continuous from $[0,T]$ *into* H. *Furthermore, the sample paths are Hölder-continuous from* $[\epsilon, T]$ *into* H *for all* $\epsilon > 0$. *The solution also satisfies* $\sup_{0 \leq t \leq T} E\|X(t)\|^{2p} < \infty$ *for all integer* $p \geq 1$.

PROOF. Let R^p be the Banach space of all measurable and adapted H-valued processes, defined on $[0,T] \times \Omega$ with norm

$$\|X\|_T \equiv \left\{ \sup_{0 \leq t \leq T} E\|X(t)\|^{2p} \right\}^{1/2p} < \infty.$$

R_0^p denotes the closed subspace $\{X \in R^p : X(0) = X_0\}$. To avoid too many brackets we write hereafter X_t instead of $X(t)$. Define the mapping ϕ: $R_0^p \longrightarrow R_0^p$ as follows:

$$\phi(Y)_t = U_t X_0 + \int_0^t U_{t-s} \sigma(s, Y_s)\,dB_s + \int_0^t U_{t-s} P h(s, X_s)\,ds.$$

The stochastic integral is well defined. U_{t-s} is a Hilbert-Schmidt operator for all $s < t$. This property and the Lipschitz condition on σ together ensure the existence of $I(t-\epsilon_n) \equiv \int_0^{t-\epsilon_n} U_{t-s}\sigma(s,Y_s)\,dB_s$ for $\epsilon_n > 0$. And, finally, $I(t-\epsilon_n)$ converges in norm as $\epsilon_n \downarrow 0$. The second integral is in fact deterministic; its interpretation is the following one:

$$\int_0^t U_{t-s}Ph(s,X_s)\,ds = \int_0^t U_{t-s}\Big(\sum_i \langle h(s,X_s), P^* e_i\rangle e_i\Big)\,ds.$$

It will become clear below that this expression is well defined.

In order to be able to apply a fixed point theorem we estimate $E\|\phi(Y)_t - \phi(Z)_t\|^{2p}$ for Y and $Z \in R_0^p$.

$$\begin{aligned} & E\,\|\phi(Y)_t - \phi(Z)_t\|^{2p} \\ \leq\ & 2^{2p-1}E\,\|\int_0^t U_{t-s}[\sigma(s,Y_s)-\sigma(s,Z_s)]\,dB_s\|^{2p} \\ +\ & 2^{2p-1}E\,\|\int_0^t U_{t-s}P[h(s,Y_s)-h(s,Z_s)]\,ds\|^{2p}. \end{aligned}$$

Let us call the two terms on the right hand side term (i) and term (ii), respectively. In the following we use that U_t can de represented as $U_t e_i = \exp(-\lambda_i t)e_i$ for all eigenfunctions e_i. $\{C_i : i \geq 1\}$ are arbitrary constants. If $p > 2$, we have for term (i)

$$\begin{aligned} & E\,\|\int_0^t U_{t-s}(\sigma(s,Y_s)-\sigma(s,Z_s))\,dB_s\|^{2p} \\ \leq\ & C_1\Big\{\int_0^t (E\|U_{t-s}(\sigma(s,Y_s)-\sigma(s,Z_s))\|_2^{2p})^{1/p}\,ds\Big\}^p \\ \leq\ & C_1\Big\{\int_0^t \|U_{t-s}\|_2^2\,(E\|\sigma(s,Y_s)-\sigma(s,Z_s)\|^{2p})^{1/p}\,ds\Big\}^p \\ \leq\ & C_2\Big\{\int_0^t \|U_{t-s}\|_2^2\,(E\|Y_s-Z_s\|^{2p})^{1/p}\,ds\Big\}^p \\ \leq\ & C_2\int_0^t \|U_{t-s}\|_2^2\,E\|Y_s-Z_s\|^{2p}\,ds\,\Big(\int_0^t \|U_{t-s}\|_2^2\,ds\Big)^{p-1} \\ \leq\ & C_3\int_0^t \|U_{t-s}\|_2^2\,E\|Y_s-Z_s\|^{2p}\,ds \\ =\ & C_3\int_0^t \Big(\sum_i \exp(-2\lambda_i(t-s))\Big)\,E\|Y_s-Z_s\|^{2p}\,ds \\ \leq\ & C_3\int_0^t \Big(\sum_i \exp(-\lambda_i(t-s))\Big)\,E\|Y_s-Z_s\|^{2p}\,ds, \end{aligned}$$

using Lemmas A.1 and A.2. If $p = 1$ we obtain the same result, because in this case we do not need Lemma A.2.

For term (ii) we get, using again Lemma A.2,

$$
\begin{aligned}
&E\,\|\int_0^t U_{t-s}P(h(s,Y_s)-h(s,Z_s))\,ds\|^{2p}\\
&= E\,\|\int_0^t U_{t-s}\Big(\sum_i \langle h(s,Y_s)-h(s,Z_s),P^*e_i\rangle e_i\Big)\,ds\|^{2p}\\
&= E\,\|\sum_i\int_0^t \exp(-\lambda_i(t-s))\,\langle h(s,Y_s)-h(s,Z_s),P^*e_i\rangle\,e_i\,ds\|^{2p}\\
&= E\Big(\sum_i\Big[\int_0^t \exp(-\lambda_i(t-s))\,\langle h(s,Y_s)-h(s,Z_s),P^*e_i\rangle\,ds\Big]^2\Big)^p\\
&\le E\sum_i\Big[\int_0^t \exp(-\lambda_i(t-s))\,\langle h(s,Y_s)-h(s,Z_s),P^*e_i\rangle\,ds\Big]^{2p}\\
&\qquad i^{ap}\,\Big(\sum_i(\frac{1}{i^a})^{p/(p-1)}\Big)^{p-1}\\
&\le C_4E\sum_i\int_0^t\Big(\exp(-\eta\lambda_i(t-s))\,\langle h(s,Y_s)-h(s,Z_s),P^*e_i\rangle\Big)^{2p}\,ds\\
&\qquad i^{ap}\,\Big(\int_0^t[\exp(-\theta\lambda_i(t-s))]^{2p/(2p-1)}\,ds\Big)^{2p-1}\\
&\le C_4E\sum_i\int_0^t \exp(-\lambda_i(t-s))\,\langle h(s,Y_s)-h(s,Z_s),P^*e_i\rangle^{2p}\,ds\\
&\qquad i^{ap}\,(\frac{1}{\lambda_i})^{2p-1}\\
&\le C_5E\sum_i\int_0^t \exp(-\lambda_i(t-s))\,\|h(s,Y_s)-h(s,Z_s)\|^{2p}\,ds\,\lambda_i^{2\alpha p}\\
&\qquad i^{ap}\,(\frac{1}{\lambda_i})^{2p-1}\\
&\le C_6\sum_i\int_0^t \exp(-\lambda_i(t-s))\,E\|Y_s-Z_s\|^{2p}\,ds\,\lambda_i^{1-2p(1-\alpha)}\,i^{ap}\\
&= C_6\int_0^t\Big(\sum_i\exp(-\lambda_i(t-s))\Big)\,E\|Y_s-Z_s\|^{2p}\,ds,
\end{aligned}
$$

where we have chosen $\eta = 1/2p$ and $\theta = 2p - 1/2p$; in the last line we put $ap = -(1+\delta)(1-2p(1-\alpha))$ from which it is seen that the factor $\lambda_i^{1-2p(1-\alpha)}\,i^{ap}$ cancels; for this choice of a $\sum_i(1/i^a)^{p/p-1}$ is finite, because of the condition $\alpha < 1 - 1/[2(1+\delta)]$.

The combined estimates of terms (i) and (ii) yield

$$E\|\phi(Y)_t - \phi(Z)_t\|^{2p} \le C_7 \int_0^t \Big(\sum_i \exp(-\lambda_i(t-s))\Big)\, E\|Y_s - Z_s\|^{2p}\, ds.$$

For ease of notation we will write

$$q(t) = \begin{cases} K\Big(\sum_i \exp(-\lambda_i t)\Big) & \text{if } t > 0, \quad K = \Big(\sum_i (1/\lambda_i)\Big)^{-1} \\ 0 & \text{otherwise.} \end{cases}$$

Iteration yields

$$\begin{aligned} & E\|\phi^2(Y)_t - \phi^2(Z)_t\|^{2p} \\ \le\; & C_7^2 \int_0^t q(t-s) \int_0^s q(s-u) E\|Y_s - Z_s\|^{2p}\, du\, ds \\ \le\; & C_7^2\, q * (q * E\|Y - Z\|^{2p})(t), \end{aligned}$$

where the '$*$' denotes the convolution of functions. Applying ϕ n times we obtain

$$E\|\phi^n(Y)_t - \phi^n(Z)_t\|^{2p} \le C_7^n\, q^{*n} * E\|Y - Z\|^{2p}(t)$$

so that

$$\sup_{0 \le t \le T} E\|\phi^n(Y)_t - \phi^n(Z)_t\|^{2p} \le C_7^n\, (q^{*n} * 1_{[0,\infty)})(T) \sup_{0 \le t \le T} E\|Y_t - Z_t\|^{2p}.$$

From [8, pp. 27,28] we take the result that for arbitrary T

$$C_7^n\, (q^{*n} * 1_{[0,\infty)})(T) < 1 \qquad \text{if } n \text{ is large enough.}$$

We conclude that ϕ is a contraction. A suitable version of the Picard-Banach fixed point theorem (see for example [10, p. 88]) tells us that there exists a unique fixed point. The fixed point, often simply denoted by X, is the unique mild solution of A.1 we were looking for.

What remains to be proven is the sample-path (Hölder-)continuity. This result will be given below as a consequence of Proposition A.1. □

Proposition A.1 (Kolmogorov) *Let $\{X_t : t \in \mathbf{R}_1^n\}$ be an S-valued stochastic process, where $\mathbf{R}_1^n$ denotes the unit cube in $\mathbf{R}^n$ ($n \ge 1$, $|\cdot|$ is the norm in $\mathbf{R}^n$) and (S,d) is a complete metric space. Suppose there are constants $k > 1$, $K > 0$ and $\epsilon > 0$ such that for all $s,t \in \mathbf{R}_1^n$*

$$E(d(X_t, X_s)^k) \le K|t - s|^{n+\epsilon}. \tag{A.2}$$

Then

(i) X has a continuous version;

(ii) there exist constants C and γ, depending only on n, k and ϵ, and a random variable Y such that with probability one for all $s, t \in \mathbf{R}_1^n$

$$d(X_t, X_s) \leq Y|t-s|^{\epsilon/k}(\log \frac{\gamma}{|t-s|})^{2/k}$$

and

$$EY^k \leq CK;$$

(iii) if S is a Banach space with $\|x\|$ corresponding to $d(0,x)$ for $x \in S$ and $E\|X_t\|^k < \infty$ for some t, then

$$E(\sup_{t \in \mathbf{R}_1^n} \|X_t\|^k) < \infty.$$

PROOF. The proposition is a slight extension of the Kolmogorov theorem on sample path continuity for real valued processes with n-dimensional time parameter that is given by Walsh [37, pp. 271–274]. In order to obtain the extension it is necessary to replace in (1.2) on p. 271 the expression $\psi([f(x) - f(y)]/p)$ by $\phi(d(f(x), f(y))/p)$, where ϕ satisfies the same conditions as ψ. After that the proof given there can be copied. □

Before we apply this proposition, we state a lemma.

Lemma A.3 *Let $\{\lambda_i : i \geq 1\}$ be a sequence of positive numbers such that $0 < \liminf_i \lambda_i/i^{1+\delta} \leq \limsup_i \lambda_i/i^{1+\delta} < \infty$ for some $\delta > 0$, for some $\delta > 0$. Then, for r and β such that $r \geq \beta > 1/(1+\delta)$ and positive t,*

$$\sum_{i=1}^{\infty} \frac{(1-e^{-\lambda_i t})^r}{\lambda_i^\beta} \leq C\, t^{\beta - 1/(1+\delta)},$$

where C is some constant. Note that on the right hand side the exponent of t is positive, so that the inequality yields a useful estimate for the sum, whenever t is small.

PROOF. In the following, C and $\{C_i : 1 \leq i \leq 5\}$ are positive constants, not depending on t. Split the summation into two parts:

$$\sum_{i=1}^{\infty} \frac{(1-e^{-\lambda_i t})^r}{\lambda_i^\beta} = \sum_{i \leq a} \frac{(1-e^{-\lambda_i t})^r}{\lambda_i^\beta} + \sum_{i > a} \frac{(1-e^{-\lambda_i t})^r}{\lambda_i^\beta},$$

where $a = t^{-1/(1+\delta)}$. For the first term we have

$$\sum_{i \leq a} \frac{(1-e^{-\lambda_i t})^r}{\lambda_i^\beta} \leq \sum_{i \leq a} \lambda_i^{r-\beta} t^r \leq C_1 t^r \int_0^a x^{(1+\delta)(r-\beta)}\, dx$$
$$= C_2\, t^r\, a^{(1+\delta)(r-\beta)+1} = C_3\, t^{r-(r+\beta)-1/(1+\delta)} = C_3\, t^{\beta-1/(1+\delta)}$$

and for the second term

$$\sum_{i > a} \frac{(1-e^{-\lambda_i t})^r}{\lambda_i^\beta} \leq \sum_{i > a} \lambda_i^{-\beta} \leq C_4 \int_a^\infty x^{-(1+\delta)\beta}$$
$$= C_5\, a^{-(1+\delta)\beta+1} = C_5\, t^{\beta-1/(1+\delta)}.$$

We conclude that

$$\sum_{i=1}^{\infty} \frac{(1-e^{-\lambda_i t})^r}{\lambda_i^\beta} \leq C\, t^{\beta-1/(1+\delta)}.$$

□

Proposition A.2 *Suppose that $\langle X_0, e_i\rangle \leq c/i^\nu$ with $\nu \geq (1+\delta)/2$. Then the sample paths of the unique mild solution $X(t)$ of (A.1) are Hölder-continuous with Hölder-coefficient up to $\delta/2(1+\delta)$. For general X_0 the sample paths are continuous from $[0,T]$ into H and the Hölder-continuity holds on $[\epsilon, T]$ for all $\epsilon > 0$.*

Proof. Let X_0 satisfy the condition mentioned in the proposition. X_t is an H-valued (measurable) stochastic process with real time parameter $t \in [0,T]$. We can convert the time interval into the unit interval by rescaling. In order to apply Proposition A.1, all that is left to be proven is inequality (A.2). Assume for simplicity that $0 \leq s \leq t \leq T$.

$$\begin{aligned}
&E\,\|X_t - X_s\|^{2p} \\
&\leq\ 5^{2p-1}\|U_t X_0 - U_s X_0\|^{2p} && \text{(i)}\\
&+\ 5^{2p-1} E\,\|\int_0^s (U_{t-u} - U_{s-u})\sigma(u, X_u)\, dB_u\|^{2p} && \text{(ii)}\\
&+\ 5^{2p-1} E\,\|\int_s^t U_{t-u}\sigma(u, X_u)\, dB_u\|^{2p} && \text{(iii)}\\
&+\ 5^{2p-1} E\,\|\int_0^s (U_{t-u} - U_{s-u}) Ph(u, X_u)\, du\|^{2p} && \text{(iv)}\\
&+\ 5^{2p-1} E\,\|\int_s^t U_{t-u} Ph(u, X_u)\, du\|^{2p}. && \text{(v)}
\end{aligned}$$

By Lemma A.3 term (i) satisfies

$$
\begin{aligned}
&\|U_tX_0 - U_sX_0\|^{2p} \\
&\quad \le \|U_s\|^{2p}\|(U_{t-s} - I)X_0\|^{2p} \le \Big(\sum_i [\exp(-\lambda_i(t-s)) - 1]^2 \langle X_0, e_i\rangle^2\Big)^p \\
&\quad \le C_1\Big(\sum_i \frac{1-\exp(-\lambda_i(t-s))}{\lambda_i^{2\nu/(1+\delta)}}\Big)^p \le C_2(t-s)^{p[2\nu/(1+\delta)-1/(1+\delta)]} \\
&\quad = C_2(t-s)^{p(2\nu-1)/(1+\delta)} \le C_3(t-s)^{p\delta/(1+\delta)}.
\end{aligned}
$$

For term (ii) we have

$$
\begin{aligned}
&E\,\|\int_0^s (U_{t-u} - U_{s-u})\sigma(u, X_u)\, dB_u\|^{2p} \\
&\quad \le \; C_4\Big\{\int_0^s (E\|(U_{t-u} - U_{s-u})\sigma(u,X_u)\|_2^{2p})^{1/p}\, du\Big\}^p \\
&\quad \le \; C_4\Big\{\int_0^s \sum_i (e^{-\lambda_i(t-u)} - e^{-\lambda_i(s-u)})^2 (C_5 + C_6 E\|X_u\|^{2p})^{1/p}\, du\Big\}^p \\
&\quad \le \; C_7\Big\{\int_0^s \sum_i (e^{-\lambda_i(t-u)} - e^{-\lambda_i(s-u)})^2\, du\Big\}^p \\
&\quad \le \; C_7\Big\{\sum_i \frac{1}{2\lambda_i}(1 + e^{-2\lambda_i(t-s)} - 2e^{-\lambda_i(t-s)})\Big\}^p \\
&\quad \le \; C_8\Big\{\sum_i \frac{1}{\lambda_i}(1 - e^{-\lambda_i(t-s)})\Big\}^p \\
&\quad \le \; C_9(t-s)^{p\delta/(1+\delta)}.
\end{aligned}
$$

To obtain the last estimates we used Lemma's A.1 and A.3 the growth condition on σ (which is implied by the Lipschitz condition) and the convexity of the function e^{-x} (so that $-e^{2\lambda_i t} + 2e^{-\lambda_i(t+s)} - e^{2\lambda_i s} < 0$).

The third term can be estimated in an equivalent way by

$$
\begin{aligned}
&E\,\|\int_s^t U_{t-u}\sigma(u, X_u)\, dB_u\|^{2p} \\
&\quad \le \; C_{10}\Big\{\int_s^t \sum_i e^{-2\lambda_i(t-u)}\, du\Big\}^p \le C_{11}(t-s)^{p\delta/(1+\delta)}.
\end{aligned}
$$

Term (iv) is bounded by

$$
E\,\|\int_0^s (U_{t-u} - U_{s-u})Ph(u, X_u)\, du\|^{2p}
$$

$$
\begin{aligned}
&= E\|\int_0^s (U_{t-u} - U_{s-u})\Big(\sum_i \langle h(u,Y_u), P^* e_i\rangle e_i\Big)\, du\|^{2p} \\
&= E\|\sum_i \int_0^s (e^{-\lambda_i(t-u)} - e^{-\lambda_i(s-u)})\langle h(u,X_u), P^* e_i\rangle e_i\, du\|^{2p} \\
&= E\Big(\sum_i \Big[\int_0^s (e^{-\lambda_i(t-u)} - e^{-\lambda_i(s-u)})\langle h(u,X_u), P^* e_i\rangle\, du\Big]^2\Big)^p \\
&\leq E\sum_i \Big[\int_0^s (e^{-\lambda_i(t-u)} - e^{-\lambda_i(s-u)})\langle h(u,X_u), P^* e_i\rangle\, du\Big]^{2p} \\
&\qquad i^{ap}\Big(\sum_i (\frac{1}{i^a})^{p/(p-1)}\Big)^{p-1} \\
&\leq C_{12} E\sum_i \int_0^s \Big(e^{-\eta\lambda_i(s-u)}\langle h(u,X_u), P^* e_i\rangle\Big)^{2p}\, du \\
&\qquad i^{ap}\Big(\int_0^s (e^{-\theta\lambda_i(s-u)} - e^{-\lambda_i(t-u)+\eta\lambda_i(s-u)})^{2p/(2p-1)}\, du\Big)^{2p-1} \\
&\leq C_{13} E\sum_i \int_0^s \exp(-\lambda_i(s-u))\langle h(u,X_u), P^* e_i\rangle^{2p}\, du \\
&\qquad i^{ap}\Big(\frac{1}{\theta\lambda_i}(1 - e^{-\lambda_i(t-s)})\Big)^{2p-1} \\
&\leq C_{14} E\sum_i \int_0^s \exp(-\lambda_i(s-u))\, du\, \lambda_i^{2\alpha p} \\
&\qquad i^{ap}\Big(\frac{1}{\theta\lambda_i}(1 - e^{-\lambda_i(t-s)})\Big)^{2p-1} \\
&\leq C_{15}\sum_i (1 - e^{-\lambda_i(t-s)})^{2p-1}\, \lambda_i^{2p(\alpha-1)}\, i^{ap} \\
&\leq C_{16}\sum_i (1 - e^{-\lambda_i(t-s)})^{2p-1}\, \lambda_i^{2p(\alpha-1)+ap/(1+\delta)} \\
&\leq C_{17}(t-s)^{2p(1-\alpha)-(ap+1)/(1+\delta)} \\
&\leq C_{18}(t-s)^{p[1+(1/2-\alpha)l]\delta/(1+\delta)} \qquad \text{for } 0 < l < 1,
\end{aligned}
$$

where we have set $ap = -(1+\epsilon p) + 2p(1-\alpha)(1+\epsilon)$ with $\epsilon = \delta(1-l/2)$. We also used that

$$
\begin{aligned}
&(e^{-\theta\lambda_i(s-u)} - e^{-\lambda_i(t-u)+\eta\lambda_i(s-u)})^{2p/2p-1} \\
&= (e^{-\theta\lambda_i(s-u)} - e^{-\theta\lambda_i(t-u)-\eta\lambda_i(t-s)})^{2p/2p-1} \\
&= (e^{-\theta\lambda_i(s-u)} - e^{-\theta\lambda_i(s-u)-\lambda_i(t-s)})^{2p/2p-1} \\
&= (e^{-\theta\lambda_i(s-u)}(1 - e^{-\lambda_i(t-s)}))^{2p/2p-1}
\end{aligned}
$$

$$\leq \; e^{-\theta\lambda_i(s-u)}(1-e^{-\lambda_i(t-s)}).$$

These choices and the condition on α yield that $ap > p - 1 + p\delta(1-l)$, which ensures the finiteness of $\sum_i (1/i^a)^{p/p-1}$. Note that here the value of a is different from the value it had in the proof of Theorem A.1.

Finally, the fifth and last term is dealt with in the same way as term four. It leads to exactly the same estimate. Combining all these estimates, we obtain that for $0 \leq s \leq t \leq T$

$$E\|X_t - X_s\|^{2p} \leq C_{19}(t-s)^{pf\delta/(1+\delta)} \qquad \text{with} \begin{cases} 0 < f < 1 & \text{if } \alpha \geq 1/2 \\ f = 1 & \text{if } \alpha < 1/2. \end{cases}$$

As before $\{C_i : i \geq 1\}$ is a family of positive constants, not depending on t.

If the condition on X_0 does not hold, the proof remains valid with respect to the interval $[\epsilon, T]$. Furthermore, as the two integrals that appear on the right hand side of the integral equation (of which X is the solution) are almost surely Hölder-continuous, independently of the condition on X_0, continuity of $U_t X_0$ for all $t \in [0,T]$ suffices for the sample path continuity of X. □

A.2

The second part of this appendix contains proofs that were omitted from Section 4.3.

Consider the stochastic evolution equation

$$dR(t) = K\frac{d^2}{dx^2}R(t)\,dt - c_0\frac{d}{dx}R(t)\,dt + \sigma\,dB_t\,, \qquad \text{(A.3)}$$
$$R(0) = R_0, \qquad 0 \leq t \leq T.$$

We set $A = Kd^2/dx^2$ and $\mathcal{D}(A) = \{f \in L^2[0,M] : f'' \in L^2[0,M], f(0) = f(M), f'(0) = f'(L)\}$. R_0 satifies $\langle R_0, e_0\rangle = 0$, i.e. $R_0 \in L^2_0[0,M]$, where $L^2_0[0,M] = \{f \in L^2[0,M] : \langle f, e_0\rangle = 0\}$. $B_t = \sum_{i\geq 1} b^s_i(t)\phi_i + b^c_i(t)\psi_i$, where $\{b^s_i(t), b^c_i(t)\}$ is a collection of independent standard Brownian motions; $\phi_i(x) = \sqrt{2/M}\sin(2\pi i x/M)$ and $\psi_i(x) = \sqrt{2/M}\cos(2\pi i x/M)$. The 's' and the 'c' refer to 'sine' and 'cosine', respectively. $\{\phi_i : i \geq 1\}$ and $\{\psi_i : i \geq 1\}$, supplemented with $\{e_0 = 1/\sqrt{M}\}$, are collectively written as $\{e_i\}$, which is an orthonormal basis for $L^2[0,M]$. Also $\{e_i\}$ is the set of eigenvectors of A. Depending on the context $\{\lambda_i\}$ denotes the eigenvalues corresponding to $\{e_i\}$ or corresponding to $\{\phi_i : i \geq 1\}$ and

$\{\psi_i : i \geq 1\}$, which is, of course, only a matter of numbering. Sometimes it is convenient to assume that B_t is a true cylindrical Brownian motion (i.e. there is also a noise component corresponding e_0). In that case we choose σ equal to a constant times the projection operator that projects onto $L_0^2[0, M]$: $\sigma f = \sigma_0 \sum_{i \geq 1} \langle f, \phi_i \rangle \phi_i + \langle f, \psi_i \rangle \psi_i$ for $f \in H$; K and σ_0 are postive constants. We note, however, that until Proposition A.4 we only use a Lipschitz condition on σ, viz. $\sigma: [0, T] \times L^2[0, M] \longrightarrow \mathcal{L}(L^2[0, M])$ is continuous and satisfies, uniformly in t, $\|\sigma(t, x) - \sigma(t, y)\| \leq C \|x - y\|$ for all $x, y \in L^2[0, M]$.

According to Theorem A.1 the equation (A.3) has a unique mild solution whose sample paths are almost surely continuous from $[0, T]$ into H. Furthermore, the sample paths are Hölder-continuous from $[\epsilon, T]$ into H for all $\epsilon > 0$. The solution also satisfies $\sup_{0 \leq t \leq T} E\|R(t)\|^{2p} < \infty$ for all integer $p \geq 1$.

We define $A_c = K d^2/dx^2 - c_0\, d/dx$ (with $\mathcal{D}(A_c) = \mathcal{D}(A)$). It is easily seen that A_c generates a semigroup, say V_t. V_t has a rather simple characterization, viz.

$$V_t e_i(x) = e^{-\lambda_i K t} e_i(x - c_0 t),$$

where $\{e_i\}$ is the same orthonormal basis as above; note that the $\{e_i\}$ are not eigenvectors of A_c. For general $f \in L^2[0, M]$ we have

$$\begin{aligned} V_t f(x) &= \int_0^M q(t, x, y) f(y) dy \qquad \text{with} \\ q(t, x, y) &= \frac{1}{M} + \sum_{i=1}^{\infty} \Big(\phi_i(x - c_0 t)\phi_i(y) + \psi_i(x - c_0 t)\psi_i(y) \Big) e^{-\lambda_i K t}. \end{aligned}$$

The correctness of this representation can be verified by direct calculation.

Of course, we must have that $R(t)$ can be written as

$$R(t) = V_t R_0 + \int_0^t V_{t-s} \sigma \, dB_s,$$

The following lemma shows that this is true.

Lemma A.4 *The solution of the equation A.3 satisfies, almost surely,*

$$R(t) = V_t R_0 + \int_0^t V_{t-s} \sigma \, dB_s, \tag{A.4}$$

PROOF. We must show that $R(t)$ as given by A.4 satisfies

$$R(t) = U_t R_0 - c_0 \int_0^t U_{t-s} \frac{d}{dx} R(s)\, ds + \int_0^t U_{t-s} \sigma\, dB_s\,,$$
$$R(0) = R_0 \in L_0^2[0,M], \qquad 0 \le t \le T.$$

Introduce the 'shift' operator T_t on $L^2[0,M]$:

$$T_t f(x) = f(x, x - c_0 t).$$

T_t is a (semi)group; its generator is $-c_0\, d/dx$ with domain $\{f \in L^2[0,M]: f' \in L^2[0,M],\ f(0) = f(M)\}$. We have the following relation

$$V_t = T_t U_t.$$

Substitution of the alternative formula of $R(t)$ into $-c_0 \int_0^t U_{t-s}\, d/dx\, R(s)\, ds$ yields

$$\begin{aligned}
&-c_0 \int_0^t U_{t-s} \frac{d}{dx} R(s)\, ds \\
&= -c_0 \int_0^t U_{t-s} \frac{d}{dx} \Big(V_s R_0 + \int_0^s V_{s-u} \sigma\, dB_u \Big)\, ds \\
&= -c_0 \int_0^t U_{t-s} \frac{d}{dx} T_s U_s R_0\, ds \\
&\qquad -c_0 \int_{s=0}^t \int_{u=0}^s T_{s-u} U_{s-u} \sigma\, dB_u\, ds \\
&= \int_0^t U_{t-s} \frac{d}{ds} T_s U_s R_0\, ds \\
&\qquad -c_0 \int_{u=0}^t \int_{w=0}^{t-u} U_{t-w-u} \frac{d}{dx} T_w U_w\, ds \sigma\, dB_u \\
&= T_t U_t R_0 - U_t R_0 + \int_{u=0}^t \int_{w=0}^{t-u} U_{t-w-u} \frac{d}{dw} T_w U_w\, ds \sigma\, dB_u \\
&= V_t R_0 - U_t R_0 + \int_0^t T_{t-u} U_{t-u} \sigma\, dB_u - \int_0^t U_{t-u} \sigma\, dB_u,
\end{aligned}$$

where we applied a Fubini theorem. The equalities hold almost surely. Rearranging of the terms gives the desired equivalence. □

Lemma A.5 *Let $e_i(x)$ denote either $\phi_i(x) = \sqrt{2/M} \sin(2\pi i x/M)$ or $\psi_i(x) = \sqrt{2/M} \cos(2\pi i x/M)$. Then, if the real numbers a, p satisfy $0 < a \le 1$ and $ap > 1$, we have for real x and y,*

$$\sum_{i=1}^{\infty} \Big| \frac{e_i(x) - e_i(y)}{i^a} \Big|^p \le C\, |x - y|^{ap-1}.$$

PROOF. $|e_i(x) - e_i(y)|$ can be bounded by $\sqrt{2/M}(2 \wedge 2\pi i|x-y|/M)$, so that

$$\begin{aligned}
&\sum_{i=1}^{\infty}\Big|\frac{e_i(x) - e_i(y)}{i^a}\Big|^p \\
&\leq C_1 \sum_{i=1}^{\infty}\Big(\frac{(2 \wedge 2\pi i|x-y|/M)}{i^a}\Big)^p \\
&\leq C_2\Big(\sum_{i=1}^{[M/(2\pi|x-y|)]}(\frac{i|x-y|}{i^a})^p + \sum_{[M/(2\pi|x-y|)]}^{\infty}(\frac{1}{i^a})^p\Big) \\
&\leq C_3\Big(|x-y|^p \int_0^{M/(2\pi|x-y|)} z^{p(1-a)}\,dz + \int_{M/(2\pi|x-y|)}^{\infty} z^{-ap}\,dz\Big) \\
&\leq C_4\Big(|x-y|^p(1/|x-y|)^{p(1-a)+1} + (1/|x-y|)^{-ap+1}\Big) \\
&\leq C_5\,|x-y|^{ap-1};
\end{aligned}$$

$[M/(2\pi|x-y|)]$ denotes the entier of $M/(2\pi|x-y|)$. □

Proposition A.3 (=Proposition 4.1) *Write $R(t,x)$ instead of $R(t)$ to indicate that we view each realization of R as a real valued function of two parameters, t and x. Then $R(t,x)$ is almost surely jointly Hölder-continuous on $[\epsilon,T] \times [0,M]$ for every $\epsilon > 0$.*

PROOF. We will use the alternative way of writing $R(t)$ (Lemma A.4). Since we want to apply once more Proposition A.1, we will try to find bounds of expressions like $E|\int_0^t V_{t-u}\sigma\,dB_u(x) - \int_0^s V_{s-u}\sigma\,dB_u(y)|^{2p}$ for $0 \leq x,y \leq M$, $0 \leq t,s \leq T$ and integer $p \geq 1$. First, we show that the expression $\int_0^t V_{t-u}\sigma\,dB_u(x)$ makes sense. By definition of the stochastic integral we have that

$$\Big\langle \int_0^t V_{t-u}\sigma\,dB_u, e_i\Big\rangle = \int_0^t \langle \sigma^* V_{t-u}^* e_i, dB_u\rangle$$

Approximate $\int_0^t V_{t-u}\sigma\,dB_u\,(x)$ by $\sum_{i=0}^n \langle \int_0^t V_{t-u}\sigma\,dB_u, e_i\rangle e_i(x)$. This expression is well defined for all n. Furthermore, by Lemma A.1,

$$\begin{aligned}
&E\Big(\sum_{i=0}^{n}\Big\langle\int_0^t V_{t-u}\sigma\,dB_u, e_i\Big\rangle e_i(x)\Big)^{2p} \\
&= E\Big\langle\int_0^t V_{t-u}\sigma\,dB_u, \sum_{i=0}^{n} e_i(x)e_i\Big\rangle^{2p}
\end{aligned}$$

$$
\begin{aligned}
&= E\Big(\int_0^t \langle \sigma^* V_{t-u}^*\Big(\sum_{i=0}^n e_i(x)e_i\Big), dB_u\rangle\Big)^{2p} \\
&\leq C_1\Big(\int_0^t (E\|\sigma^* V_{t-u}^*\Big(\sum_{i=0}^n e_i(x)e_i\Big)\|^{2p})^{1/p}\, du\Big)^p \\
&\leq C_2\Big(\int_0^t \Big(\sum_{i=0}^n e_i(x)^2 e^{-2\lambda_i K(t-u)}\Big)\, du\Big)^p \\
&\leq C_3\Big(\sum_{i=1}^n \frac{e_i(x)^2}{\lambda_i K}\Big)^p + C_4.
\end{aligned}
$$

The summation converges for $n \to \infty$, so that the interpretation of the expression $\int_0^t V_{t-u}\sigma dB_u(x)$ is obvious. Replacing $e_i(x)$ by $e_i(x) - e_i(y)$ immediately yields, using Lemma A.5,

$$
\begin{aligned}
& E|\int_0^t V_{t-u}\sigma dB_u(x) - \int_0^t V_{t-u}\sigma dB_u(y)|^{2p} \\
&\leq C_5\Big(\sum_{i=1}^\infty \frac{(e_i(x) - e_i(y))^2}{\lambda_i K}\Big)^p \\
&\leq C_6|x-y|^p.
\end{aligned}
$$

For fixed x we calculate (assuming that $s \leq t$)

$$
\begin{aligned}
& E\,|\int_0^t V_{t-u}\sigma\, dB_u(x) - \int_0^s V_{s-u}\sigma\, dB_u(x)|^{2p} \\
&\leq 2^{2p-1} E\,|\int_0^s (V_{t-u} - V_{s-u})\,\sigma\, dB_u(x)|^{2p} \\
&+ 2^{2p-1} E\,|\int_s^t V_{t-u}\sigma\, dB_u(x)|^{2p} \\
&\leq C_7\Big(\int_0^s (E\|\sigma^*(V_{t-u}^* - V_{s-u}^*)\Big(\sum_{i=0}^n e_i(x)e_i\Big)\|^{2p})^{1/p}\, du\Big)^p \\
&+ C_8\Big(\int_s^t (E\|\sigma^* V_{t-u}(\sum_{i=0}^n e_i(x)e_i)\|^{2p})^{1/p} du)^p \\
&\leq C_9\Big(\int_0^s \Big(\sum_{i=0}^\infty (e^{-\lambda_i K(t-u)} - e^{-\lambda_i K(s-u)})^2\Big)\, du\Big)^p \\
&+ C_{10}\Big(\int_s^t \sum_{i=0}^\infty e^{-2\lambda_i K(t-u)}\, du\Big)^p \\
&\leq C_{11}|t-s|^{p/2}.
\end{aligned}
$$

In the last line we used results obtained earlier in the proof of Proposition A.2.

Combining these results we have

$$\begin{aligned}
& E\,|\int_0^t V_{t-u}\sigma\,dB_u(x) - \int_0^s V_{s-u}\sigma\,dB_u(y)|^{2p} \\
\leq\; & 2^{2p-1}E\,|\int_0^t V_{t-u}\sigma\,dB_u(x) - \int_0^s V_{s-u}\sigma\,dB_u(x)|^{2p} \\
+\; & 2^{2p-1}E\,|\int_0^s V_{s-u}\sigma\,dB_u(x) - \int_0^s V_{s-u}\sigma\,dB_u(y)|^{2p} \\
\leq\; & C_{12}|x-y|^p + C_{13}|t-s|^{p/2} \leq C_{14}(|x-y| + |t-s|^{1/2})^p.
\end{aligned}$$

The numbering of the constants has meaning only within this proof. Because of the restriction to the time interval to $[\epsilon, T]$ we can easily obtain the same estimates for the expressions $V_t R_0(x) - V_t R_0(y)$ and $V_t R_0(x) - V_s R_0(x)$.

We provide the square $[\epsilon,T]\times[0,M]$ with the metric $d((t,x),(s,y)) = |x-y|+|t-s|^{1/2}$, so that the condition (A.2) is satisfied with $k = 2p$ and $\epsilon = p-2$. The coefficient of Hölder-continuity can be chosen in the interval $(0, 1/2)$. Note that for fixed x we have a real valued proces with one dimensional time parameter; with the usual metric (i.e. $d(t,s) = |t-s|$) the Hölder coefficient is smaller than $1/4$. For fixed t we find almost the same situation: now the Hölder coefficient is smaller than $1/2$. □

Before we proceed we make some comments concerning this result on the joint continuity of the sample paths. Rather analogous results can be found in work by Funaki [9, pp. 180–183] and Walsh [37, pp. 323–326]. The proposition and proof given here intend to be a combination of both these results, being as general as the first and as simple as the second.

The joint continuity has been proven for $(t,x) \in [\epsilon,T]\times[0,M]$, for every $\epsilon > 0$. This was done to avoid unnecessary conditions on R_0.

One might wonder what can be said in the case of a more general, non-linear convection term as in example 4.3. In this situation the convection effect cannot easily be incorporated into the semigroup, so that the convection term must be dealt with separately. It appears that for this term joint Hölder-continuity can also be proven under some extra condition on the function h. For instance the following will do. Let as before $A = Kd^2/dx^2$ with suitably chosen domain: $\mathcal{D}(A) = \{f \in L^2[0,M] : f'' \in L^2[0,M],\ f(0) = f(M) \text{ and } f'(0) = f'(M)\}$, but other choices are also possible (as the one in example 4.3). Now define, for positive α, $\mathcal{D}_\alpha \equiv \{f \in L^2[0,M] : \sum_i \lambda_i^\alpha \langle f, e_i\rangle^2 < \infty\}$. Or, if we first define the Hilbertian norm $\|f\|_\alpha \equiv \|A^\alpha f\|$ for $f \in \text{span}\{e_i\}$, then $\mathcal{D}_\alpha$ can also be defined as the closure of the span of $\{e_i\}$ with respect to the norm $\|\cdot\|_\alpha$. It is obvious that $\mathcal{D}_\alpha$ is a subset of $L^2[0,M]$ consisting of (rather) smooth functions. The larger α is the smoother are the functions. Of course $\mathcal{D}_0 = L^2[0,M]$. This procedure may also be carried out for negative α. The result is a family of spaces of distributions of which $L^2[0,M]$ is a subspace. For a more elaborate treatment of these ideas we refer to a monograph of Itô [15]. The stochastic heat equation with non-linear convection term can be solved considering $\mathcal{D}_\alpha$-valued processes $(0 \le \alpha < 1/4)$ by exactly the same method as was used above. We can conclude that the solution we already obtained is not only $L^2[0,M]$-valued, but also $\mathcal{D}_\alpha$-valued. The 'regularity' of h we need to prove the joint continuity now can be captured by requiring that $h: \mathcal{D}_\alpha \longrightarrow \mathcal{D}_\alpha$ and $\|h(t,x)\|_\alpha \le C_1 + C_2\|x\|_\alpha$ for all $x \in \mathcal{D}_\alpha$.

In the following we will explicitly use that σ is a constant and that B_t is a modification of truly white noise.

Proposition A.4 (=Proposition 4.2) *Suppose R_0 is not a fixed element of $L_0^2[0,M]$, but an $L_0^2[0,M]$-valued random variable that is $\mathcal{F}_0$-measurable. Then there is a unique measure μ for R_0 such that $R(t)$ is a stationary $L_0^2[0,M]$-valued process. Moreover, if R_0 is distributed according to μ, $R(t,x)$ is a real-valued process which is stationary under shifts in $\mathbf{R}^2$.*

PROOF. Again we make use of the 'shift' operator T_t on $L^2[0,M]$:

$$T_t f(x) = f(x, x - c_0 t)$$

(see also the proof of Lemma A.4). Define

$$Z(t) = U_t R_0 + \sigma \int_0^t U_t - u\, dB_u.$$

It can be verified that $R(t) = T_t Z_t$. Proving the stationarity of Z_t is rather easy. Using the decomposition of the (modified) cylindrical Brownian motion we obtain

$$\int_0^t U_{t-u}\sigma\, dB_u = \sigma \int_0^t \sum_{i\ge 1} \exp(-\lambda_i K(t-u)) e_i\, db_i,$$

so that

$$\left\langle \int_0^t U_{t-u}\sigma\, dB_u, e_i \right\rangle = \sigma \int_0^t \exp(-\lambda_i K(t-u))\, db_i,$$

where we (essentially) use the Fubini theorem for finite-dimensional stochastic integrals. Our first conclusion is that the stochastic integral $\int_0^t V_{t-u}\sigma\, dB_u$ is a centred Gaussian process. This follows from the properties of the standard one-dimensional Brownian motion. The covariance functional of the integral is given by

$$\begin{aligned} r(t,s;e_i,e_j) &= \delta_{i,j}\,\sigma^2 \int_0^s \exp(-\lambda_j K(t+s-2u))\, du \\ &= \delta_{i,j}\,\sigma^2 \exp(-\lambda_j K(t-s))\, \frac{1-\exp(-2\lambda_j K s)}{2\lambda_j K}, \end{aligned}$$

where we assumed $s \le t$.

Now it is easy to define R_0 in such a way that the process $Z(t)$ becomes (weakly) stationary. Denoting for the moment the stochastic integral by $SI(t)$, we find that the covariance functional of the process $Z(t)$ is given by

$$\begin{aligned} &E(\langle Z(t), e_i\rangle \langle Z(s), e_j\rangle) \\ &\quad = E(\langle V_t R_0, e_i\rangle \langle V_s R_0, e_j\rangle) + E(\langle SI(t), e_i\rangle \langle SI(s), e_j\rangle). \end{aligned}$$

The first term equals

$$e^{-(\lambda_i t + \lambda_j s)K}\, E(\langle R_0, e_i\rangle \langle R_0, e_j\rangle)$$

If we choose R_0 such that $\langle R_0, e_i\rangle = 0$ and $\langle R_0, e_i\rangle \sim \mathrm{N}(0, \sigma^2/(2\lambda_i K))$ for $i \ge 1$ (such that all these real random variables are independent), then this expression simplifies to

$$\delta_{i,j}\,\sigma^2\, e^{-(\lambda_j(t-s)K}\, \frac{e^{-2\lambda_j K s}}{2\lambda_j K}.$$

For this choice of R_0 the covariance functional of $R(t)$ is clearly invariant under time shift, as it only depends on the time difference $(t-s)$. It is also clear from the explicit construction of R_0 that the probability measure that makes the process $Z(t)$ stationary is unique.

When we return to the process $R(t)$, we split the set $\{e_i\}$ into $\{\phi_i : i \ge 1\}$ and $\{\psi_i : i \ge 1\}$. We get, for $t \ge s$,

$$r(t,s;\phi_i,\phi_j) = E\langle R(t), \phi_i\rangle \langle R(t), \phi_j\rangle$$

$$
\begin{aligned}
&= E\langle Z(t), T_{-t}\phi_i\rangle\langle Z(t), T_{-t}\phi_j\rangle \\
&= E\Big(\langle Z(t),\phi_i\rangle\psi_i(c_0t) + \langle Z(t),\psi_i\rangle\phi_i(c_0t)\Big) \\
&\qquad \Big(\langle Z(t),\phi_j\rangle\psi_j(c_0t) + \langle Z(t),\psi_j\rangle\phi_j(c_0t)\Big) \\
&= \delta_{i,j}\,\sigma^2\, e^{-(\lambda_j(t-s)K}\ \cos(2\pi i c_0(t-s)/M).
\end{aligned}
$$

Equivalently, we have

$$
r(t,s;\phi_i,\psi_j = \delta_{i,j}\,\sigma^2\, e^{-(\lambda_j(t-s)K}\ \cos(2\pi i c_0(t-s)/M);
$$

$r(t,s;\psi_i,\psi_j) = r(t,s;\phi_i,\phi_j)$ and $r(t,s;\psi_i,\phi_j) = r(t,s;\phi_i,\psi_j)$. We see that these covariance functionals also depend only on the time difference $(t-s)$.

The second assertion of the proposition is proven in much the same way as the first one. It can be shown that

$$
EZ(t,x)Z(s,y) = \frac{2\sigma^2}{M}\sum_{i=1}^{\infty}\frac{1}{2\lambda_i K}\exp(-\lambda_i K(t-s)),
$$

so that

$$
\begin{aligned}
&EZ(t,x)Z(s,y) \\
&= \frac{2\sigma^2}{M}\sum_{i=1}^{\infty}\frac{1}{2\lambda_i K}\exp(-\lambda_i K(t-s))\ \cos(2\pi i\frac{x-y-c_0(t-s)}{M}),
\end{aligned}
$$

for $0 \le x, y \le M$ and $0 \le s \le t$. □

Corollary A.1 (=Proposition 4.3) *The stationary solution of the equation (A.3) can be written as*

$$
R(t,x) = \sum_{i=1}^{\infty} a_i^s(t)\phi_i(x-c_0t) + a_i^c(t)\psi_i(x-c_0t),
$$

where $\phi_i(x) = \sqrt{2/M}\sin(2\pi ix/M)$ *and* $\psi_i(x) = \sqrt{2/M}\cos(2\pi ix/M)$. $\{a_i^s(t), a_i^c(t) : i \ge 1\}$ *is a family of mutually independent Ornstein-Uhlenbeck processes, i.e. each* $a_i^s(t)$ *or* $a_i^c(t)$ *is a centred stationary Gaussian process having covariance (function)*

$$
Ea_i^s(t)a_i^s(s) = Ea_i^c(t)a_i^c(s) = \frac{\sigma^2 M^2}{8\pi^2 K i^2}\exp(-\frac{4\pi^2 K i^2}{M^2}|\Delta|),
$$

where $\Delta = (t-s)$. Thus $R(t,x)$ is also a centred stationary Gaussian process. Its covariance function, $r(\Delta, z)$, reads

$$r(\Delta, z) = \frac{\sigma^2 M}{4\pi^2 K} \sum_{i=1}^{\infty} \frac{1}{i^2} \exp(-\lambda_i K|\Delta|) \cos(2\pi i \frac{z - c_0\Delta}{M}), \tag{A.5}$$

where $\Delta = (t-s)$ and $z = x - y$; (t,x) and (s,y) are two points in the plane.

PROOF. From the previous proposition we have that

$$\begin{aligned} Z(t) &= U_t R_0 + \sigma \int_0^t U_{t-u}\, dB_u \\ &= \sum_{i \geq 1} \Big[\langle R_0, e_i\rangle e_i + \sigma \int_0^t \exp(-\lambda_i K(t-u))\, db_i(u) e_i\Big] \\ &= \sum_{i \geq 1} \Big[\langle R_0, e_i\rangle + \sigma \int_0^t \exp(-\lambda_i K(t-u))\, db_i(u)\Big] e_i \\ &= \sum_{i \geq 1} a_i(t) e_i, \end{aligned}$$

using the fact that $\langle R_0, e_i\rangle + \int_0^t \sigma \exp(-\lambda_i K(t-u)) db_i(u)$ is a one-dimensional Ornstein-Uhlenbeck a_i process having covariance function

$$E a_i(t) a_i(s) = \frac{\sigma^2}{2\lambda_i K} \exp(-\lambda_i K|t-s|).$$

Therefore,

$$R(t,x) = T_t Z(t,x) = \sum_{i \geq 1} a_i(t) e_i(x - c_0 t).$$

In terms of ϕ and ψ we obtain the result stated in the corollary.

The covariance function of $R(t,x)$ follows either from this representation of R or from the proof of the previous proposition. □

Finally, we consider the stationary, centred Gaussian process

$$R_m(t,x) = \sum_{i=m}^{\infty} a_i^s(t)\phi_i(x - c_0 t) + a_i^c(t)\psi_i(x - c_0 t),$$

where $\{a_i^s(t), a_i^c(t) \colon i \geq m\}$ is a family of mutually independent Ornstein-Uhlenbeck processes, i.e. each $a_i^s(t)$ or $a_i^c(t)$ is a centred stationary Gaussian process having covariance (function)

$$E a_i^s(t) a_i^s(s) = E a_i^c(t) a_i^c(s) = \frac{\sigma^2 M^2}{8\pi^2 K i^2} \exp(-\frac{4\pi^2 K i^2}{M^2}|\Delta|),$$

for $\Delta = (t-s)$. As before $\phi_i(x) = \sqrt{2/M}\sin(2\pi i x/M)$ and $\psi_i(x) = \sqrt{2/M}\cos(2\pi i x/M)$. K, c_0 and M are positive constants; m is a large integer such that $M/m = S$.

The covariance function, $r_m(\Delta, z)$, reads

$$r_m(\Delta, z) = \frac{\sigma^2 M}{4\pi^2 K} \sum_{i=m}^{\infty} \frac{1}{i^2} \exp(-\lambda_i K|\Delta|)\cos(2\pi i \frac{z - c_0\Delta}{M}), \tag{A.6}$$

where $\Delta = (t-s)$, $z = x - y$ and (t, x) and (s, y) are two points in the plane.

We now let M and m tend to infinity, keeping S constant. First we note that the covariance function r_m concerges pointwise to the covariance function r, which is defined as

$$r(\Delta, z) = \frac{\sigma^2 S}{4\pi^2 K} \int_1^{\infty} \frac{1}{l^2} \exp(-\frac{4\pi^2 l^2}{S^2} K|\Delta|)\cos(2\pi l \frac{z - c_0\Delta}{S})\, dl. \tag{A.7}$$

Theorem A.2 (=Theorem 4.5) *Let R be the stationary, centred, Gaussian process characterized by the covariance function r. Then R_m converges in distribution to R as m tends to infinity, if R_m as well as R are viewed as $C([0,T] \times [0,L])$-valued random variables, where $C([0,T] \times [0,L])$ is endowed with the Borel σ-algebra.*

PROOF. We already know that for every finite collection $\{(t_i, x_i)\}$ the covariance matrix of $\{R_m(t_i, x_i)\}$ converges to the covariance matrix of $\{R(t_i, x_i)\}$. Therefore, $\{R_m(t_i, x_i)\}$ converges in distribution to $\{R(t_i, x_i)\}$ or, in other words, the finite dimensional distributions of R_m converge to the finite dimensional distributions of R. Thus, if we can prove tightness of $\{R_m\}$, we can apply Prohorov's theorem and conclude that R_m converges in distribution to R.

A sufficient condition for tightness is that there are positive constants α, β, γ and C such that, uniformly in m,

$$\begin{aligned} E\,|R_m(0,0)|^{\gamma} &< \infty \\ E\,|R_m(t,x) - R_m(s,y)|^{\alpha} &\leq C\,|(t,x) - (s,y)|^{2+\beta} \end{aligned}$$

for all (t, x), $(s, y) \in [0,T] \times [0,L]$ (see [9, p. 132]). Write

$$R(t) = V_t R_0 + \sigma \int_0^t V_{t-u}\, dB_u,$$

where R_0 has the invariant distribution (see Proposition A.4). From the proof of Proposition 4.1 we have that

$$\begin{aligned} & E\,\Big|\int_0^t V_{t-u}\sigma\, dB_u(x) - \int_0^s V_{s-u}\sigma\, dB_u(y)\Big|^{2p} \\ \leq \;& C_1\Big(\sum_{i=1}^{\infty}\frac{(e_i(x)-e_i(y))^2}{\lambda_i K}\Big)^p + C_2\,|t-s|^{1/2})^p. \end{aligned}$$

C_1 and C_2 depend only on p. Direct calculation or careful application of Lemma A.5 (with $a = 1$ and $p = 2$) shows that

$$\Big(\sum_{i=1}^{\infty}\frac{(e_i(x)-e_i(y))^2}{\lambda_i K}\Big)^p \leq C_3\,|x-y|,$$

where C_3 does not depend on m.

The same estimate is obtained for $E\,|V_t R_0(x) - V_s R_0(y)|^{2p}$. Furthermore,

$$\sup_{m\in\mathbf{N}} E\,|R_m(0,0)|^2 = \sup_{m\in\mathbf{N}}\Big(\frac{\sigma^2 m S}{4\pi^2 K}\sum_{i=m}^{\infty}\frac{1}{i^2}\Big) < \infty.$$

We conclude that the sequence $\{R_m\}$ is tight. □

For later use we give the expansion of r around $(0,0)$.

Proposition A.5 *The covariance function r is expanded around $(0,0)$ as follows:*

$$\begin{aligned} r(0,z) &= A\,(1 - \frac{\pi^2}{S}|z| + \frac{2\pi^2}{S^2}z^2 + o(z^2)) \\ r(\Delta,0) &= A\,(1 - \sqrt{\pi a|t|} + a|t| + o(|t|)). \end{aligned}$$

where $A = \dfrac{\sigma^2 S}{4\pi^2 K}$ *and* $a = K\dfrac{4\pi^2}{S^2}$.

PROOF. We have from Appendix B, Proposition B.2 (cf. also Subsection 6.3.2)

$$r(0,z) = A\,\Big(\cos(p) + p\Big[\frac{\pi}{2} - \sum_{j=0}^{\infty}\frac{(-1)^j p^{2j+1}}{(2j+1)(2j+1)!}\Big]\Big),$$

where $p = 2\pi z/S$. This result immediately implies the expansion given above. Further, $r(\Delta,0)$ can be written as

$$r(\Delta_i,0) = A\,I_2(a\Delta_i, \frac{2\pi c_0}{aS}),$$

where

$$I_2(p,b) = \int_1^\infty \frac{1}{l^2} \exp(-pl^2) \cos(pbl)\, dl.$$

Using the results given in Appendix B (Proposition B.1 and also the subsequent remarks) we obtain the expansion of $r(\Delta, 0)$. □

Appendix B

Appendix B contains the proofs omitted in Chapter 6.

B.1

The following series expansions are needed in Subsection 6.3.2.

Proposition B.1 *Adopt the following notation:*

$$
\begin{aligned}
I_0 &= \int_1^\infty \exp(-pl^2)\cos(pbl)\,dl \\
I_1 &= \int_1^\infty \frac{1}{l}\exp(-pl^2)\sin(pbl)\,dl \\
I_2 &= \int_1^\infty \frac{1}{l^2}\exp(-pl^2)\cos(pbl)\,dl \\
\dots & \quad \dots \\
\dots & \quad \dots \\
I_{2n} &= \int_1^\infty \frac{1}{l^{2n}}\exp(-pl^2)\cos(pbl)\,dl \quad \text{and} \\
I_{2n+1} &= \int_1^\infty \frac{1}{l^{2n+1}}\exp(-pl^2)\sin(pbl)\,dl,
\end{aligned}
$$

where b is some fixed real constant, p is positive and $n \geq 1$. These integrals can be written as series expansions, in the following way:

$$
\begin{aligned}
I_0 &= \sum_{j=0}^{\infty} \frac{(-1)^j (pb)^{2j}}{(2j)!} K_j \\
I_1 &= \sum_{j=0}^{\infty} \frac{(-1)^j (pb)^{2j+1}}{(2j+1)!} K_j \\
I_2 &= \exp(-p)\cos(pb) - 2pI_0 - pbI_1
\end{aligned}
$$

$$
\begin{aligned}
\cdots \quad & \quad \cdots \\
\cdots \quad & \quad \cdots \\
I_{2n} &= \frac{1}{2n-1}[\exp(-p)\cos(pb) - 2pI_{2n-2} - pbI_{2n-1}] \\
I_{2n+1} &= \frac{1}{2n}[\exp(-p)\sin(pb) - 2pI_{2n-1} + pbI_{2n}].
\end{aligned}
$$

The numbers $\{K_j : j \geq 0\}$ *are given by*

$$
\begin{aligned}
K_0 &= \int_1^\infty \exp(-pl^2)dl \\
K_j &= \frac{1}{2p}[\exp(-p) + (2j-1)K_{j-1}].
\end{aligned}
$$

PROOF. Using the Taylor expansion of the cosine we get

$$I_0 = \sum_{j=0}^{\infty} \frac{(-1)^j (pb)^{2j}}{(2j)!} \int_1^\infty l^{2j} \exp(-pl^2)\, dl.$$

We abbreviate the last integral with K_j. Partial integration shows that, for $j \geq 1$,

$$
\begin{aligned}
K_j &= l^{2j-1}\frac{\exp(-pl^2)}{-2p}\Bigg|_1^\infty + \int_1^\infty \frac{\exp(-pl^2)}{2p}(2j-1)l^{2j-2}\, dl \\
&= \frac{1}{2p}[\exp(-p) + (2j-1)K_{j-1}].
\end{aligned}
$$

This leaves us with the calculation of K_0. K_0 equals $\sqrt{\pi/4p}\,\mathrm{erfc}(\sqrt{p})$, where $\mathrm{erfc}(x) = 1 - \mathrm{erf}(x)$ and 'erf' denotes the error function. Abramowitz and Stegun [1, p. 297] give the following series expansion for the error function:

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \sum_{j=0}^{\infty} \frac{(-1)^j x^{2j+1}}{j!(2j+1)}.$$

It seems dangerous to use an approximating series in the first term (and all subsequent terms) of a second approximating series. Let us see how a small error in K_0 propagates itself. Applying repeatedly the recursion formula yields that

$$K_j = \exp(-p)\sum_{i=1}^{j} \frac{1}{(2p)^i} \frac{(2j-1)!!}{(2j-2i+1)!!} + (2j-1)!!\, K_0,$$

where $(2j-1)!! \equiv 1 \cdot 3 \cdot 5 \cdots (2j-1)$. Clearly, the influence of an error in K_0 decays as j tends to infinity.

For very small p the factor K_j explodes as j increases. To avoid this it is useful to write the expansion in a slightly different way:

$$I_0 = \sum_{j=0}^{\infty} \frac{(-1)^j p^j b^{2j}}{(2j)!} (p^j K_j),$$

where $p^j K_j$ is determined by the recursion

$$p^j K_j = \frac{1}{2}[p^{j-1} \exp(-p) + (2j-1) p^{j-1} K_{j-1}].$$

The series expansion of I_1 is obtained in exactly the same manner.

The rest of the proof consists in partial integration. The formula that reduces I_{2n} $(n \geq 1)$ to integrals with lower index is obtained as follows

$$\begin{aligned}
&\int_1^{\infty} \frac{1}{l^{2n}} \exp(-pl^2) \cos(pbl)\, dl \\
&\quad = -\frac{1}{2n-1} \frac{1}{l^{2n-1}} \exp(-pl^2) \cos(pbl)\Big|_1^{\infty} + \\
&\quad \int_1^{\infty} \frac{1}{2n-1} \frac{1}{l^{2n-1}} [-2pl \exp(-pl^2) \cos(pbl) - pb \exp(-pl^2) \sin(pbl)]\, dl \\
&\quad = \frac{1}{2n-1} \exp(-p) \cos(pb) - \frac{2p}{2n-1} I_{2n-2} - \frac{pb}{2n-1} I_{2n-1};
\end{aligned}$$

I_{2n+1} is dealt with in exactly the same way. □

Let us add some remarks to this proposition. First we note that for I_0 an alternative expansion is available.

$$\begin{aligned}
I_0 \;=\; & \sqrt{\frac{\pi}{4p}} \exp(-pb^2/4) \qquad \text{(B.1)} \\
& [1 - \frac{2}{\sqrt{\pi}} \sum_{j=0}^{\infty} \frac{(-1)^j (\sqrt{p})^{2j+1} |\theta|^{2j+1}}{j!(2j+1)} \cos((2j+1)\arg(\theta))],
\end{aligned}$$

where θ is shorthand for $(1 - \frac{1}{2}ib)$. This formula is proved by writing the cosine as an exponential having complex argument.

$$\begin{aligned}
I_0 &= \Re \int_1^{\infty} \exp(-pl^2 + ipbl)\, dl \\
&= \Re \int_1^{\infty} \exp(-p(l - \frac{1}{2}ib)^2 - \frac{1}{4}pb^2)\, dl
\end{aligned}$$

$$
\begin{aligned}
&= \exp(-\frac{1}{4}pb^2)\ \Re \int_{1-\frac{1}{2}ib}^{\infty} \exp(-pz^2)\, dz \\
&= \exp(-\frac{1}{4}pb^2)\ \Re \frac{1}{2}\sqrt{\frac{\pi}{p}} \mathrm{erfc}(\sqrt{p}\,(1-\frac{1}{2}ib)),
\end{aligned}
$$

where in the last integral the path of integration extends from $z = \sqrt{p}(1 - \frac{1}{2}ib)$ to infinity along the line $\Im z = \frac{1}{2}ib$. Now we can cite again the series expansion given in Abramowitz and Stegun (this time for a complex argument):

$$
\mathrm{erfc}(z) = 1 - \frac{2}{\sqrt{\pi}} \sum_{j=0}^{\infty} \frac{(-1)^j z^{2j+1}}{j!(2j+1)},
$$

so that, with $z = \sqrt{p}(1 - \frac{1}{2}ib)$, we obtain (B.1). A second remark concerns I_2. If we need not calculate I_0 and I_1 for their own sake, it seems natural to evaluate I_2 directly, using the series expansion

$$
I_2 = \sum_{j=0}^{\infty} \frac{(-1)^j (pb)^{2j}}{(2j)!} K_{j-1}.
$$

K_{-1} is given by $K_{-1} = \exp(-p) - 2pK_0$ and the subsequent factors $p^j K_j$ are again obtained by the recursion formula

$$
p^j K_j = \frac{1}{2}[p^{j-1}\exp(-p) + (2j-1)p^{j-1}K_{j-1}].
$$

Finally we note that the series I_0, I_1 and I_2 convere (too) slowly, when b is large, even if we asssume that the product pb remains constant. In this case, however, Riemann-sum like approximations of the integrals can be used. For example

$$
\begin{aligned}
I_0 &\approx \sum_{i=0}^{\infty} \exp(-p[1 + (i+0.5)\frac{\epsilon}{pb}]^2) \int_{1+i\epsilon/(pb)}^{1+(i+1)\epsilon/(pb)} \cos(pbl)dl \\
&= \frac{2}{pb} \sum_{i=0}^{\infty} \exp(-p[1 + (i+0.5)\frac{\epsilon}{pb}]^2) \cos(pb + \epsilon(i+0.5)) \sin(\epsilon/2),
\end{aligned}
$$

where $\epsilon \ll pb$.

Proposition B.2 *Define*

$$
\begin{aligned}
J_1(p) &= \int_1^{\infty} \frac{1}{l} \sin(pl)\, dl \\
J_{2n}(p) &= \int_1^{\infty} \frac{1}{l^{2n}} \cos(pl)\, dl \\
J_{2n+1}(p) &= \int_1^{\infty} \frac{1}{l^{2n+1}} \sin(pl)\, dl,
\end{aligned}
$$

where $n \geq 1$ and p is real (and not equal to zero in the case of J_1). These integrals can be expressed as series expansions in the following way

$$\begin{aligned} J_1(p) &= \frac{\pi}{2} - \sum_{j=0}^{\infty} \frac{(-1)^j p^{2j+1}}{(2j+1)(2j+1)!} \\ J_{2n}(p) &= \frac{1}{2n-1}[\cos(p) - pJ_{2n-1}] \\ J_{2n+1}(p) &= \frac{1}{2n}[\sin(p) + pJ_{2n}]. \end{aligned}$$

PROOF. The integral J_{2n} is reduced to an integral with lower index by partial integration

$$\begin{aligned} &\int_1^{\infty} \frac{1}{l^{2n}} \cos(pl)\, dl \\ &= \frac{-1}{2n-1} \frac{-1}{l^{2n-1}} \cos(pl)\Big|_1^{\infty} - p\frac{1}{2n-1} \int_1^{\infty} \frac{1}{l^{2n-1}} \sin(pl)\, dl \\ &= \frac{1}{2n-1} \cos(p) - p\frac{1}{2n-1} J_{2n-1}. \end{aligned}$$

The integral J_{2n+1} is reduced in the same way. Thus, all these integrals can eventually be written in terms of J_1. But this integral is nothing else than an instance of the sine integral

$$J_1(p) = \mathrm{si}(p) = \frac{\pi}{2} - \mathrm{Si}(p)$$

and Abramowitz and Stegun [1, p. 232] give the following series expansion for $\mathrm{Si}(z)$:

$$\mathrm{Si}(z) = \sum_{j=0}^{\infty} \frac{(-1)^j z^{2j+1}}{(2j+1)(2j+1)!}.$$

□

B.2

In this section we discuss some results concerning the discretization of the stationary, centred, Gaussian process, $R(t, x)$, which is determined by the covariance function

$$r(\Delta, z) = \frac{\sigma^2 S}{4\pi^2 K} \int_1^{\infty} \frac{1}{l^2} \exp\left(-\frac{4\pi^2 l^2}{S^2} K|\Delta|\right) \cos\left(2\pi l \frac{z - c_0\Delta}{S}\right) dl.$$

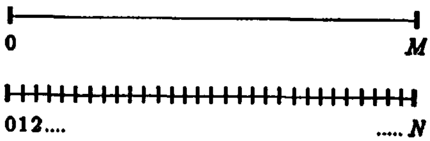


Figure B.1: The discretization of the interval $[0, M]$.

The discretization is performed in three steps. First, we return to the process $R_m(t,x)$. Secondly, the process $R_m(t,x)$ is discretized with respect to the space coordinate, approximating it by the N-dimensional process $R_m^N(t)$, that is the solution of the N-dimensional stochastic differential equation

$$\begin{aligned} dR_m^N(t) &= K\Delta R_m^N(t)\,dt - c_0\nabla R_m^N(t)\,dt + \sqrt{\frac{N}{M}}\,\sigma\,dw(t), \\ R_m^N(0) &\sim \mathrm{N}(0,V), \end{aligned} \tag{B.2}$$

where Δ and ∇ are the difference operators defined as

$$\begin{aligned} \Delta f(k) &= \left(\frac{N}{M}\right)^2 [f(k+1) - 2f(k) + f(k-1)] \\ \nabla f(k) &= \frac{N}{2M}[f(k+1) - f(k-1)] \end{aligned}$$

for any vector-valued function f. Recall that M is the total length of the space interval on which the process $R_m(t,x)$ is defined. Figure B.1 shows the discretization of the interval $[0, M]$. The covariance matrix V is given by

$$V_{kl} = \sum_{i=m}^{N/2-1} \frac{\sigma^2}{(M\lambda_i K)} \cos(2\pi i \frac{k-l}{N}) + \frac{\sigma^2}{(2M\lambda_{N/2}K)}(-1)^{k-l},$$

where $\lambda_i = 4N^2\sin^2(\pi i/N)/M^2$ are the eigenvalues of Δ, and $w(t)$ is an N-dimensional Brownian motion, having covariance function

$$Ew_t(k)w_s(l) = (t \wedge s)\left(\frac{2}{N}\sum_{i=m}^{N/2-1} \cos(2\pi i(k-l)/N) + \frac{1}{N}(-1)^{k-l}\right).$$

And, thirdly, this N-dimensional stochastic differential equation is discretized with respect to the time coordinate, using the standard Euler-scheme. We obtain the following:

$$\begin{aligned} R_d(i+1,j) &= K\Delta R_d(i,j)\,\delta - c_0\nabla R_d(i,j)\,\delta + \sigma\sqrt{\frac{N}{M}}\sqrt{\delta}\,w(i,j), \\ &R_d(0) \sim \mathrm{N}(0,V) \end{aligned} \tag{B.3}$$

where $\delta = T/N_t$, N_t being the number of time steps, and $w(i,j)$ is a small Gaussian error term with zero mean and covariance

$$Ew(h,k)w(i,l) = \delta_{h,i}\left(\delta_{k,l} - \frac{1}{N} - \frac{2}{N}\sum_{i=1}^{m-1}\cos(2\pi i\frac{(k-l)}{N})\right).$$

Here $\delta_{h,i}$ is the Kronecker symbol. In (B.3) d means 'discretized'.

In order to speak of convergence of R_d to $R_m^N(t)$ we embed R_d into the space $C([0,T]\times[0,M])$ by identifying $R_d(i,j)$ and $R_d(t,x)$ for $t = iT/N_t$ and $x = jM/N$ and by subsequent linear interpolation. Note that we use the symbol 'R_d' for two different, though very related, processses. Following the three discretization steps in reversed order, we claim that

$$\lim_{m\to\infty}\lim_{N\to\infty}\lim_{N_t\to\infty} R_d \stackrel{d}{=} R.$$

The convergence in distribution of R_d to R_m^N follows from Theorem 1.10 in [26, p. 32]. Convergence of R_m to R was already proven in Appendix A.

Thus, the remaining step, needed to prove the claim, is proving convergence in distribution of R_m^N to R_m.

Theorem B.1 *Let R_m^N denote the solution of the equation (B.2) and R_m the stochastic process whose representation is given by (4.15). R_m can be viewed as a $C([0,T]\times[0,M])$-valued process, while R_m^N can be embedded into the same space by identifying $R_m^N(t,k)$ and $R_m^N(t,x)$ for $x = kM/N$ and by subsequent linear interpolation:*

$$\begin{aligned} R_m^N(t,x) &= (\frac{Nx}{M} - k + 1)\,R_m^N(t,k) + (k - \frac{Nx}{M})\,R_m^N(t,k-1),\\ & \qquad x \in [\frac{(k-1)M}{N}, \frac{kM}{N}], \qquad 1 \le k \le N. \end{aligned}$$

Then R_m^N converges in distribution to R_m.

Proof. Throughout the proof m is fixed. We will drop this subscript. Most of the time N is fixed too and, by assumption, it is always even. Therefore, we will drop this superscript as long as no confusion can arise.

Going again through many calculations of Appendix A, but now for the discrete-space equation, we obtain that

$$R(t,k) = \sum_{i=m}^{N/2-1} a_i^s(t)\phi_i(k - c_i t) + a_i^c(t)\psi_i(k - c_i t) + a_{N/2}^c(t)\sqrt{\frac{1}{M}}(-1)^k,$$

where $\phi_i(k) = \sqrt{2/M}\sin(2\pi ik/N)$ and $\psi_i(k) = \sqrt{2/M}\cos(2\pi ik/N)$. Further, $\{a_i^s(t), a_i^c(t) : i \geq 1\}$ is a family of mutually independent Ornstein-Uhlenbeck processes, i.e. each $a_i(t)$ or $b_i(t)$ is a centred stationary Gaussian process having covariance function

$$Ea_i^s(t)a_i^s(s) = Ea_i^c(t)a_i^c(s) = \frac{\sigma^2}{2\lambda_i K}\exp(-\lambda_i K|\Delta|),$$

where $\lambda_i = 4N^2\sin^2(\pi i/N)/M^2$ and $\Delta = t - s$. Finally,

$$c_i = \frac{N}{M}\frac{\sin(2\pi i/N)}{2\pi N}c_0.$$

Using this we see that the covariance function, $r(\Delta, z)$, reads

$$\begin{aligned} r(\Delta, z) &= \sum_{i=m}^{N/2-1}\frac{\sigma^2}{M\lambda_i K}\exp(-\lambda_i K|\Delta|)\cos(2\pi i\frac{z - c_i\Delta}{N}) \\ &+ \frac{\sigma^2 M}{8N^2 K}\exp(-4N^2 K|\Delta|/M^2)(-1)^z, \end{aligned}$$

where $\Delta = t - s$, $z = k - l$ and (t, k) and (s, l) are two points of the space $\mathbf{R} \times \{1, \ldots, N\}$.

We note in passing that the analogue of the semigroup V introduced in Subsection 4.3.2 is

$$\begin{aligned} V_t f(k) &= \frac{M}{N}\sum_{l=1}^{N} q(t,k,l)f(l) \qquad \text{with} \\ q(t,k,l) &= \frac{1}{M} + \sum_{i=1}^{N/2-1}\Big(\phi(k - c_i t)\phi(l) + \psi(k - c_i t)\psi(l)\Big)e^{-\lambda_i K t} \\ &+ \frac{1}{M}(-1)^{k-l}. \end{aligned}$$

The rest of the proof essentially consists in the same argument as the one put forward in the proof of Theorem A.2. Let R_m^N denote the interpolated process derived from the discrete one. It is easy to see that the covariance function of the interpolated process converges pointwise to the covariance function of the continuous process. As both processes are Gaussian, this implies convergence of all finite dimensional distributions.

Furthermore, using estimates obtained before, it can be shown in a straightforward manner that (with C_1 and C_2 independent of N)

$$E\,|R_m^N(t,x) - R_m^N(s,y)|^{2p} \leq C_1\,|t-s|^{p/2} + C_2\,|x-y|^p,$$

which, as in the proof of Theorem A.2, implies tightness. □

Appendix C

In this appendix we state two extreme value theorems. First we introduce some notation. In the following $X(t,x)$ will be a two parameter, real-valued, zero-mean and stationary Gaussian process, that is completely determined by its covariance function, $r(\Delta, z)$, where $\Delta = t - s$ and $z = x - y$; (t,x) and (s,y) are two points in the parameter plane. We assume that the covariance function has an expansion around $(0,0)$ of the form

$$r(\Delta, z) = C_0(1 - C_1|\Delta|^\alpha - C_2|z|^\beta + o(|\Delta|^\alpha) + o(|z|^\beta)), \qquad \text{(C.1)}$$

where $0 < \alpha, \beta \leq 2$. Further, we define

$$M(T,L) \equiv \sup_{0 \leq t \leq T,\ 0 \leq x \leq L} X(t,x).$$

As usual ϕ denotes the density of the standard normal distribution.

Theorem C.1 *Let $X(t,x)$ be a two parameter, real-valued and stationary Gaussian process with zero mean, whose covariance function satisfies (C.1), as (Δ, z) tends to $(0,0)$. If L and T are such that, for all $\epsilon_1, \epsilon_2 > 0$,*

$$\begin{aligned} \sup_{\epsilon_1 \leq \Delta \leq T, 0 \leq z \leq L} r(\Delta, z) &< 1 \\ \sup_{0 \leq \Delta \leq T, \epsilon_2 \leq z \leq L} r(\Delta, z) &< 1, \end{aligned}$$

then

$$\lim_{u \to \infty} \frac{1}{u^{2/\alpha + 2/\beta - 1}\phi(u)} P\{M(T,L) > uC_0\} = TLC_1^{1/\alpha}C_2^{1/\beta}H_{\alpha,\beta},$$

where the constant $H_{\alpha,\beta}$ depends only on α and β.

Theorem C.2 *Let $X(t,x)$ be a two parameter, real-valued and stationary Gaussian process with zero mean, whose covariance function satisfies (C.1), as (Δ, z) tends to $(0,0)$, and*

$$r(\Delta,0)\log(\Delta) \to 0 \qquad \text{as } \Delta \to \infty.$$

Furthermore, $\sup_{\epsilon \le z \le L} r(\Delta, z) \le \eta < 1$, *uniformly in* Δ. *It holds that, if* $T \to \infty$ *and* $u \to \infty$, *such that* $\lim_{T\to\infty} T\mu = \tau > 0$, *where* $\mu(u) = u^{2/\alpha+2/\beta-1}\phi(u)C_1^{1/\alpha}C_2^{1/\beta}H_{\alpha,\beta}$,

$$P\{M(T,L) \le uC_0\} \to \exp(-\tau L).$$

$H_{\alpha,\beta}$ *is a constant depending only on* α *and* β.

The proofs are almost copies of the proofs of Theorems 12.2.9 and 12.3.4 of the book of Leadbetter, Lindgren and Rootzén [18].

It is clear that the stochastic process $R(t,x)$ satisfies the conditions of the theorems. From Proposition A.5 we have that $C_0 = A$, $C_1 = \sqrt{\pi a}$, $C_3 = \pi^2/S$, $\alpha = 1/2$ and $\beta = 1$, so that Theorem C.1 yields

$$\lim_{u\to\infty} \frac{1}{u^5\phi(u)} P\{M(T,L) > uA\} = \pi^3 TL\frac{a}{S}H_{\frac{1}{2},1}.$$

This means that for large values of u

$$P\{M(T,L) > uA\} \approx u^5\phi(u)\pi^3 TL\frac{a}{S}H_{\frac{1}{2},1}.$$

We conclude that the probability of extreme values, relative to A, is approximately proportional to a/S.

Bibliography

[1] Milton Abramowitz and Irene S. Stegun (Eds.), *Handbook of Mathematical Functions*, with Formulas, Graphs and Mathematical Tables, Dover Publications, New York, 1965.

[2] E. Alberti and G. Belli, Contributions to the Boltzmann-like approach for traffic flow — a model for concentration dependent driving programs, *Transpn. Res.*, vol. 12 (1978), pp. 33–42.

[3] Patrick Billingsley, *Probability and Measure*, second edition, Wiley, New York, 1986.

[4] Leo Breiman and Robert Lawrence, A simple stochastic model for freeway traffic, *Transpn. Res.*, vol. 11 (1977). pp. 177–182.

[5] Paul Bratley, Bennett L. Fox and Linus E. Schrage, *A Guide to simulation*, Springer, New York, 1983.

[6] Ruth F. Curtain, Stochastic Evolution Equations with General White Noise Disturbance, *J. of Math. An. and Appl.*, vol. 60 (1977). pp. 570–595.

[7] E. B. Davies, *One-parameter semigroups*, Academic Press, London, 1980.

[8] D. A. Dawson, Stochastic evolution equations and related measure processes, *Journal of Multivariate Analysis*, vol. 5 (1975), pp. 1–52.

[9] Tadahisa Funaki, Random motion of strings and related stochastic evolution equations, *Nagoya Math. J.*, vol. 89 (1983), pp. 129–193.

[10] J. A. Goldstein, *Semigroups of Linear Operators and Applications*, Oxford University Press, New York, 1985.

[11] P. Groeneboom, *Onderzoek volgtijden op de 2×2-strooks autosnelweg*, Report MS-N8401, Centre for Mathematics and Computer Science, Amsterdam, 1984.

[12] Robert Herman and R. B. Potts, Single-lane traffic theory and experiment, in: *Theory of Traffic Flow* (Ed. Robert Herman), pp. 120–146, Elsevier, Amsterdam, 1961.

[13] T. Hida, *Brownian Motion*, Springer-Verlag, New York, 1980.

[14] N. Ikeda and S. Watanabe, *Stochastic Differential Equations and Diffusion Processes*, North-Holland/Kodansha, Tokyo, 1981.

[15] K. Itô, *Foundations of Stochastic Differential Equations in Infinite Dimensional Spaces*, SIAM, Philadelphia, 1984.

[16] Robert I. Jennrich and Mary L. Ralston, Fitting nonlinear models to data, *Ann. Rev. Biophys. Bioeng*, vol. 8 (1979), pp. 195–238.

[17] Erwin Kreyszig, *Introductory Functional Analysis with Applications*, Wiley, New York, 1978.

[18] M. R. Leadbetter, Georg Lindgren and Holger Rootzén, EXTREMES *and Related Properties of Random Sequences and Processes*, Springer-Verlag, New-York, 1983.

[19] M. J. Lighthill and G. B. Whitham, On kinematic waves II: A theory of traffic flow on long crowded roads, *Proc. Roy. Soc. London*, Series A, vol. 229 (1955), pp. 317–345.

[20] Adolf D. May, Models for freeway corridor analysis, in: *TRB Special Report 194*, Transpn. Res. Board, Washington, D.C., approx. 1982.

[21] Panos G. Michalopoulos, Dimitrios E. Beskos and Yasuji Yamauchi, Multilane traffic flow dynamics: some macroscopic considerations, *Transpn. Res.*, vol. 18B (1984), pp. 377–395.

[22] P. K. Munjal and J. Pahl, An analysis of the Boltzmann-type statistical models for multi-lane traffic flow, *Transpn. Res.*, vol. 3 (1969), pp. 151–163.

[23] Michel Metivier and J. Pellaumail, *Stochastic integration*, Academic Press, New York, 1980.

[24] E. W. Montroll, Acceleration noise and clustering tendency of vehicular traffic, in: *Theory of Traffic Flow* (Ed. Robert Herman), pp. 147–157, Elsevier, Amsterdam, 1961.

[25] G. F. Newell, A theory of traffic flow in tunnels, in: *Theory of Traffic Flow* (Ed. Robert Herman), pp. 193–206, Elsevier, Amsterdam, 1961.

[26] E. Pardoux and D. Talay, Discretization and Simulation of Stochastic Differential Equations, *Acta Applicandae Mathematicae*, vol. 3 (1985), pp. 23–47.

[27] S. L. Paveri-Fontana, On Boltzmann-like treatments for traffic flow: a critical review of the basic model and an alternative proposal for dilute traffic analysis, *Transpn. Res.*, vol. 9 (1975), pp. 225–235.

[28] Harold J. Payne, Models of freeway traffic and control, *Mathematical Models of Public Systems, Simulation Council Proceedings*, vol. 1 (1971), pp. 51–61.

[29] I. Prigogine, A Boltzmann-like approach to the statistical theory of traffic flow, in: *Theory of Traffic Flow* (Ed. Robert Herman), pp. 158–164, Elsevier, Amsterdam, 1961.

[30] Ilya Prigogine and Robert Herman, *Kinetic Theory of Vehicular Traffic*, American Elsevier, New York, 1971.

[31] C. Radhakrishna Rao, *Linear Statistical Inference and Its Applications* (second edition), Wiley, New York, 1973.

[32] Ajay K. Rathi and Zoltan A. Nemeth, An application of variance reduction techniques in freeway simulation, *Transpn. Res.*, vol. 19B (1985), pp. 209–216.

[33] Murray Rosenblatt, *Stationary Sequences and Random Fields*, Birkhäuser, Boston, 1985.

[34] Paul Ross, Traffic dynamics, *Transpn. Res.*, vol. 22B (1988), pp. 421-435.

[35] G. A. F. Seber and C. J. Wild, *Nonlinear Regression*, Wiley, New York, 1989.

[36] Stef Smulders, *Control of Freeway Traffic Flow*, Thesis, Twente University of Technology, Enschede, 1988.

[37] J. B. Walsh, An introduction to stochastic partial differential equations, in: *Lecture notes in mathematics*, vol. 1180, Springer-Verlag, Berlin Heidelberg, 1986.

[38] G. B. Whitham, *Linear and Nonlinear waves*, Wiley, New York, 1974.

[39] M. Yor, Existence et unicité de diffusions à valeurs dans un espace de Hilbert, *Annales de l'Institut Henri Poincaré*, section B, vol. 10 (1974), pp. 55–88.

Samenvatting

De capaciteit van een snelweg is geen vaststaand gegeven, alleen bepaald door bijvoorbeeld het aantal rijstroken en de geometrische vormgeving van de weg. Zij hangt mede af van veranderlijke zaken als de weersgesteldheid en het tijdstip van de dag. Verder is ook de mate van homogeniteit van de verkeersstroom een belangrijke (veranderlijke) factor. Een gelijkmatige, homogene verkeersstroom zal minder congestiegevoelig zijn dan een wilde, inhomogene verkeersstroom. Deze dissertatie beschrijft een stochastisch model voor een verkeersstroom op een snelweg, dat de mogelijkheid biedt om de homogeniteit van de verkeersstroom met behulp van een klein aantal parameters te karakteriseren.

Het model is gebaseerd op een aantal aannames, waarvan er één al wordt aangeduid door de gebruikte terminologie. Immers de term verkeers*stroom* wordt slechts dan terecht gebezigd, als we te maken hebben met een aaneengesloten stroom van voertuigen, die elk kunnen worden getypeerd als 'volger'. Elk voertuig rijdt binnen volgafstand van zijn voorganger. De verkeersstroom heeft een hoge 'dichtheid' (uitgedrukt in aantal voertuigen per km per rijstrook). Verder is om het model hanteerbaar te houden besloten om alleen naar stationaire verkeersstromen te kijken. Deze beperking is niet zwaarwegend, omdat het bij beschouwingen over de (in)homogeniteit van verkeersstromen meestal over min of meer stationaire verkeersstromen gaat. De poging om een verkeersstroom te kenschetsen als homogeen of inhomogeen impliceert al dat de stroom gedurende langere tijd een gelijkblijvend karakter heeft, d.w.z stationair is.

De eerste aanname leidt tot het beschrijven van de verkeersstroom op ongeveer dezelfde manier als ook één-dimensionale vloeistofstromen worden beschreven. Hierbij speelt het 'behoud van voertuigen' een belangrijke rol. Voorts kunnen we de resulterende stelsels van vergelijkingen op grond van de tweede aanname lineariseren. Aan de aldus verkregen lineaire vergelijking (voor de dichtheid van de verkeersstroom) wordt een passende ruisterm toegevoegd. Het eindresultaat is een zogenaamde stochastische warmtevergelijking, die de fluctuaties van de

dichtheid rond een zeker gemiddelde beschrijft. Een tamelijk groot deel van het proefschrift gaat over het op correcte wijze opstellen en analyseren van deze vergelijking. De aard van de fluctuaties, die bepalend is voor de mate van homogeniteit van het verkeer, kan worden samengevat in een drietal parameters:

- de amplitude van de fluctuaties,
- de afstand waarover de fluctuaties op één bepaald tijdstip ongeveer uitmiddelen,
- de snelheid waarmee de fluctuaties uitdempen in de ogen van een waarnemer die met de stroom meereist.

Het ligt voor de hand een nauwe samenhang tussen de congestiegevoeligheid en het optreden van extreme waarden van de dichtheid te veronderstellen. Het voorkomen van extreme waarden wordt bepaald door een viertal parameters. In de eerste plaats de gemiddelde dichtheid en in de tweede plaats de drie parameters die de aard van de fluctuaties vastleggen.

De validatie van het model is geschied met behulp van gegevens verstrekt door Rijkswaterstaat. De uitkomsten hiervan zijn bemoedigend, al bleek één ad hoc aanpassing noodzakelijk. Voor directe toepassing van het model is het nodig dat de parameters worden gecalibreerd, d.w.z. dat van de parameters de kritieke waarden worden bepaald. Het vermoeden bestaat, op grond van de bestudeerde gegevens, dat hierbij vooral de waarde van de derde parameter weergeeft of de verkeersstroom homogeen of inhomogeen is.

Curriculum vitae

Ello Aart Gijsbert Weits is geboren op 6 januari 1959 in Emmen. In de jaren 1971–1977 doorliep hij het gymnasium in Stadskanaal.

Vervolgens verhuisde hij naar Groningen, waar hij aan een studie natuurkunde begon. Tevens werd hij lid van de VCS 'Hendrik de Cock'. Een groeiende interesse in levensbeschouwelijke onderwerpen leidde er in 1979 toe dat hij theologie als tweede studie koos.

In 1980 behaalde hij het kandidaatsexamen technische natuurkunde, in 1985 het doctoraal examen natuurkunde (gemengde afstudeerrichting; specialisaties: vaste stof fysica en statistische mechanica; bijvakken: wiskunde en wetenschapsfilosofie).

Terwijl hij zich bezon op de mogelijkheden voor een pas afgestudeerd natuurkundige, rondde hij de kandidaatsstudie theologie af en behaalde onderwijsbevoegdheden voor wis- en natuurkunde.

Geïnspireerd door het voorbeeld van enkele goede vrienden maakte hij in 1986 de overstap naar de toegepaste wiskunde (met name de mathematische statistiek). Hij begon in Amsterdam onder leiding van prof. dr. P. Groeneboom aan een STW-project getiteld "Statistische analyse van verkeersstromen".

Onderwijl verloor hij zijn interesse in de theologie en de filosofie niet. In Leiden vatte hij een doctoraal studie theologie (nieuwe stijl) op. De filosofie van Emmanuel Levinas stond hierbij centraal. In 1989 werd het examen afgelegd.

Gestaag werkte hij verder aan het ontwerp en de statistische analyse van een stochastisch model voor een verkeerstroom op een snelweg. Toen eenmaal de balans tussen de complexiteit van het model en de hanteerbaarheid ervan gevonden was, kwam er duidelijk vaart in het onderzoek.

Het grootste deel van de dissertatie werd geschreven in Delft, waar hij vanaf de zomer van 1988 als NWO-medewerker is gedetacheerd. Overigens was hij er al in begin 1987 met zijn levensgezellin gaan wonen. De verandering van werkplek bespaarde zodoende veel reistijd.

Over de activiteiten na de promotie bestaat op dit moment nog geen zekerheid.