

# Proceedings of the forty-fifth European Study Group with Industry

Leiden, The Netherlands, 17 - 21 February 2003

Editors:  
Derk Pik  
Vivi Rottschäfer

Study groups with industry are meetings where people from industry join forces with mathematicians to tackle industrial problems. The first such meeting was held in the sixties at the University of Oxford. Nowadays study groups with industry are organised in many countries, e.g. Australia, the USA and several European countries.

Since 1998 study groups with industry are also organised in The Netherlands. In 1998 a study group was hosted by the University of Leiden, in 1999 by the University of Eindhoven, in 2000 by Twente University, in 2002 by the CWI and the University of Amsterdam, and in 2003 the study group was hosted by the University of Leiden again. The next study group will take place in Delft, in March 2004. For up-to-date information please follow the links [www.wiskgenoot.nl/swi](http://www.wiskgenoot.nl/swi) or [www.stw.nl](http://www.stw.nl). The 45th European Study Group with Industry was organised by the University of Leiden at the Lorentz Center. The programme was supported by the section *Industriële en Toegepaste Wiskunde* (ITW) of the *Wiskundig Genootschap*, with financial support from the technology programme *Wiskunde Toegepast* of the technology foundation STW and the EW (Exact Sciences) programme of The Netherlands Organisation for Scientific Research (NWO), and with additional support from the Lorentz Center in Leiden and the European Consortium for Mathematics in Industry (ECMI).

Organising committee:

N.C. Ovenden

D.R. Pik

V. Rottschäfer

2000 Mathematics Subject Classification:

00B25, 35QXX, 62H35, 68Q15, 92D25, 90C27, 90C90, 62P30, 76A05.

ISBN 90-9017846-5

Copyright ©2004

Printed in the Netherlands

## Preface

The forty-fifth study group with industry is one in a series of very successful workshops. This series started in the sixties at the University of Oxford. Since 1998, some of these study groups have been held in the Netherlands, and this one is the fifth in a row. The purpose of these groups is to join the effort of mathematicians to solve real-life mathematical problems faced in industry.

This year, the study group was held at the Lorentz Center in Leiden from February 17 until February 21. Five problems were studied: the behavior of a droplet of a polymer solution, the spreading of Japanese oysters in Dutch waters, the measure of agreement between a tool and marks left at a crime scene, the optimal way of hanging bells and wires in a carillon, and forecasting the performance and flexibility of flight schedules.

Each of these problems was studied by one group, varying in size between 7 and 14 persons. At the Lorentz Center in Leiden they were provided with any kind of facility they needed, such as office space, computers, library access, and, a very friendly environment.

All participants were very enthusiastic. This could be seen at many occasions: of course at the study group diner, but also at the visit to the Museum Boerhaave in Leiden, the visit of the carillon group to the Zuiderkerk in Amsterdam to observe a carillon, the Friday morning carillon concert in Leiden and the proud participants of the crime scene marks group (and many others) having a special made print of marks on their own shoes.

The results were presented on the last Friday of the workshop to the other participants, to the participating companies and to other interested persons. These were very interesting presentations which clearly explained the problems faced from both the societal and mathematical point of view, and the solutions found during the preceding week. All possible techniques were used to convince the public of the results: among others full simulations of various sizes of falling droplets and the effects on the ‘tail’ of such a drop.

This study group couldn’t have been organized without the help of the sponsors, first the participating companies providing the problems who have been assisting the mathematicians during the week of the study group, the Lorentz Center in Leiden providing all means of facilities, and, as always, NWO-EW and the Technology Foundation STW with their special program ‘Wiskunde Toegepast’. The main aim of Wiskunde Toegepast is to finance applied mathematical research, but another important aim is to stimulate contact between mathematicians and contact between mathematicians and society. The study group is an excellent way of doing that, as

can be seen from the various newspaper clippings and from an interview for the Dutch radio with the carillon group.

Last but certainly not least, this study group would not at all have been held without the inspiration and activities of the organizing committee: Nick Ovenden (University of Eindhoven), Derk Pik and Vivi Rottschäfer (both University of Leiden). Together, we all can promote mathematics for a very broad audience.

Rik D.T. Janssen  
Technology Foundation STW  
October 2003

## Contents

Preface	iii
Summaries	1
Participants	5
Chapter 1. Solution Methodologies for generating robust Airline Schedules	9
Chapter 2. Japanese Oysters in Dutch Waters	21
Chapter 3. Non-Newtonian Effects on Ink-Jet Droplet Formation: A Mathematical Model	45
Chapter 4. Catch them ... if you can	57
Chapter 5. Hanging a Carillon in a Broek-system	73



## Summaries

For the 45th European Study Group with Industry five problems were selected. The summaries of these problems are presented here.

### 1. KLM – forecasting the performance and flexibility of a flight schedule

KLM flies to over 150 destinations with 97 aircraft. Four times a year, a new flight schedule is developed aimed principally at maximising the number of seats which can be sold. The schedule design takes into account operational feasibility to some extent, but concentrates on the commercial aspects such as expected demand per destination and the number of transfer connections at Schiphol.

Each day during operation, many adaptations are made to this schedule to minimise delays caused by for example problems with the aircraft or weather. If it is known that an aircraft will arrive at Schiphol Airport with a delay, they try to assign its next flight to another aircraft so that that flight can still leave on time. Usually, a couple of other adaptations are needed to have all flights fit again. Some schedules prove more flexible and robust than others in coping with delays.

Question:

Can a simple fast method be found to test the performance and flexibility of a given flight schedule when incorporating the adaptations that are made to the schedule during operation? So far they have been doing simulations with real data and the KLM wants to know whether this can be reduced or avoided.

### 2. RIVO – modelling the expanding Pacific Oyster population in the Eastern Scheldt

Pacific oysters, native to Japan, were introduced into the Eastern Scheldt Estuary in The Netherlands after the severe winter of 1962/63 diminished the stock of European flat oysters. At that time, it was believed that the Pacific Oyster could not breed at such latitudes. However, during the hot summer of 1976, the first settling of larvae was observed on dike foets and jetties after which importation of the Pacific Oyster was immediately halted. In 1982, a second larvae outburst permanently established the wild Pacific oysters in the waters of the Eastern Scheldt and the population has been growing rapidly ever since.

The Pacific oyster represents a serious environmental problem in the Eastern Scheldt due to lack of natural predators and because they compete with other species such as cockles, mussels and cultivated oysters for space and food.

Question:

Can the study group design a mathematical model of the oyster population that predicts the spreading of the oyster population? Can a suggestion be given to how the population growth can be brought under control?

### 3. PHILIPS – the behaviour of a droplet of polymer solution

Light-emitting polymer displays are a new, interesting flat display principle. The active material is a very thin layer of semi-conducting polymer. To make a full colour display red, green and blue polymer solutions must be positioned in pixels of the order of  $66 \times 200$  micron. Philips is currently trying to adapt ink-jet technology to position these small drops of polymer solution effectively.

The polymers involved in this process have a high-molecular weight which causes the droplet formation from the ink-jet to be highly non-Newtonian; this is unlike the behaviour of normal inks.

Question:

Can the study group model the visco-elastic behaviour of the jetted polymer solution from which information about the velocity and droplet formation of the polymer liquid can be obtained?

### 4. Dutch Forensic Institute – probability models for tool marks and shoe prints

When a tool has been used to commit a crime (such as using a screwdriver to open a door during a burglary) the tool leaves certain marks which are unique to that particular tool. The question when handling a case in court is of course whether a tool of a suspect could have left these traces. A similar question can also be asked for shoe prints found at a crime scene. These questions can be answered by comparing the traces to test traces of the tool or shoe of the suspect. The position of certain lines, curves, dents or more distinguishing marks on the tool or shoe help to *match* the trace left at a crime scene. However, so far, a good subjective judgment can't be made.

Question:

Design a probability model that gives the probability that the trace found at a scene of crime was made by the tool or shoe of a certain suspect.

### 5. National Carillon Museum – optimise the hanging of carillon bells and wires in a tower

A carillon consists of around 20 to 45 bells hung in a tower. The bells are played using a keyboard situated below and a wire connects each bell clapper to its relevant key on the keyboard. The oldest and simplest wire-connection system is the 'broek-system' connecting three wires to the 'broekring': one from the clapper, one from the key and one from a fixed point on the wall.

To construct a carillon, using this system, such that all the designing properties are satisfied, and that all the bells sound as they should is incredibly difficult. The carillon-builder attempts to place the bells in a geometrically balanced way, but it



is hard to prevent all the wires from touching or to make every key and bell play equally well.

Question:

Can the bells and wires be positioned in a tower in an optimal way such that designing properties are satisfied?



## Participants

Sergei Anisov  
Universiteit Utrecht  
anisov@math.uu.nl

Jan Bouwe van den Berg  
University of Nottingham  
Jan.Bouwe@nottingham.ac.uk

Dragan Bezanovic  
Technische Universiteit Eindhoven  
d.bezanovic@tue.nl

Fuping Bian  
Tianjin University, Tianjin, China  
fpbian@public.tpt.tj.cn

Gideon Bodden  
Het Molenpad Expertise, Gouda  
gideonbodden@molenpad.com

Chris Budd  
University of Bath  
cjb@maths.bath.ac.uk

Edmund Burke  
University of Nottingham  
ekb@cs.nott.ac.uk

John Mc Darby  
University College London  
makemineaguinness@hotmail.com

Fieke Dekkers  
Universiteit Utrecht  
dekkers@math.uu.nl

Frits Dijkstra  
Philips Research Laboratories,  
Eindhoven  
frits.dijkstra@philips.com

Johan Dubbeldam  
Technische Universiteit Eindhoven  
j.l.a.dubbeldam@tue.nl

Paul Duineveld  
Philips Research Laboratories,  
Eindhoven  
paul.duineveld@philips.com

Luca Ferracina  
Universiteit Leiden  
ferra@math.leidenuniv.nl

Robbert Fokkink  
Technische Universiteit Delft  
r.j.fokkink@its.tudelft.nl

Miguel Frasson  
Universiteit Leiden  
frasson@math.leidenuniv.nl

Joris Geurts van Kessel  
RIKZ, Middelburg  
j.geurtsvankessel@rikz.rws.minvenw.nl

Yodi Gunawan  
Technische Universiteit Eindhoven  
a.y.gunawan@tue.nl

Huub Hardy  
Netherlands Forensic Institute  
h.hardy@nfi.minjus.nl

Geertje Hek  
Universiteit van Amsterdam  
ghek@science.uva.nl

Andre Hoogstrate  
Netherlands Forensic Institute  
a.hoogstrate@nfi.minjus.nl

Rein van der Hout  
Universiteit Leiden  
houtr@math.leidenuniv.nl

Huaxiong Huang  
York University, Toronto  
hhuang@yorku.ca

Sudhir Jain  
Aston University, Birmingham  
S.Jain@aston.ac.uk

Rik D.T. Janssen  
Technologiestichting STW, Utrecht  
rik@stw.nl

Peter van der Kamp  
Vrije Universiteit Amsterdam  
peter@few.vu.nl

Belinda Kater  
RIVO-CSO, Yerseke  
b.j.kater@rivo.dlo.nl

Ies Keereweer  
Netherlands Forensic Institute  
i.keereweer@nfi.minjus.nl

Graham Kendall  
University of Nottingham  
gxk@cs.nott.ac.uk

Ger Koole  
Vrije Universiteit Amsterdam  
koole@cs.vu.nl

Gregory Kozyreff  
OCIAM, Oxford  
kozyreff@maths.ox.ac.uk

Simon Kronemeijer  
Universiteit van Amsterdam  
kronemj@science.uva.nl

Dinard van der Laan  
Universiteit Leiden  
dvdlaan@math.leidenuniv.nl

Hai Xiang Lin  
Technische Universiteit Delft  
h.x.lin@its.tudelft.nl

Kamyar Malakpoor  
Technische Universiteit Eindhoven  
k.malakpoor@TUE.nl

Ludolf Meester  
Technische Universiteit Delft  
L.E.Meester@EWI.TUdelft.nl

Leila Mohammadi  
Universiteit Leiden  
leila@math.leidenuniv.nl

Jaap Molenaar  
Technische Universiteit Eindhoven  
j.molenaar1@tue.nl

Jeroen Mulder  
KLM, Amstelveen  
jeroen.mulder@klm.com

Misja Nuyens  
Universiteit van Amsterdam  
mnuyens@science.uva.nl

Nick Ovenden  
Technische Universiteit Eindhoven  
n.c.ovenden@tue.nl

Marc Paelinck  
KLM, Amstelveen  
marc.paelinck@klm.com

Marc Paulhus  
Den Haag  
paulhus@wanadoo.nl

Mark Peletier  
CWI, Amsterdam  
peletier@cwi.nl

Derk Pik  
Universiteit Leiden  
drpik@math.leidenuniv.nl

Bob Planqué  
CWI, Amsterdam  
rplanque@cwi.nl

Harmen van der Ploeg  
Universiteit van Amsterdam  
hvdploeg@science.uva.nl

Colin Reeves  
Coventry University  
C.Reeves@coventry.ac.uk

José Antonio Rodríguez  
Universiteit Leiden  
rodriguez@math.leidenuniv.nl

Nicole Rommelse  
Midden Brabant College, Tilburg  
nic\_76@hotmail.com

Bart van de Rotten  
Universiteit Leiden  
barotten@math.leidenuniv.nl

Vivi Rottschäfer  
Universiteit Leiden  
vivi@math.leidenuniv.nl

Yves Rozenholc  
Technion University, Haifa  
rozen@math.jussieu.fr

Firdian Rusdi  
Technische Universiteit Delft  
f.rusdi@ITS.TUdelft.nl

Marjan Sjerps  
Netherlands Forensic Institute  
m.sjerps@nfi.minjus.nl

Flora Spieksma  
Universiteit Leiden  
spieksma@math.leidenuniv.nl

Muhammad Suleman  
University of Oxford  
omer.suleman@exeter.ox.ac.uk

Mikhail Tchesnokov  
Universiteit Twente, Enschede  
m.tchesnokov@cs.utwente.nl

Karin Troost  
RIVO, Yerseke  
K.Troost@rivo.dlo.nl

Nienke Valkhoff  
Universiteit van Amsterdam  
nienke@science.uva.nl

JF Williams  
University of Bath  
J.F.Williams@maths.bath.ac.uk

Phillip Wilson  
University College London  
plw@liquidus.net

Koos van Winden  
Vrije Universiteit Amsterdam  
Koos@cybercom.net

Dmitri Znamenski  
Vrije Universiteit Amsterdam  
dznamen@cs.vu.nl

## CHAPTER 1

# Solution Methodologies for generating robust Airline Schedules

F.Bian, E.Burke, S.Jain, G.Kendall, G.M.Koole, J.Mulder, M.C.E.Paelinck, C.Reeves, I.Rusdi, M.O.Suleman

**ABSTRACT.** Aircraft fleet can have a major effect on the efficiency and smooth running of an airline. Constructing good quality schedules is vital for an airline to operate in an effective and efficient way in order to accomplish high levels of consumer satisfaction and to maximise profits. The robustness of an airline schedule is an indicative measure of how good the schedule is because a robust plan allows the airline to cope with the unexpected disturbances which normally occur on a daily basis. Here we describe a technique to measure the robustness of schedules for aircraft fleet scheduling within KLM airlines. The method is based on the 'Aircraft on Ground (ACOG)' measure, it employs statistical methods (although alternative methods were also considered) and it is shown to provide a good estimation of the robustness of a given schedule.

**KEYWORDS:** Airline Scheduling, Modelling, Schedule Quality Measures

## 1. Introduction

The problem of generating fleet schedules is crucially important to the efficiency of an airline [3, 4]. An effective schedule can lead to significant savings. It can also, and perhaps more importantly, contribute to higher levels of customer satisfaction. Customers who experience regular delays with a particular airline are likely to take their custom elsewhere. Of course, delays are inevitable for a wide range of reasons (e.g. technical breakdowns, security alerts, adverse weather etc.). However, an indicative measure of the quality of an airline schedule is its level of robustness: How well can a schedule cope with a delay(s) to a particular aircraft(s)? Is there enough slack in the schedule to minimise the knock on effect of a delay to a particular aircraft. If there is no slack in the schedule then a delay to one aircraft could affect a significant proportion of the fleet. This could have major resource implications. If passengers miss connecting flights then the airline has to cover the incurred costs. However, building slack into the schedule is expensive. It essentially involves aircraft standing idle. One of the goals in trying to generate a high quality fleet schedule is to build in enough slack to ensure that the schedule has an acceptable level of robustness while, at the same time, attempting to keep costs at an effective level. It would be very easy indeed to build a very robust schedule. However, it would

be too expensive to implement. It would also be possible to build a schedule which minimises cost by decreasing aircraft idle time. However, this could easily lead to an increase in the overall incurred costs if one minor delay to one aircraft leads to a chain of delays. In summary, the goal is to provide an effective balance between robustness and aircraft idle time.

The integration of schedule optimisation algorithms and other systems in an airline is crucial to achieve an effective scheduling environment that considers all functions of the airline [23]. Reviews of research on airline scheduling are presented in [28]. A more recent survey on models and solution methods for a range of problems in aircraft scheduling was carried out by [16].

Aircraft scheduling is often addressed simultaneously with other associated problems. An example is provided by fleet assignment with time windows where the assignment of aircraft is carried out simultaneously to scheduling flight departures in order to improve flight connection opportunities and minimise costs [26]. The scheduling of maintenance operations and of aircraft are considered simultaneously using network models and a two phase heuristic by [14] while crew availability and maintenance operations are taken into account while tackling the fleet assignment problem in [10]. The additional constraint of equal aircraft utilisation when tackling fleet assignment and aircraft routing problems is considered by [2]. A network model for large scale fleet assignment problems that permits the expression of constraints within a unified framework was presented by [31].

Integer linear programming techniques have been applied by several researchers to tackle fleet assignment, aircraft routing and related problems [17, 27]. Dynamic programming and heuristics have also been investigated for the problem of fleet assignment [12]. In recent years, modern metaheuristics have also been used to tackle airline scheduling problems. For example, simulated annealing was applied to the optimisation of airline schedules by [22]. Also, [32] showed that by applying simulated annealing to the fleet assignment and aircraft routing, improvements of about 10 to 20 percent over the method used by the company could be achieved. A genetic algorithm was applied to generate alternative routes for air traffic by [25]. Also recently, genetic search methods have been applied to solve the problem of sequencing the arrival of aircraft in airports [9, 18].

Re-scheduling is a crucial activity for airlines and it has to be carried out on a daily basis due to a number of uncertainties and unforeseen events. Disruptions of planned schedules can result in a chain of events that can cause major disruptions throughout the system. A survey of techniques employed to recover from these disruptions is presented by [15]. A stochastic model is employed by [29] to show that the actual performance of an airline differs greatly from the planned performance while [1] propose a GRASP method to reconstruct schedules while minimising costs and satisfying constraints. Network models and Lagrangian relaxation were used by [37] for aircraft re-scheduling given a specific disruption that affects the airline operations greatly and causes substantial decrements in profits and levels of service: the temporary closure of airports (see also [33]). The problem of changing the assigned aircraft to specific flights while satisfying existing constraints is addressed by [19]. A steepest ascent local search heuristic was applied by [21] to re-schedule



aircraft and it was capable of finding good quality schedules in a short amount of time.

The allocation of arrival slots in airports affects the efficient implementation of airline schedules and this activity can be disrupted by many factors such as bad weather, cancelled flight and other unforeseen events [34]. Simulation models for these type of operations in airports are described in [24] while [11] proposed one of the earliest algorithms for the automation of these operations. Instead of the traditional FCFS (first-come-first-served) system, delay exchanges in arrival sequencing and scheduling permit airlines to express relative arrival priorities so that these can be taken into account for the arrival slot allocation [8]. The problem of scheduling aircraft when multiple runways are available has been addressed using queuing theory by [7]. A population heuristic was applied by [6] for the optimisation of the arrival sequence of aircraft to a UK airport in order to improve runway utilisation. Linear programming and an alternative heuristic were applied to the arrival sequencing problem with single and multiple runways by [5]. A detailed description of the dynamic planner used to carry out the scheduling, sequencing, runway allocation and other operations related to the scheduling of aircraft arrivals is given by [35].

Other related airline scheduling issues that have been investigated are for example:

- ★ The airline scheduling problem in charter companies which is different mainly because the market is well-known and the schedule can be changed completely from period to period [13].
- ★ The assignment and routing of a fleet of aeromedical airlifts in military sectors [30].
- ★ The impact that the rotation of aircraft has on the construction of schedules [36].
- ★ The construction of weekend fleet assignments [20].

The problem that is addressed in this paper is discussed in the next section. It represents a real world problem that faces KLM Airlines on a daily basis.

## 2. Problem description

Within KLM, two departments are responsible for the fleet schedule. The network planning department produce schedules which are then passed to the operations department who have the responsibility for implementing them and running them on a day-to-day basis. These two departments have conflicting objectives. The network department aims to produce a schedule which is as cost effective as possible. This, essentially means, maximising aircraft usage by minimising their idle time. The operations department have the reverse objective. The overall schedule has to achieve the kind of balance between these two objectives that is briefly described above.

The aim for KLM is to introduce a method that checks the robustness of a schedule, from the network department, before it is passed to the operations department for implementation. One way to achieve this is to run a simulation. However, this is seen as too time consuming and other methods are sought to test for the robustness of the schedule.

KLM flies to over 150 destinations using 97 aircraft. Four times a year, a new flight schedule is developed. Though the operational feasibility is taken into account to a certain degree during the development process, the aim at that stage is largely to maximise the number of seats that can be sold. During schedule development, KLM considers various commercial aspects such as the expected demand per destination and the number of possible transfer connections at Schiphol Airport.

The realisation of a flight schedule involves a number of parties. As described above, the initial plan is developed by KLM's Network department. The initial plan is based on commercial and strategic insights and long term plans for the fleet composition, cabin crew and baggage handling.

Two months before the beginning of a schedule plan the plan is handed over to the operational department, the Operation Control Center. From that moment on they are the owners of the plan and small adaptations have to be evaluated and approved by them. This department will try to prevent and solve problems such as emergencies and bottlenecks and, in case of unsolved problems, try to minimize the effects on succeeding flights. A final plan is created two weeks before the beginning of the plan where passenger bookings are matched with aircraft capacities.

In order to monitor the performance of a flight schedule, some critical performance indicators are defined. These are:

- \* The departure and arrival punctuality, that is the percentage of flights that departed or arrived on time.
- \* The completion factor, that is the percentage of accomplished flights. These are all flights that were not cancelled.
- \* The No Connection Passenger factor, that is the percentage of transfer passenger that missed their connections due to operational problems.
- \* The Irregularity-rate, that is the number of bags that were not delivered on time.

For the punctuality performance indicator the contribution of each of the involved parties is also monitored. This introduces the concept of building blocks. The whole operational process is divided into sub processes, (the so called building blocks). Each building block is owned by a capacity and service provider, these being Ground Services, Front Office, Air Traffic Management, Engineering and Maintenance, Cabin and Cockpit Crew, Cargo and Operations Control.

Seven Building Blocks have been established, these are called:

- BB1: Flight
- BB2: Arriving aircraft
- BB3: Layover aircraft
- BB4: Departing aircraft
- BB5: BB5.1 Transferring passengers BB5.2 Transferring baggage
- BB6: BB6.1 Arriving passengers BB6.2 Arriving baggage
- BB7: BB7.1 Departing passengers BB7.2 Departing baggage

The doors being opened and closed are the points at which responsibility passes from one capacity and service provider to another. The distinction of the 1st door being opened is made because a door can either be the passenger door(s) or a baggage door(s). For example, once a plane has physically landed it is not actually

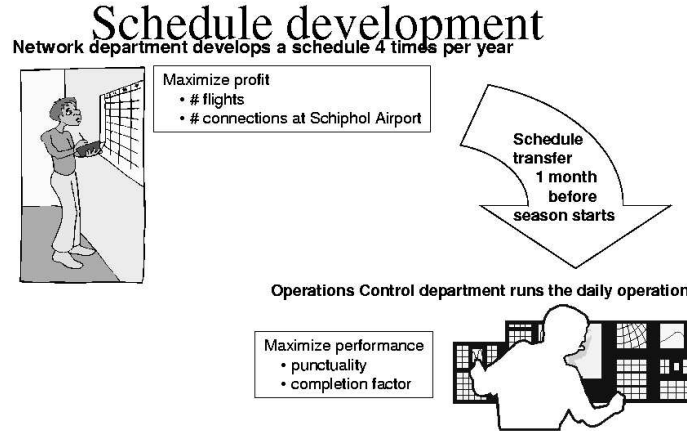


FIGURE 1. Schedule development

considered to have landed (i.e. responsibility passed to the ground staff) until one (passenger OR baggage) door has been opened. In contrast, responsibility changes again when all doors have been closed, not just one door. All these agreements and the flight schedule itself comes together into an operational plan (see figure 1). This functions as a contract between Network, Operations Control and the Building Blocks (Capacity and Service Providers). The plan covers an operational plan period of between 2 to 4 months spread over the year. It consists of agreements concerning a schedule plan and a capacity plan position for each specific period. It contains a demand driven schedule that has been fully checked with the Building Block representatives (Capacity and Service Providers) and Operations Control by means of an operational check. Eventually the agreements enable each provider to deliver an operational performance forecast. This could deviate from the targets as laid down in the corresponding Business Plan. Each operational plan will be finalized ultimately 2 months prior each operational plan period.

The schedule is usually published as an Aircraft Rotation Schedule, which is different each week. This is due to the fact that each day many adaptations are made so as to minimise delays. For instance, if KLM know that an aircraft will arrive at Schiphol Airport with a delay, they could assign its next flight to another aircraft so that that flight can still leave on time. Usually, KLM will also need other adaptations to have all flights fit into the Rotation Schedule again. When a schedule is first published, KLM do not know the exact layout of the Rotation Schedule, so they publish a hypothetical 'average' one instead.

Before a schedule is published, an estimation of the expected punctuality (that is the percentage of 'on time' flights) is performed using a simple deterministic model. As this model lacks accuracy, a simulation model is currently being developed in order to enable a better forecast. This model simulates aircraft movements according to a given schedule. The model subjects the schedule to a "stress test" by generating various disruptions such as air traffic congestion, delays during the boarding

process or unexpected problems during maintenance. Throughout the simulation, a Problem Solver algorithm attempts to resolve delays by swapping flights in the Rotation Schedule, or in extreme cases, by canceling flights. More successful runs of the simulation are considered as better schedules for implementation.

A simulation, though, has several disadvantages. Processing times are usually too long, which limits the number of schedules that can be assessed. Also, KLM need to collect a lot of data about the processes that are being simulated. For the simulation model currently under development they need statistics about the variation in the actual flight duration, the variation in the time it takes to handle an aircraft on the ground (boarding, fuelling, catering, etc.), break down times of each aircraft type etc. Each of these statistics must constantly be updated to reflect the change in flight routes, working methods, fleet, etc. KLM are currently seeking a more simple model that would enable them to make a comparative statement, such as: 'Of a number of alternative schedules, schedule X will provide the best performance'.

### 3. Models for the problem

It was anticipated that there should be some features of any schedule that would be correlated with its performance. The first question is then what features should be investigated? A brainstorming session with representatives of KLM led to some suggestions. It was expected that the number of potential swaps available to a delayed flight would be an important factor, but measuring this value was not easy. In practice, it might also be necessary to undertake a cascade of swaps, so another possible measure of performance would be the length of time and/or the number of swaps needed to restore the schedule to its normal condition. However, this is also complicated to determine, although the Problem Solver module of the simulation could be invoked if necessary.

After further discussion, it was agreed to look at a simpler measure, which could easily be found, and is arguably a surrogate for some of the more complex measure suggested. This is the 'Aircraft on Ground' (ACOG) measure discussed in the next section. Having obtained some features related to this measure, the next step is the identification of a suitable model for purposes of prediction. Candidates here include multiple linear regression methods, regression trees, neural nets and other pattern recognition techniques. However, the fact that the amount of data available was small meant that data-hungry methods should be avoided if at all possible. Thus it was resolved to begin the investigation with traditional statistical methods.

### 4. Experimental results

Eleven schedules were available (Summer/Winter 2000-2002, apart from the last 13 weeks of 2002). KLM's operation at Schiphol is such that the activity occurs in 4 major waves - a deliberate strategy to maximise passengers' opportunities for making onward connections. Graphing the number of aircraft available on the ground reveals this pattern clearly. These can be counted in 2 ways: the more accurate picture is obtained by subtracting the lengths of BB2 and BB4, leaving just those aircraft that are actually idle at a given moment. However, it is a simpler

calculation to count the whole of the time on the ground from 'First Door Open' to 'Last Door Closed', which comprises the whole of BBs 2,3 and 4. In the case of European operations, each day is more or less identical, so peaks can be defined quite easily. For each peak, the first 4 moments of the 'Aircraft on Ground' (ACOG) values were calculated for each day, using both definitions - BB3 and BB234. As days are so alike (apart from the very first day of a new schedule), one day can be selected at random as a representative of a schedule. As there are 4 peaks daily, we have 16 features as inputs, which we need to associate with the performance indicators (PIs) already calculate by KLM. The ones used for the models developed here were simply the departure and arrival punctualities: the fraction of planes (of those scheduled) that departed or arrived on time.

As a first step, correlations were calculated between the PIs and the 16 input variables. The 6 or 7 most highly correlated input variables were than used in a stepwise regression procedure (using S-plus) to determine the best balance between parsimony and explanatory power. (S-plus uses the Akaike Information Criterion for this purpose.) The table below summarises the models determined by this approach.

<b>PI – Departure Punctuality</b>		
	Using BB3 only	Using BB234
Predictor sets	p4m, p1sd, p1sk, p1k	p2m, p4m, p2sd, p4sd, p3sk
R-squared	95.6%	91.6%
P value (F-test)	0.00032	0.01028
<b>PI-Arrival Punctuality</b>		
Predictor sets	p4m, p1sk, p3sk, p3k	p1m, p4m
R-squared	95.2%	84.1%
P value (F-test)	0.00042	0.00064

**Table of experimental results; see text for further details.**

In this table, 'p1' means the 1st peak, 'm' is the 1st moment (mean), 'sd' the 2nd (standard deviation), 'sk' the 3rd (skewness) and 'k' the 4th (kurtosis). Of interest is the fact that 'p4m' - the mean number of ACOG - is important for all 4 models, but the other predictors seem to be an eclectic bunch. From KLM's point of view, this doesn't matter if the predictions are good enough, but from a modeller's perspective we would like to see more consistency. However, all models are based on just 11 data points, so perhaps the lack of consistency is not surprising. Prediction intervals can easily be obtained on the assumption of Normally distributed errors: these vary from +/- 2% for punctualities in the middle of the range to +/- 3% at the edges.

It was quite surprising that the R-squared values were as high as they were – we were anticipating that a linear model would be too simple, yet it seems quite powerful. Of course regression analysis makes certain assumptions about the errors, and it is necessary to check the residuals to see if these assumptions are plausible. The plot of residuals against fitted values was obtained for each model; in no case does a systematic pattern seem plausible, and a random scatter is obtained, as in figure 2. QQ plots of the residuals against Normal quantiles were also obtained. The tails of the distribution in particular are not well fitted, so the assumption that

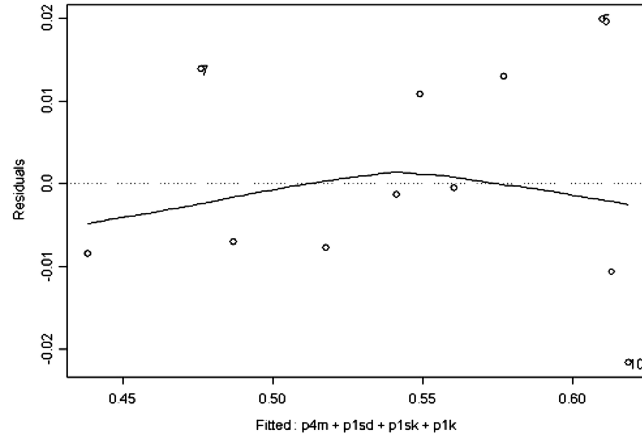


FIGURE 2. Residuals against fitted values for Departure Punctuality using BB3 only

the errors are Normally distributed is perhaps questionable. Thus any confidence intervals should be treated cautiously.

### 5. Logistic regression

In any case, the response variable in all 4 models is actually a ratio that is confined to remain between 0 and 1. This means that a better theoretical model would be based on a logistic transformation, since it is theoretically possible that a simple linear model could generate predictions outside the possible range of values. (We can hardly have a punctuality of greater than 100%!) Such a model would also be based on a more plausible probability model than the Normal distribution.

However, attempts to fit such a model did not produce an improvement. A possible explanation is that the data available are all in the region of approximate linearity of the logistic curve. Consequently, any attempt to identify the turning points of the curve is likely to be rather speculative. In any case, on inspecting the coefficients of the models, it seems unlikely that we would predict bizarre fractions in practice. For example, using the most extreme values observed in the 1st model above would predict only 80% departure punctuality, and in the opinion of KLM's experts it is hard to imagine physical circumstances in which these values could be exceeded simultaneously. (There is just not enough space to put many more planes, for example.)

Thus, despite the attractions of a more plausible theoretical model, the airline is comfortable with the predictive ability of a simpler linear model.

## 6. Conclusions

An analysis of the expected number of aircraft on ground has been shown to provide a good prediction for the robustness of a given schedule. Further refinements are possible - and desirable - but even this work has given the operations department a better insight into what makes a schedule easier or harder to implement effectively. Some of the work that still needs to be done includes an analysis of the effect of day-to-day variations in the schedule - these variations are small, but preliminary work has suggested that the definition of activity peaks needs to be tighter, and the possibility of a day-of-the-week effect should also be explored. Furthermore, the schedules examined so far have concentrated only on the European operations, where fleet homogeneity is substantial and diurnal variation is small. Incorporating the effects of the inter-continental timetable may lead to some changes in these conclusions.

## 7. Acknowledgements

We would like to thank the following bodies for supporting the collaboration described in this paper:

The Lorentz Centre in Leiden : 45th European Study Group for Mathematics with Industry.

The EU Mathematics, Computing and Simulation for Industry Network (MACSInet).

## Bibliography

- [1] Argüello M.F. and Bard, J.F., A GRASP for aircraft routing in response to groundings and delays, *Journal of Combinatorial Optimization*, 1, 3, pp 211–228, 1997.
- [2] Barnhart C., Boland N.L., Clarke L.W., Johnson E.L., Nemhauser G.L. and Sheno R.G., Flight String Models for Aircraft Fleeting and Routing, *Transportation Science, Focused Issue on Airline Optimization*, **32**, No. 3, pp 208–220, 1998.
- [3] Barnhart C., Lu F. and Sheno R., Integrated Airline Scheduling, *Operations Research in the Air Industry, International Series in Operations Research and Management Science*, **9**, edited by Prof. Gang Yu, Kluwer Academic Publishers, pp 384–403, 1997.
- [4] Barnhart C. and Talluri K., Airline Operations Research, *Design and Operations of Civil and Environmental Engineering Systems*, pp 435–469, 1997.
- [5] Beasley J., Krishnamoorthy M., Sharaiha Y. and Abramson D., Scheduling Aircraft Landings - The Static Case, *Transportation Science*, 34, pp 180–197, 2000.
- [6] Beasley J., Sonander J. and Havelock P., Scheduling Aircraft Landings at London Heathrow using a Population Heuristic, *Journal of the Operations Research Society*, 52, pp 483–493, 2001.
- [7] Bolender M.A. and Slater G.L., Evaluation of scheduling methods for multiple runways, *Journal of Aircraft*, **37**, pp 410–416, 2000.

- [8] Carr G.C., Erzberger H. and Neuman F., Fast-time study of airline-influenced arrival sequencing and scheduling, *Journal of Guidance, Control and Dynamics*, **23**, pp 526–531, 2000.
- [9] Ciesielski V. and Scerri P., Real time genetic scheduling of aircraft landing times, *Proceedings of the 1998 IEEE International conference on Evolutionary Computation (ICEC98)* Fogel D (editor), pp 360–364, 1998.
- [10] Clarke L.W., Hane C.A., Johnson E.L. and Namhauser G.L., Maintenance and crew considerations in fleet assignment, *Transportation Science*, 3, 3, pp 249–260, 2001.
- [11] Dear R.G. and Sherif Y.S., An algorithm for computer assisted sequencing and scheduling of terminal area operations, *Transportation Research Part A - Policy and Practice*, **25A**, pp 129–139, 1991.
- [12] El Moudani W. and Mora-Camino F., A dynamic approach for aircraft assignment and maintenance scheduling by airlines, *Journal of Air Transport Management*, 6, pp 233–237, 2000.
- [13] Erdmann A., Noltemeier A. and Schrader R., Modeling and solving an airline schedule generation problem, *Annals of Operations Research*, 107, 1-4, pp 117–142, 2001.
- [14] Feo T.A. and Bard J.F., Flight scheduling and maintenance base planning, *Management Science*, 35, 12, pp 1415–1432, 1989.
- [15] Filar J.A., Manyem P. and White K., How airlines and airports recover from schedule perturbations: a survey, *Annals of Operations Research*, 108, 1-4, pp 315–333, 2001.
- [16] Gopalan R. and Talluri K.T., Mathematical models in airline schedule planning: a survey, *Annals of Operations Research*, 76, 1, pp 155–185, 1998.
- [17] Hane C.A., Barnhart C., Johnson E.L., Marsten R.E., Nemhauser G.L. and Sigismondi G., The Fleet Assignment Problem: Solving a Large-Scale Integer Program, *Mathematical Programming*, 70,2, pp 211–232, 1995.
- [18] Hansen J.V., Genetic search methods in air traffic control, *Computers and Operations Research*, In press, 2003.
- [19] Jarrah A.I., An efficient airline re-fleetting model for the incremental modification of planned fleet assignments, *Transportation Science*, 34, 4, pp 349–363, 2000.
- [20] Kontogiorgis S. and Acharya S., US Airways automates its weekend fleet assignment, *Interfaces*, 29, 3, pp 52–62, 1999.
- [21] Love M., Sorensen K.R., Larsen J. and Clausen J., Disruption management for an airline - rescheduling of aircraft, *Applications of evolutionary computation: Proceedings of the EvoWorkshops 2002, Kinsale Ireland, Lecture notes in computer science*, **2279**, Springer, pp. 315–324, 2002.
- [22] Mashford J.S. and Marksjo B.S., Airline base schedule optimisation by flight network annealing, *Annals of Operations Research*, 108, 1-4, pp 293–313, 2001.
- [23] Mathaisel D.F.X., Decision support for airline schedule planning, *Journal of Combinatorial Optimization*, 1, 3, pp 251–275, 1997.
- [24] Milan J., The flow management problem in air traffic control: a model of assigning priorities for landings at a congested airport, *Transportation Planning and Technology*, **20**, pp 131–162, 1997.



- [25] Oussedik S., Delahaye D. and Schoenauer M., Flights alternative routes generator by genetic algorithms, *Proceedings of the 2000 congress on evolutionary computation*, CEC 2000, pp 896–901, 2000.
- [26] Rexing B., Barnhart C., Kniker T., Jarrah A. and Krishnamurthy N., Airline Fleet Assignment with Time Windows, *Transportation Science*, **34**, No. 1, pp 1–20, 2000.
- [27] Richardson R., An optimization approach to routing aircraft, , pp 52–71, 1975.
- [28] Richter H., Thirty Years of Airline Operations Research, *Interfaces*, 19, pp 3–9, 1989.
- [29] Rosenberger J.M., Schaefer A.J., Goldsman D., Johnson E.L., Kleywegt A.J. and Nemhauser G.L., A Stochastic Model of Airline Operations, *Transportation Science*, **36**, No. 4, 2003.
- [30] Ruland K.S., A model for aeromedical routing and scheduling, *International Transactions in Operational Research*, 6, 1, pp 57–73, 1999.
- [31] Rushmeier R.A. and Kontogiorgis S.A., Advances in the optimization of airline fleet assignment, *Transportation Science*, 31, 2, pp 159–169, 1997.
- [32] Sosnowska D. and Rolim J., Fleet scheduling optimization: a simulated annealing approach, *The Practice and Theory of Automated Timetabling III: Selected Papers from the 3rd International Conference on the Practice and Theory of Automated Timetabling PATAT 2000, Konstanz Germany, Lecture Notes in Computer Science, Springer*, **2079**, pp 227–241, 2001.
- [33] Thengvall B.G., Yu G. and Bard J.F., Multiple fleet aircraft schedule recovery following hub closures, *Transportation Research Part A*, 35, pp 289–308, 2001.
- [34] Vazquez-Marquez A., American airlines arrival slot allocation system (ASAS), *Interfaces*, 21, 1, pp 42–61, 1991.
- [35] Wong G.L., The dynamic planner: the sequencer, scheduler, and runway allocator for air traffic control automation, *Report NASA/TM-2000-209586, NASA Ames Research Center, Moffett Field, CA 94035-1000, USA. Available from <http://www.ctas.arc.nasa.gov/publications>*, 2000.
- [36] Wu C.L. and Caves R.E., Towards the optimisation of the schedule reliability of aircraft rotations, *Journal of Air Transportation Management*, 8, pp 419–426, 2002.
- [37] Yan S. and Lin C.G., Airline scheduling for the temporary closure of airports, *Transportation Science*, 31,1, pp 72–82, 1997.



## CHAPTER 2

# Japanese Oysters in Dutch Waters

Jan Bouwe van den Berg, Gregory Kozyreff, Hai-Xiang Lin, John McDarby,  
Mark A. Peletier, Robert Planqué, Phillip L. Wilson

Other participants:

Dragan Bezanovic, Luca Ferracina, Joris Geurts van Kessel<sup>1</sup>, Belinda Kater<sup>1</sup>,  
Kamyar Malakpoor, Harmen van der Ploeg, José A. Rodríguez, Bart  
van de Rotten, Karin Troost<sup>1</sup>, Nienke Valkhoff, and J.F. Williams

**ABSTRACT.** We study a number of aspects of the colonisation of the Eastern Scheldt by the Japanese Oyster. We formulate and analyse some simple models of the spatial spreading, and determine a rough dependence of the spreading behaviour on parameters. We examine the suggestion of reducing salinity by opening freshwater dams, with the aim of reducing oyster fertility, and make predictions of the effect of such measures. Finally, we present an outline of a large-scale simulation taking into account detailed data on the geometry and sea floor properties of the Eastern Scheldt.

**KEYWORDS:** Japanese Oyster, *Crassostrea gigas*, population dynamics

## 1. Introduction

In 1964 the Japanese Oyster (*Crassostrea gigas*) was introduced into the Eastern Scheldt. It was believed that this species could not breed in the colder Dutch climate; each generation would have to be set out by hand. Therefore this introduction was expected to have limited impact on the local ecosystem. Additionally, at that time the plan was to close off (part of) the Eastern Scheldt, so that even if spawning occurred, the problem would remain local.

Unfortunately the brief hot spells of some Dutch summers allowed the Japanese Oysters to spawn. With a maximal life span of thirty years the population proved able to spawn in the rare hot years and simply survived in other years. As a result, the Japanese Oyster is now a dominant species in the Eastern Scheldt; the indigenous flat oyster (*Ostrea edulis*) has disappeared almost completely (mainly due to disease and the very cold winter of 1963), the cockles are declining in number, and mussels have been confronted by the appearance of a strong competitor for food. In addition, the Japanese Oyster has spread beyond the Eastern Scheldt and

---

<sup>1</sup>We would like to direct special thanks to the proposers of the problem for the information and data provided, the corrections suggested and the hospitality in Yerseke.



FIGURE 1. A bank of Japanese Oysters

settled in parts of the Western Scheldt and the Wadden Sea. At present, the main negative impact is that the Japanese Oysters compete with cockles for space and food. In turn, the decline in cockles causes problems for the birds that feed on them.

At the Study Group two Dutch institutes, the Nederlands Instituut voor Visserijonderzoek (The Netherlands institute for fisheries research<sup>2</sup>) and Rijksinstituut voor Kust en Zee (RIKZ, The Netherlands institute for coastal and marine management) presented the issue of the spreading of the Japanese Oysters. The following questions were formulated:

- How do oysters spread?
- Can development in the past be reconstructed?
- Can a prediction be made for the future?
- Can the spreading of the oysters be stopped? How?

## 2. Overview

In this report we address these questions from a number of different viewpoints. Let us give an overview of the different models and their outcomes up front.

- In Section 3 we present a first study of the spreading of oysters. We assume a favourable (i.e. hard) substrate and model the spread of oysters from year to

---

<sup>2</sup>Formerly known as RIVO, presently part of Animal Sciences Group.



FIGURE 2. Map of the Eastern Scheldt

year. We indicate how to extend this model to perform simulations to reconstruct the spreading of oysters from 1964 onwards.

- In Section 4 we simplify the model of Section 3 by passing to continuous time, and compare spreading velocities between soft and hard substrates. We find a significant difference in spreading velocity, and argue that the ratio of the two velocities is relatively independent of the important parameters, which are difficult to estimate.
- In Section 5 we study a proposed remedy of re-opening the dams that currently prevent river water from entering the Eastern Scheldt. The reduction in salinity that results from the influx of fresh water may reduce the growth rate of the oysters. In extension of a recent simulation at RIKZ we consider a scenario of partial, seasonal re-opening of the dams, and find that the effect on salinity is similar (to a permanent reopening scenario), but with some advantages. The main conclusion, however, is that the proposed regulation of salinity is not sufficient to significantly control the growth of oysters in the Eastern Scheldt.
- In Section 6, finally, we present an outline of a large-scale simulation that takes into account the detailed geometry and geology (e.g. substrate hardness) of the Eastern Scheldt. This simulation might be implemented as an extension of a code that is currently in use at Rijkswaterstaat.

### 3. Diffusion of larvae

At the end of July of a hot year, the rising of the sea temperature over a certain threshold triggers a massive production of oyster larvae. During their 15–30 day life span, these larvae are passively transported by the flow in random directions until they settle. This transport is probably the main mechanism by which *Crassostrea gigas* invaded the whole Eastern Scheldt. In this section, we will analyse this mechanism, in conjunction with a simple model of the interaction between the oyster and larvae populations. We will restrict our attention to the eastern part of the Eastern Scheldt, where the *Crassostrea gigas* was introduced, in 1964. Indeed, this region is characterised by relatively shallow water and weak currents, which hampers the geographic progression of oysters. Because of this, it took years before the oyster population reached the central part of the Eastern Scheldt. Afterwards, larvae became subjected to much stronger currents and were therefore prone to colonise the rest of the Eastern Scheldt in a relatively short period. At least, this is one of the possible scenarios. Slow adaptation of the Japanese Oysters to the local environment may also have contributed to the time delay before the central part was reached. Probably a combination of factors, partly due to the construction of the Delta works, has caused the explosion of Japanese Oysters in the Eastern Scheldt.

The simplest description of the geographical spreading of oysters should comprise two independent variables: one for the larvae population and one for the oysters. Hence we introduce:

- $\mathcal{O}_n(x)$ : the density of oysters, expressed in  $\text{m}^{-2}$ , in year  $n$ .  $\mathcal{O}_n$  depends on the position  $x$  (which is one-dimensional, for simplicity);
- $\mathcal{L}_n(x, t)$ : the density, also expressed in  $\text{m}^{-2}$ , of larvae in the summer of year  $n$ .  $\mathcal{L}_n$  depends on position  $x$  and time  $t$ .

During their short life, we model the transport of larvae by a reaction-diffusion equation of the form:

$$(1) \quad \frac{\partial \mathcal{L}_n}{\partial t} = D_T \frac{\partial^2 \mathcal{L}_n}{\partial x^2} + Q(\mathcal{L}_n, \mathcal{O}_n).$$

In this equation, the first two terms describe the diffusive transport averaged over an entire tidal cycle. The last term  $Q$  accounts both for the production and the disappearance of larvae, either by death or by settlement on the ground. A crucial parameter is the diffusion coefficient  $D_T$  and we will estimate it below. As for the oysters, their density varies from year to year according to:

$$(2) \quad \mathcal{O}_{n+1} = \mathcal{O}_n + G(\mathcal{L}_n, \mathcal{O}_n),$$

where  $G$  is the number of newly born oysters minus the deceased ones per unit area. For the moment, we do not specify the functionals  $Q$  and  $G$ . Several choices of  $Q$  and  $G$  will be presented in this report and many variations are possible. The choice between these requires a delicate balancing of the questions that are to be addressed, on one hand, with the available data on the other. More complex models, often used for the purpose of tracking growth in cultivated oysters, sort the individuals by sizes and introduce as many oyster variables as there are size-classes [3]. In

this work, motivated by the relative lack of data<sup>3</sup>, we discard such aspects of the dynamics and, with the exception of the final Section, focus on models of minimal complexity.

In the following sections we first model the spreading of oysterlarvae to determine  $D_T$ . (in §3.1). Then we discuss the factors that influence the growth, survival and reproduction of oysters and larvae, and we present the outcome of the model (in §3.2). Finally, in §3.3, a continuous time limit is derived, which serves as a connection to the continuous time model discussed in Section 4.

**3.1. Diffusion-convection over a single tidal cycle.** On the time scale of a tidal cycle the larvae are subject to a tidal flow. With  $u$  the water velocity generated by the tides, the transport equation for the larvae becomes (setting  $Q = 0$  for the present discussion)

$$(3) \quad \frac{\partial \mathcal{L}}{\partial t} + u \frac{\partial \mathcal{L}}{\partial x} = D \left( \frac{\partial^2 \mathcal{L}}{\partial x^2} + \frac{\partial^2 \mathcal{L}}{\partial z^2} \right),$$

where  $D$  is the coefficient of diffusion in the absence of tide and where  $\mathcal{L}$  is assumed to depend on the vertical coordinate  $z$  too.

As was first recognised by Taylor [15], a nonuniform vertical distribution of  $u$  accelerates the dispersion of particles. This can be understood by noting that at a depth  $z$  where  $u$  is maximal, particles (e.g. larvae) are likely to travel over much longer distances than those at depths where  $u$  is small.

If  $u = u(z)$ , one can show that equation (3) can be approximated in the long time limit by the following, simpler equation:

$$(4) \quad \frac{\partial \mathcal{L}}{\partial t} + U_0 \frac{\partial \mathcal{L}}{\partial x} = (D + D_T) \frac{\partial^2 \mathcal{L}}{\partial x^2},$$

where  $U_0$  is the average velocity of the flow over the  $z$ -direction. While equation (4) applies to the rising tide, the falling tide is described by

$$(5) \quad \frac{\partial \mathcal{L}}{\partial t} - U_0 \frac{\partial \mathcal{L}}{\partial x} = (D + D_T) \frac{\partial^2 \mathcal{L}}{\partial x^2}.$$

Hence, averaging over many tides, one obtains

$$(6) \quad \frac{\partial \mathcal{L}}{\partial t} = (D + D_T) \frac{\partial^2 \mathcal{L}}{\partial x^2}.$$

Since we are primarily interested in the diffusion of larvae in the horizontal directions, and owing to its relative simplicity, equation (6) represents progress from equation (3).

With the sea level and sea bed respectively denoted by  $h$  and  $z_0$ , the new diffusion coefficient is given by [11]

$$(7) \quad D_T = \frac{U_0}{D(h - z_0)} \int_{z_0}^h \left[ \int_{z_0}^z \left( 1 - \frac{u(z')}{U_0} \right) dz' \right]^2 dz.$$

Our first task will therefore be to assess the velocity profile  $u$  resulting from the tidal flow. Actually,  $u$  depends on  $x$  as well as  $z$ . Hence,  $D_T = D_T(x)$  and equation (4) has to be slightly modified, as shown in the appendix.

---

<sup>3</sup>Data are hard to obtain since experiments are difficult and time consuming.

3.1.1. *Tidal flow.* For a shallow part of the sea the determination of  $u$  is relatively simple. Let the elevation of the sea bed be denoted by  $z_0(x)$ . As a result of the tides, the sea level is a function of time and given by  $z = h(t)$ . From the “shallowness” hypothesis, and assuming low velocities, the Navier-Stokes equations for the flow reduce to:

$$(8) \quad 0 = -\frac{dp}{dx} + \mu \frac{\partial^2 u}{\partial z^2}.$$

In this equation,  $p$  is the pressure and  $\mu$  is the viscosity of water. This equation must be supplemented by two boundary conditions. One is that the velocity vanishes on the sea bed,  $u(z_0) = 0$ . The other is that the sea surface is free of any applied stress, which translates into  $\frac{\partial u}{\partial z}(h) = 0$ . This allows to write the solution of (8) as:

$$(9) \quad \begin{aligned} u(x, z) &= \frac{1}{2\mu} \frac{dp}{dx} (z - z_0)(z - 2h + z_0), \\ &= C(z - z_0)(z - 2h + z_0). \end{aligned}$$

In this expression the constant  $C$  is determined from the conservation of mass. Considering a slice  $dx$  of fluid, the rise or fall of its level,  $\frac{dh}{dt}$ , is only due to incoming and outgoing flux of water on either sides of the slice. This leads to the equation

$$(10) \quad \frac{dh}{dt} + \frac{\partial}{\partial x} \int_{z_0}^h u(x, z) dz = 0.$$

Substituting (9) into (10), we find:

$$(11) \quad u(x, z) = \frac{3}{2} U_0 \frac{(z - z_0)(2h - z - z_0)}{(h - z_0)^2},$$

where  $U_0$  is the average velocity, given by

$$(12) \quad U_0 = \frac{dh/dt}{dz_0/dx}.$$

3.1.2. *Estimation of  $D_T$ .* The value of  $D$  is estimated to be  $10^{-4} \text{ m}^2 \text{ s}^{-1}$  for stratified flow and  $10^{-3} \text{ m}^2 \text{ s}^{-1}$  for well mixed flows [17]. These values are obtained by measuring the diffusion in the vertical direction, which is not affected by the tidal flow. During the rising tide, the sea level rises by 3 meters in 6 hours, and we assume the slope of the sea bed to be 1%. Hence,  $U_0$  is estimated to be

$$U_0 \approx \frac{3 \text{ m} / (6 \cdot 3600 \text{ s})}{0.01} \approx 0.02 \text{ ms}^{-1}.$$

Then, substituting expression (11) into (7), we obtain:

$$(13) \quad D_T = \frac{2U_0(h - z_0)^2}{105D}.$$

Hence, with a sea depth of 3-4 m, a vertical diffusion  $D$  of  $10^{-3} \text{ m}^2 \text{ s}^{-1}$  and our estimate of  $U_0$ , we get

$$D_T \approx 0.1 \text{ m}^2 \text{ s}^{-1},$$

which is considerably larger than  $D$ .

As we already noted,  $D_T$  is a function of  $x$ . It is therefore tempting to simply rewrite the diffusion term in the right hand side of (6) as  $\frac{\partial}{\partial x} D_T(x) \frac{\partial}{\partial x} \mathcal{L}$ . However,



one must bear in mind that the effective parameter  $D_T(x)$  encompasses more than just Fick's law of transport. The actual reduced diffusion equation turns out to be (see the appendix)

$$(14) \quad \frac{\partial \mathcal{L}}{\partial t} = D_T \left[ \left( 1 + \frac{D}{D_T} \right) \frac{\partial^2 \mathcal{L}}{\partial x^2} - \frac{12 (dz_0/dx)}{(h - z_0)} \frac{\partial \mathcal{L}}{\partial x} \right].$$

For simplicity, in what follows, we will neglect spatial variations of  $D_T$ .

**3.2. Completion of the model and outcome.** In order to be able to forecast the expansion of the oysters, one needs to choose sensible and simple forms for  $Q$  and  $G$ .

Female and male oysters are able to detect the presence of eggs and sperm in the water [12]. In order to maximise fecundation, they all release their gametes at the same time. Accordingly, the production of larvae is modelled by the initial condition

$$(15) \quad \mathcal{L}_n(x, 0) = F \mathcal{O}_n,$$

where  $F$  is the fecundity of an oyster and is the combination of several factors, such as the size of the oyster, the salinity of water and the likelihood of an egg to be fertilised. An expression derived from observation is [9]

$$F = F_l F_s F_f$$

with

- $F_l = 0.4 l^{2.8}$ : size factor, where  $l$  is the size of the oyster in cm.
- $F_s = \frac{s-8}{5.5}$ : salinity factor, where  $s$  is the salinity of water, expressed in grams of chloride per liter,
- $F_f = 0.005 \mathcal{O}_n^{0.72}$ : fertilisation efficiency.

Besides, since the larvae only have an approximate 15–30 day life span, one includes a death rate in the larvae equation:

$$(16) \quad \begin{aligned} \frac{\partial \mathcal{L}_n}{\partial t} &= D_T \frac{\partial^2 \mathcal{L}_n}{\partial x^2} - \frac{\mathcal{L}_n}{t^*}, \\ t^* &= 30 \text{ days.} \end{aligned}$$

Of the larvae that are transported according to this equation, only a tiny fraction  $\lambda$  will actually be able to settle. The magnitude of this fraction depends on a number of factors, among which the hardness of the substrate and the presence of other oysters. For the length of this section we will assume that the substrate is hard, providing a good environment for settling larvae. In Section 4 we study the effect of substrate hardness in detail.

Presence of other oysters on the substrate will result in competition for nutrient and predation of larvae from the mature oysters. This overcrowding factor is supposed to be only effective when the oyster density passes a certain threshold  $\mathcal{O}_{sat}$ . Finally, approximately one tenth of the oyster population dies each year [3]. These considerations lead to the following equation for the oyster population

$$(17) \quad \mathcal{O}_{n+1}(x) = \mathcal{O}_n(x) + \frac{\lambda \mathcal{L}_n(x, t^*)}{1 + \mathcal{O}_n(x) / \mathcal{O}_{sat}} - \frac{\mathcal{O}_n(x)}{10}$$

From the observation of actual oyster banks,  $\mathcal{O}_{sat}$  ranges between  $30 \text{ m}^{-2}$  and  $100 \text{ m}^{-2}$ .

Prior to the integration of equations (15)-(17), it is useful to introduce dimensionless variables as

$$T = \frac{t}{t^*}, \quad X = \frac{x}{(D_T t^*)^{1/2}} \approx \frac{x}{400 \text{ m}},$$

$$u_n = \frac{\lambda \mathcal{L}_n}{\mathcal{O}_{sat}}, \quad v_n = \frac{\mathcal{O}_n}{\mathcal{O}_{sat}}.$$

The system of equations to solve for each year thus becomes

$$(18) \quad \frac{\partial u_n}{\partial T} = \frac{\partial^2 u_n}{\partial X^2} - u_n,$$

$$(19) \quad u_n(X, 0) = \Lambda v_n(X),$$

$$(20) \quad v_n = 0.9v_{n-1} + \frac{u_{n-1}(X, 1)}{1 + v_{n-1}}.$$

In these equations, the only free parameter that remains is

$$\Lambda = \lambda F.$$

It represents the number of newly born oysters per oyster per year in the absence of overcrowding effect. Each year,  $n$ , the diffusion equation (18) is integrated with initial condition (19) and the oyster population is updated according to (20).

From 1964 onwards, oysters were introduced and cultivated on 100 m wide squares approximately 8 km away from the central region. In the new space scale,  $100 \text{ m} \rightarrow 0.25$  and  $8 \text{ km} \rightarrow 20$ . We assume now that there has been (a small amount of) spawning every year since the introduction in 1964. The first settling of oyster larvae, called spat at this stage, on dike foets and jetties was recorded in 1976. Hence, we need to integrate the system above for 12 years,  $n = 1, \dots, 12$ , and check whether a significant number of oyster/larvae could reach the central region. Initially, the oyster distribution is given by

$$v_0(X) = \begin{cases} 1, & 0 < X < 0.25, \\ 0, & 0.25 < X. \end{cases}$$

It is a simple task to integrate numerically the system (18)-(19) with the initial condition above. The result is shown in Figure 3 and suggests that the development in the past can indeed be constructed with this simple model provided that  $\Lambda \geq 8$ -10.

**3.3. Continuous time dependence for oysters.** To close this section, let us remark that the system (18)-(19) can be made amenable to further analytical development by turning the difference equation (20) into a differential one. Indeed, assuming that the variation of oyster population density is small from year to year, one can write

$$v(n+1) \approx v(n) + \frac{\partial v}{\partial n},$$

where  $n$  is now considered as a continuous variable that measures time in units of years. One thus has:  $n = \frac{t}{\text{year}} = \frac{t}{\text{month}} \frac{\text{month}}{\text{year}} = \varepsilon T$ , so that, generally, the

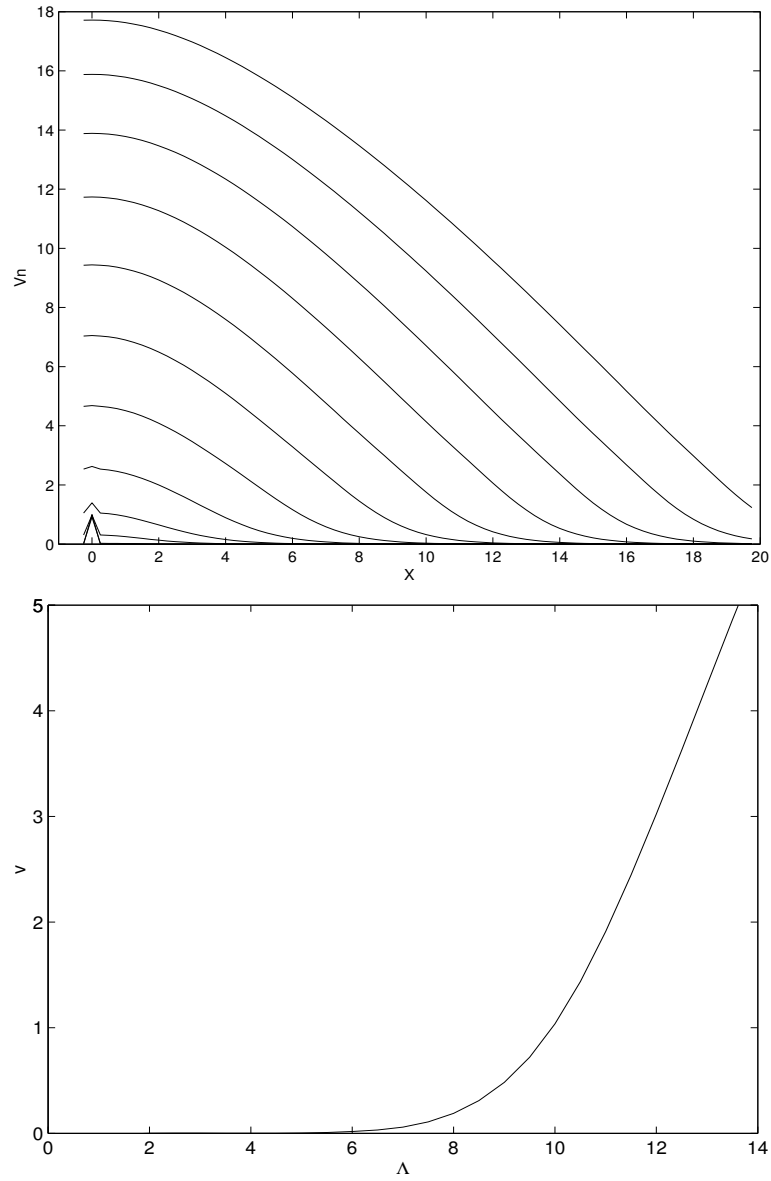


FIGURE 3. Yearly evolution of the oyster distribution  $v_n$  with  $\Lambda = 10$  (top) and oyster density  $v_{12}$  at the border of the Eastern region for various values of  $\Lambda$  after 12 years (bottom).

larvae/oyster model can be cast in the following form

$$(21) \quad \varepsilon \frac{\partial u}{\partial n} = \frac{\partial^2 u}{\partial X^2} + Q,$$

$$(22) \quad \frac{\partial v}{\partial n} = G.$$

**3.4. Discussion.** In this section we have sketched a way to model the oyster expansion in the Eastern Scheldt. The simplest choice of the generation rates  $Q$  and  $G$  of larvae and oysters, respectively, already proves quite informative. After rescaling variables, only one free parameter remains:  $\Lambda$ , the number of descendants per oyster per year in the absence of any overcrowding effects. It was found that  $\Lambda$  needs to be in the order of 10 in order to reconstruct a specific aspect of the history of oyster proliferation, the time lapse between the introduction in 1964 and the first large-scale sightings in 1976. Although this value may appear large, one must bear in mind that it does not account for overcrowding effects. Moreover, the more accurate diffusion model (19) would probably require smaller values of  $\Lambda$ , owing to the dependence of  $D_T$  on the sea depth.

Although the proliferation of oysters was only observed in 1976, the present analysis suggests that it had actually been taking place right from the beginning of their implantation in 1964. Because it was under water and at some distance from the shores and coasts, it is possible that the process went unnoticed before 1976.

Let us finally note the large values attained by the oyster density in figure 3. This results from the relatively long life span of oysters. However, overcrowding should be taken into account in the death rate of oysters. It is indeed observed that oysters grow on top of each other, so that only the top layer is alive. This point is (at least partially) addressed in the next sections.

#### 4. The effect of substrate hardness on the spreading of an oyster bank

In this section we study the spreading of an oyster bank where we concentrate on the phenomenon that oysters prefer to settle on a hard substrate rather than a sandy one.

**4.1. The Fisher-Kolmogorov equation.** The classical model [10] for spatial spreading of any biological species is the Fisher-Kolmogorov (FK) equation

$$(23) \quad \frac{\partial u}{\partial t} = D\Delta u + f(u),$$

where  $\Delta$  is the Laplacian (in the spatial coordinates),  $D$  the diffusion constant and  $f$  is a nonlinear function, typically

$$(24) \quad f(u) = u(1 - u)$$

or

$$(25) \quad f(u) = u(u - a)(1 - u) \quad \text{with } 0 \leq a \leq 1.$$

The Fisher-Kolmogorov equation has been analysed in excruciating detail. We will only recall a few results here to serve as a guide for a more specialised model presented in §4.2.

In the present context  $u$  can be interpreted as the biomass of the oysters (or their number) per unit area. The crucial difference between the nonlinearities (24) and (25) is that for the former the trivial equilibrium  $u = 0$  is unstable while for the latter it is stable (at least in the absence of diffusion). The equilibrium  $u = 1$  is stable in both cases. We will refer to (24) as the monostable case and to (25) as the bistable case. One interpretation is that the monostable case corresponds

to oyster growth on a hard substrate (rocks or concrete), whereas the bistable case corresponds to oyster growth on a soft substrate (a sand bank). Without going into a biological interpretation we now state some mathematical results. In §4.2 we discuss the choice of nonlinearities in more detail.

The dynamics of solutions of (23) are dominated by travelling wave solutions, i.e., solutions of the form  $u = U(x - ct)$ , where  $c$  is the speed of the wave, and  $x$  is the direction of propagation. The waves of interest are those connecting the solutions  $u = 0$  (no oysters) and  $u = 1$  (thriving oyster population).

In the monostable case there exist travelling waves with arbitrarily large speed, hence a priori the oysters could spread at arbitrarily large rates. However, “most” solutions, in particular those starting from compactly supported initial data (corresponding to a well-defined bank of oysters), select the velocity  $c_{\text{mono}} = 2\sqrt{D}$ , which is also the minimal speed among the everywhere positive travelling waves.

In the bistable case there is a unique travelling wave; it has velocity  $c_{\text{bi}} = (1 - 2a)\sqrt{\frac{D}{2}}$ . Hence for  $a < \frac{1}{2}$  the oysters spread, i.e.  $u = 1$  “invades”  $u = 0$ . Clearly  $c_{\text{mono}} < c_{\text{bi}}$  for any  $a \in [0, 1/2)$ . Although this information has limited value since we have ignored implicit scalings in the argument, the idea is that the oysters spread more rapidly on a hard than on a soft substrate. With these differences in mind we now turn to a more detailed model which incorporates both oysters and larvae.

**4.2. An oyster-larvae model.** We consider a model which takes into account two stadia in the life cycle of an oyster with obviously different dynamic capabilities, namely oysters which are fixed to the seabed and larvae which float around in the sea. We thus disregard (or assume insignificant) the fact that oysters may detach from the seabed and move to more favourable grounds. Of course there are many other features that we do not include in our model either.

From Section 3 we pick up the discussion at the continuous-time system of equations (21-22). We briefly return to the dimensional variables  $\mathcal{O}$  and  $\mathcal{L}$  for oysters and larvae. Assuming that the larvae diffuse (with diffusion constant  $D$ ), die at rate  $E$  (per unit time) and that (as in Section 3) each oyster produces  $F$  larvae per unit time, we obtain the equation (analogous to equation (1), with a particular choice of  $Q$ )

$$(26) \quad \mathcal{L}_t = D\mathcal{L}_{xx} - E\mathcal{L} + F\mathcal{O}.$$

Here subscripts denote partial derivatives. For simplicity (and since we are going to look at travelling waves anyway) we take into account only one spatial dimension.<sup>4</sup>

The more interesting part of the model is the choice of the nonlinearity  $G$  which describes the transition of larvae to oysters. We assume that the increase in oyster population is proportional to the amount of larvae (with proportionality constant  $A$ ), and that the growth saturates when the oysters reach some maximal density

---

<sup>4</sup>As explained in §3.3, this equation represents a smoothed version of the discrete-time equation of the previous section. Assuming a reference time scale of a year, the coefficients  $D$  and  $E$  in this equation are the natural coefficients associated with diffusion and death of the larvae, per year. The coefficient  $F$ , on the other hand, should be viewed as the production of larvae, per oyster, averaged over a year.

$\mathcal{O}_{sat}$  due to competition for food and/or space (which makes it harder for larvae to settle). Additionally, we include a death rate  $C$ . This leads to

$$(27) \quad \mathcal{O}_t = A\mathcal{L}(1 - \mathcal{O}/\mathcal{O}_{sat}) - C\mathcal{O}.$$

Since the oysters are immobile there is no diffusion term. Alternatively, when the sea bed is sandy, the larvae prefer to settle on existing oysters (dead or alive), which we model by

$$(28) \quad \mathcal{O}_t = A\mathcal{L}(\mathcal{O}/\mathcal{O}_{sat} + \delta)(1 - \mathcal{O}/\mathcal{O}_{sat}) - C\mathcal{O}$$

with  $0 \leq \delta \ll 1$  a (dimensionless) measure for the relative preference of larvae to settle on a soft compared to a hard substrate. We note that the choice  $\delta = 0$  prevents spreading of oysters to previously unoccupied territory: since equation (28) contains neither diffusion nor convection,  $\mathcal{O}(x, 0) = 0$  implies  $\mathcal{O}(x, t) = 0$  for all  $t \geq 0$ .

For (27), in combination with (26), the trivial equilibrium  $(\mathcal{L}, \mathcal{O}) = (0, 0)$  is unstable provided  $AFE^{-1} < C$ , while for (28) it is stable provided  $\delta AFE^{-1} < C$  (the inequalities have an obvious interpretation). In the following we will assume that

$$(29) \quad \delta < \frac{CE}{AF} < 1.$$

The situation is thus very similar to the comparison between the monostable and bistable cases for the scalar equation in §4.1.

In true study group spirit, some educated guesses for the parameters are

$$(30) \quad \begin{array}{llll} A : & 10^{-5} \text{ y}^{-1} & \mathcal{O}_{sat} : & 10^2 \text{ m}^{-2} \\ D : & 10^{-1} \text{ m}^2 \text{ s}^{-1} & E : & 3 \cdot 10^1 \text{ y}^{-1} \end{array} \quad \begin{array}{ll} C : & 10^{-1} \text{ y}^{-1} \\ F : & 10^7 \text{ y}^{-1} \end{array} \quad \delta : \quad 10^{-2}$$

The death rates  $C$  and  $E$  follow from the life span of the oysters and of the larvae; the maximal oyster density  $\mathcal{O}_{sat}$  is estimated from existing oyster banks; the diffusion coefficient  $D$  was estimated in Section 3 (and then called  $(D + D_T)$ ); the larvae production per oyster per year  $F$  and the ratio  $\delta$  were estimated by experts from the Animal Sciences Group. The larvae-to-oyster transformation rate  $A$  (under optimal conditions) is difficult to estimate. The second inequality in (29) implies the bound  $A > 3 \cdot 10^{-7} \text{ y}^{-1}$ ; we chose the value  $10^{-5} \text{ y}^{-1}$  to accommodate this inequality.

Introduce the dimensionless variables

$$\tilde{u} = E(\mathcal{O}_{sat}F)^{-1}\mathcal{L}, \quad \tilde{v} = \mathcal{O}_{sat}^{-1}\mathcal{O}, \quad \tilde{x} = \sqrt{E/D}x, \quad \tilde{t} = Et,$$

and the dimensionless parameters

$$\alpha = AFE^{-2} \quad \text{and} \quad \beta = CE(AF)^{-1}.$$

(The parameter  $\alpha$  is closely related—approximately equal—to the combined parameter  $\varepsilon\Lambda$  of the previous section. It is the growth rate of the system without diffusion, without oyster morbidity, without taking crowding into account, and with  $\delta = 0$ ). After dropping the tildes from the notation we obtain

$$(31) \quad \begin{cases} u_t = u_{xx} - u + v, \\ v_t = \alpha[u(v^k + k\delta)(1 - v) - \beta v]. \end{cases}$$

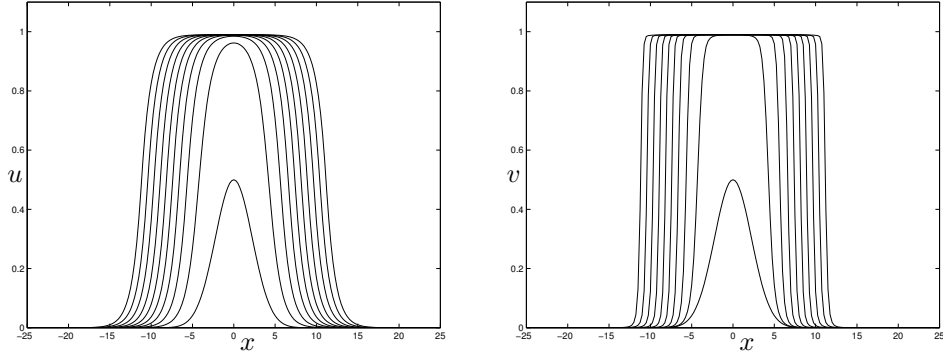


FIGURE 4. A solution of (31) with  $k = 1$ ,  $\alpha = 1$ ,  $\beta = 10^{-2}$ ,  $\delta = 10^{-3}$  showing the development towards travelling waves.

Here  $k = 0$  or  $k = 1$ , corresponding to a hard and a soft substrate respectively. We recall that we assume  $0 \leq k\delta < \beta < 1$  (which is satisfied if the estimated values (30) are approximately correct, namely  $\beta = 3 \cdot 10^{-2} \ll 1$ ).

As for the FK equation in §4.1 we expect the long term dynamics to be dominated by travelling waves and this is confirmed by numerical simulations, see Figure 4.

Our discussion of (31) now follows the lines of that of the FK equation. For the hard substrate ( $k = 0$ ) we look at the asymptotic case  $\beta = 0$ . The trivial state is unstable and we can expect there to be a one parameter family of travelling waves, one of which is selected by sufficiently localised initial data. The expected asymptotic velocity  $c_h$  can be calculated explicitly; here we did so by locating the value of  $c$  for which two eigenvalues coincide. Under the condition that  $c$  should be real, this value is unique. The result is depicted in Figure 5 as a function of the parameter  $\alpha$ . In the limits of small and large  $\alpha$  the behaviour is

$$c_h \sim \sqrt{\frac{27}{4}}\alpha \quad \text{as } \alpha \rightarrow 0, \quad c_h \sim \sqrt[4]{\frac{27}{4}}\alpha^{1/4} \quad \text{as } \alpha \rightarrow \infty.$$

Numerical calculation show that the speed  $c_h$  thus calculated analytically is indeed the selected wave speed.

For the soft substrate ( $k = 1$ ) the limit case  $\beta \rightarrow 0$  also implies  $\delta \rightarrow 0$  because of the requirement  $\delta < \beta$ . However, it is not clear that this limit is well-defined. Therefore, for the moment we will use the values  $\beta = 3 \cdot 10^{-2}$  and  $\delta = 10^{-2}$  which follow from (30). In any case it is impossible to calculate the (unique) wave speed analytically, so we have to rely on numerical computations. The asymptotic speed  $c_s$  is depicted in Figure 5, again as a function of the parameter  $\alpha$ .

**4.3. Discussion.** For  $\alpha = 10^{-1}$ , the value corresponding to (30), we have  $c_h = 0.2$  and  $c_s = 0.01$ . These numbers can be interpreted in two ways:

- A speed  $c = 0.2$  in dimensionless variables corresponds to  $0.2\sqrt{ED} \approx 2 \cdot 10^3$  meter per year (which seems rather fast). This confirms that the estimates

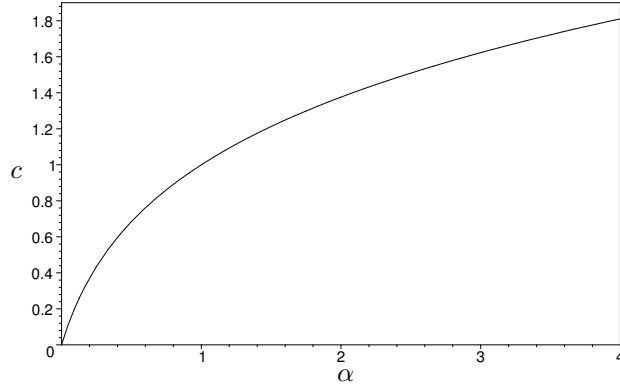


FIGURE 5. The wave speed  $c_h$  on a hard substrate as a function of  $\alpha$  (with  $\beta = 0$ ).

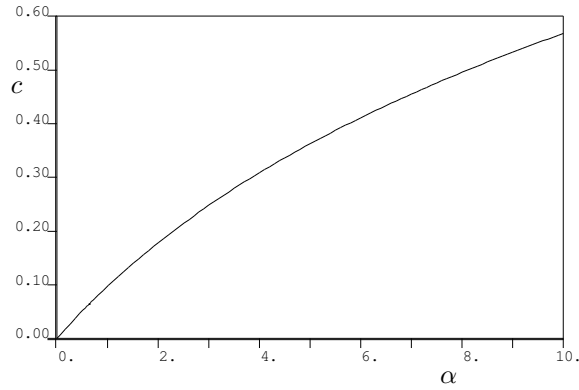


FIGURE 6. The wave speed  $c_s$  on a soft substrate as a function of  $\alpha$  (with  $\beta = 3 \cdot 10^{-2}$  and  $\delta = 10^{-2}$ ).

of (30) may be rather inaccurate, and that  $\alpha$  may well differ significantly from  $10^{-1}$ .

- On the other hand, the two numbers  $c_h$  and  $c_s$  suggest that an oyster bank spreads about twenty times faster on hard substrate than on soft substrate. Despite the inaccuracy in the coefficients  $\alpha$ ,  $E$ , and  $D$ , the linear behaviour of  $c_h$  and  $c_s$  for small  $\alpha$  (see Figs. 5 and 6) implies that the ratio

$$c_h : c_s \approx 20 : 1$$

remains valid as long as the parameters are such that the dimensionless parameter  $\alpha = AFE^{-2}$  is not too large (roughly  $\alpha < 1$ ).

While this biological conclusion is relatively easily drawn, many mathematical niceties/issues are so far unresolved, partly due to time limitations. In particular, the asymptotic regime  $\beta \rightarrow 0$  for the soft substrate model has not been analysed



in depth. Besides, the limit of large  $\alpha$  would be interesting to study from a mathematical point of view.

### 5. Influencing the population by reduction of salinity

The possibility of lowering the salinity of all or part of the Eastern Scheldt by once again allowing freshwater from local rivers to drain into the basin has been proposed as a method to hinder oyster population growth. It should be noted that for *raising* salinity levels (by for example turning compartments into lagoons by building more dams) to have an impact on larval production or mortality, or on oyster mortality, the salinity level would have to be taken above the tolerance of many other indigenous organisms that inhabit the bottom of the estuary, collectively called benthics.

Specifically, we have already seen (p. 27) that oyster fecundity is a function of salinity, both because oysters produce fewer larvae in less salty water and because oyster larvae have higher mortality (proportion of population succumbing to disease) in less salty water [8]. However, other species in the Eastern Scheldt are also sensitive to salinity levels, so for example a stratagem involving closing the storm surge barrier and flooding the Eastern Scheldt with freshwater would doubtlessly remove the vast majority of oysters, but would also presage an ecological disaster. Furthermore, since the freshwater dams were constructed between 1964 and 1987, agriculture upstream of the dams has become dependent on the current river conditions. This is in addition to the still pertinent reason for building the dams: flood prevention. With these facts in mind, a permanent reopening of the dams would not be without its side effects. In this section we will therefore consider *seasonal salinity reduction*, or *SSR*, in which freshwater would be allowed into the Eastern Scheldt only in a period coinciding with the spawning season and the life span of the larvae. In addition, and for reasons to be explained below, only SSR in the Eastern Compartment (Kom) of the Eastern Scheldt will, see figure 2, be considered. This implies that the dams in other compartments would remain intact, and the consequences for agriculture and flood prevention would be minimised.

**5.1. The Eastern Compartment.** The two locations for direct freshwater input are The Northern and Eastern Compartments. The Kom was chosen for analysis for several reasons. Important among these is the canal which now runs along the Eastern bank of the compartment. It splits into two canals: the Schelde-Rijn-kanaal for shipping and the Bathse Spuikanaal, both fed by the Zoommeer, a lake bordering on the Eastern Scheldt, from which it is assumed large volumes of water can be pumped due to the capacity of the lake. Agriculture is believed to be more dependent on the rivers which feed the Northern Compartment. Furthermore, the Kom is the shallowest compartment and has the most regular topography and is thus amenable to a simplified analysis.

Prior to construction of the Kom dam, RIKZ measured [5] the influx of freshwater into the Kom at between 50 and 70 cumecs (cubic metres per second). As a seasonal measure, freshwater flowing through sluice gates in the dam would not approach these flow rates, and so the following analysis assumes that the dam remains intact and freshwater is pumped from the canal at a rate of 100 cumecs. The

assumption is that the large body of water in the Zoommeer can provide the extra capacity, and this source can be replenished in the majority of the year when the pumps are not in operation.

**5.2. Effects of lower salinity.** The primary trigger for spawning is water temperature [8, p. 312], [4]: *C. gigas* are known to spawn when the temperature exceeds 20°C over a period of several weeks, although they are capable of spawning at 16°C. With large concentrations of gametes in the water, spawning tends to be highly coordinated with large colonies spawning almost in synchrony, to maximise chances of reproduction. Larvae develop in the water phase for between 15 and 30 days, during which time they are spread by water currents and diffusion.

The approximately 100 day period covering the spawning and larval phases is seen as the optimal time to reduce salinity as larvae are more susceptible to reduced salinity levels than oysters. Oysters can survive water as brackish as 10ppt (parts per thousand of chloride), while larvae will not develop in salinity levels below 11ppt. This is far below the tolerance of other benthics in the area and, even if it were possible, lowering the salinity to kill the oysters and/or larvae is therefore not an option. However, mortality, oyster respiration rates, and oyster filtration rates are affected by even a small change in conditions, through the following formulae quoted in [8]:

$$(32a) \quad \text{mortality} \propto \frac{S - 8}{5.5} ,$$

$$(32b) \quad \text{respiration rate} \propto 1 + \left( \frac{R_r - 1}{5} \right) (20 - S) ,$$

$$\text{where } R_r = \begin{cases} 0.0915T + 1.324 & T \geq 20^\circ\text{C} , \\ 0.007T + 2.099 & T < 20^\circ\text{C} , \end{cases}$$

$$(32c) \quad \text{filtration rate} \propto \frac{S - 10}{10} \quad \text{for } 10 < S < 20 ,$$

where  $S$  denotes salinity and  $T$  temperature. Note that oysters are filter feeders, actively pumping water through their feeding organs and filtering out phytoplankton, bacteria and protozoa for consumption. Filtration rate can be measured in millilitres per minute.

**5.3. Pumping strategies.** We introduce a time system modulo 4 (i.e. 4 corresponds to one day), with 0 at low tide, 1 at the midpoint between low and high tide, and so on, with each time phase defining a *Pumping Period* ( $PP$ ). We assume that the tidal convection is the primary mixing mechanism for fresh and salt water in the Kom, and that mixing is highly efficient: in any  $PP$ , volumes of fresh and salt water are assumed to mix fully. Though this assumption may not hold in all circumstances, in a relatively shallow basin such as the Kom, with a tidal difference of approximately 3 metres, temperature and salinity stratification, for example, are perhaps unlikely.

As discussed above, we normalise on a pumping rate of 100 cumecs by supposing that the capacity of the canal and lake permit this rate of continual supply. Thus

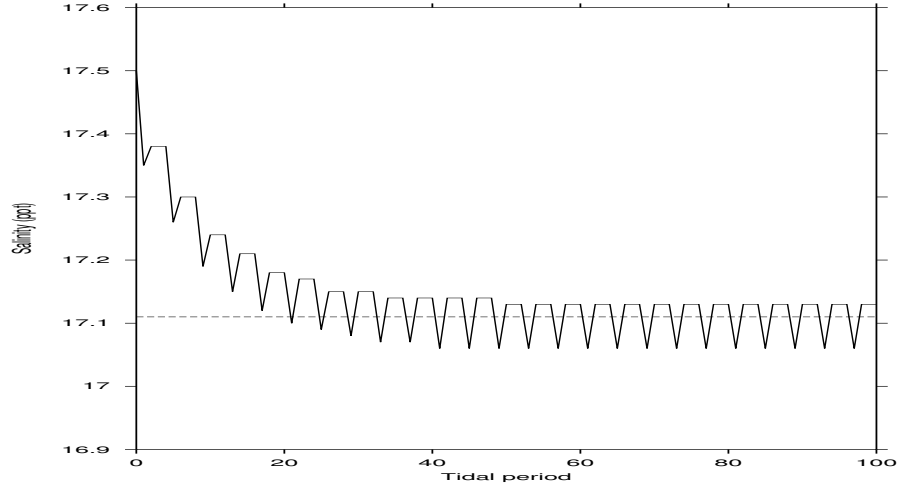


FIGURE 7. Strategy 0400

Concept	Symbol	Value
Time step	$t$	—
Volume at $t = n$	$V_n$	—
Global salinity at $t = n$	$\rho_n$	—
Weighting at $t = n$	$W_n$	Depends on strategy (mod 4)
Volume of freshwater influx in one PP	$V_f$	1.08
Freshwater salinity	$\rho_f$	0.5
Sea water salinity	$\rho_s$	17.5
Water volume in Kom at high tide	$V$	600
Water volume in Kom at low tide	$v$	400

TABLE 1. Table of concepts and data

a normalised freshwater input would be 1 unit per PP, and we can generate  $2^4$  pumping strategies within the constraint of a total of 4 units pumped during the four PPs. We shall be using the concepts, symbols and values listed in table 1, where volume is in units of  $10^6 \text{m}^3$ , and salinity is in ppt. The data were provided by RIKZ.

The formula we use to calculate the mass of salt in the Kom at time  $n + 1$  is the following:

$$(33a) \quad V_{n+1}\rho_{n+1} = V_n\rho_n + W_{n+1}V_f\rho_f + (V_{n+1} - V_n - W_{n+1}V_f)M_{n+1} ,$$

$$(33b) \quad \text{where } M_{n+1} = \begin{cases} \rho_s & \text{if } V_{n+1} - V_n - W_{n+1}V_f > 0. \\ \rho_n & \text{if } V_{n+1} - V_n - W_{n+1}V_f < 0. \end{cases}$$

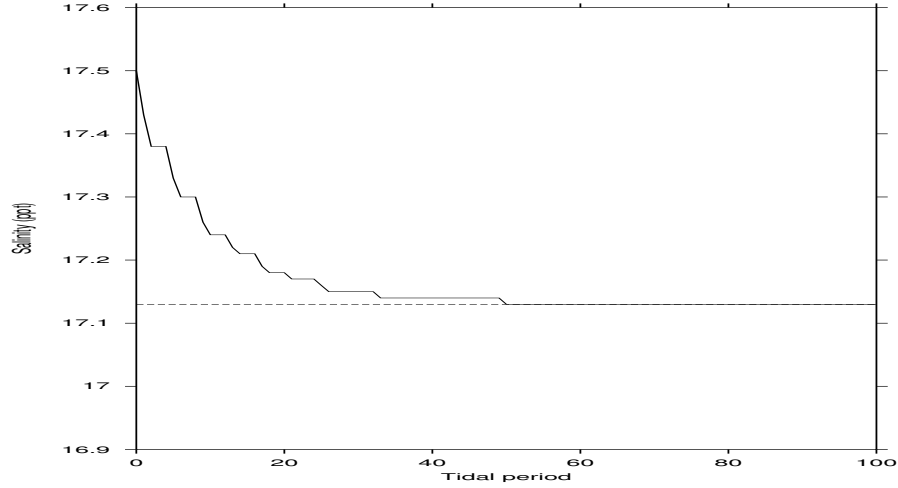


FIGURE 8. Strategy 0220

The origin of this formula is clear when it is considered term by term:

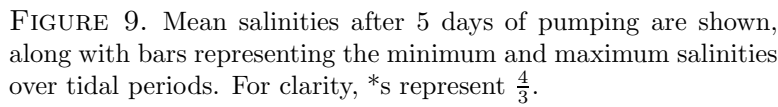
$$\begin{aligned}
 (34) \quad & V_{n+1}\rho_{n+1} && \text{— mass of salt at } t = n + 1; \\
 & V_n\rho_n && \text{— mass of salt at } t = n; \\
 & W_{n+1}V_f\rho_f && \text{— mass of salt from freshwater input;} \\
 & V_{n+1} - V_n - W_{n+1}V_f && \text{— saltwater volume change due to tidal flux.}
 \end{aligned}$$

The two cases in (33b) correspond to a rising and falling tide, respectively.

**5.4. Results and Discussion.** Figures 7 & 8 show SSR for two different pumping strategies. The periodic behaviour apparent here after approximately 40 time steps is a common feature of the 16 strategies considered, and suggests that pumping need only begin 5 days before spawning is predicted to start. In these two examples, the baseline (that is, mean over tidal periods) salinities after 5 days are close to each other in value. Figure 9 shows a comparison in baseline salinities for all 16 strategies, and it is clear that there is not much to choose between them in terms of baseline salinities. It is very likely that each strategy has a different engineering cost depending on the rate of pumping required, and also that some strategies may place too great a strain on water supply, and that these may become the dominant factors in deciding strategy. All strategies recover to the original baseline salinity of 18 ppt within 5 days of the freshwater supply being cut off.

It should be noted that RIKZ has conducted numerical simulations to study the effects on salinity of a *permanent* reopening of the compartment dams in both the Northern Compartment and the Eastern Compartment ([5]). The long-time maximum baseline salinity they predicted for the Kom is within the range covered by our SSR strategies.

We conclude by noting that SSR in the Kom has the advantages of maintaining the flood protection afforded by the compartment dams, of providing a short-term,



The above analysis rests on many simplifications of the system, and any final recommendation on the viability of SSR would have to rest on a more complete numerical model similar to that used by RIKZ in their long-time simulations. However, we believe that the data obtained gives reasonable estimates for the likely effects of SSR in the Kom, and it seems clear that the salinity cannot be reduced sufficiently to make this approach feasible.

## 6. Outline of a large-scale simulation

For a more detailed answer to the questions of the Introduction we propose the use of large-scale computer simulation to reconstruct and to predict the development of the oysters.

**6.1. How are the oyster clusters formed?** In the following we give a sketch of the life cycle of a Japanese oyster. Most data are taken from the report [7].

- 1 eggs (in July-August an adult oyster produces between  $10^6$  and  $10^8$  eggs);
- 2 larvae (an egg together with sperm develop into a larvae within 1 day);
- 3 veliger larvae (in one or two days the larvae develop into veliger larvae (with larval shell));
- 4 spat (in 15 to 30 days they settle down on appropriate hard substrate, bigger larvae can crawl quite a distance searching for appropriate settlement);
- 5 juvenile oysters develop into adult oysters after one year (estimated mortality of juvenile eastern oysters of more than 64% in seven days after settlement and more than 86% in one month after settlement on sub-tidal and inter-tidal plates);
- 6 death (caused by aging (it has been estimated a Japanese oyster can live approximately 20 years); the other major causes of death for oysters in general, such as predators and diseases, are not substantial in the Eastern Scheldt).

A simplified view of how a new cluster/colony arises is the following:

The larvae floating in water are displaced by diffusion and carried with the tidal movement. Within 3 weeks they have to settle down on appropriate substrates (hard surface) or otherwise die. The settling down occurs at the moment when the water is quiet (e.g., at the turn of a tide, approximately 20 minutes, or in quiet surroundings). After settling down they will stay there, however, a large percentage of them die before reaching adulthood. After one year, those surviving (juvenile) oysters will enter the reproduction system and start to produce gametes and eggs. Oyster clusters are formed or grow when (veliger) larvae settle down on appropriate substrates and survive there.

**6.2. Diffusion and convection of larvae due to tides.** Two main factors, determining the place for the larvae to settle down, are: 1) advection and diffusion of the tidal flow; and 2) to be at the appropriate place while the water is quiet. If we ignore the swimming effect of the larvae, the transport of larvae can be described by the shallow water equations,

$$(35) \quad \frac{\partial u}{\partial t} = -u \frac{\partial u}{\partial x} - v \frac{\partial u}{\partial y} - w \frac{\partial u}{\partial z} + fv - g \frac{\partial \zeta}{\partial x} + \nu_h \frac{\partial^2 u}{\partial x^2} + \nu_h \frac{\partial^2 u}{\partial y^2} + \frac{\partial}{\partial z} \left( \nu_v \frac{\partial u}{\partial z} \right)$$

$$(36) \quad \frac{\partial v}{\partial t} = -u \frac{\partial v}{\partial x} - v \frac{\partial v}{\partial y} - w \frac{\partial v}{\partial z} - fu - g \frac{\partial \zeta}{\partial y} + \nu_h \frac{\partial^2 v}{\partial x^2} + \nu_h \frac{\partial^2 v}{\partial y^2} + \frac{\partial}{\partial z} \left( \nu_v \frac{\partial v}{\partial z} \right)$$

$$(37) \quad w = -\frac{\partial}{\partial x} \left( \int_{-d}^z u dz' \right) - \frac{\partial}{\partial y} \left( \int_{-d}^z v dz' \right)$$

$$(38) \quad \frac{\partial \zeta}{\partial t} = -\frac{\partial}{\partial x} \left( \int_{-d}^{\zeta} u dz \right) - \frac{\partial}{\partial y} \left( \int_{-d}^{\zeta} v dz \right)$$

and the transport equation with  $c$  being the concentration of larvae:

$$(39) \quad \frac{\partial c}{\partial t} = -\frac{\partial uc}{\partial x} - \frac{\partial vc}{\partial y} - \frac{\partial wc}{\partial z} + \frac{\partial}{\partial x} \left( D_h \frac{\partial c}{\partial x} \right) + \frac{\partial}{\partial y} \left( \frac{\partial c}{\partial y} \right) + \frac{\partial}{\partial z} \left( D_v \frac{\partial c}{\partial z} \right)$$

Here the triplet  $(u, v, w)$  is the velocity vector,  $f$  the Coriolis parameter,  $g$  the gravitational acceleration, and  $\nu_h/\nu_v$  and  $D_h/D_v$  the horizontal and vertical viscosities and diffusion (or dispersion) coefficients. The function  $\zeta$  is the height of the water surface.

The shallow water equations governing the water movement can be simulated using the existing simulation model TRIWAQ (which is in operational use at Rijkswaterstaat [18, 16]), and the transport equation can be simulated with a particle simulation model SIMPAR (a 2-dimensional version is also in operational use at Rijkswaterstaat [2, 6], and a 3-dimensional model is currently being developed at TU Delft [13, 14]) using the flow velocity data produced by TRIWAQ. Instead of looking at the concentration of larvae, we consider the discrete (likely aggregated into groups of larvae) quantity which are ‘particles’ carried with the water movement. For that the continuous transport equation (in concentration) is transformed into a stochastic partial differential equation (e.g., Fokker-Planck equation).

With the above simulation models we can determine the position of the larvae at the time when the water is quiet. If a larva happens to be positioned above an appropriate (preferably hard) surface, then there is a certain chance that it will settle down on that surface. There exist quite detailed geometry and geological data about the Eastern Scheldt which can be used for the simulation. The geometry data (e.g. depth) are already used by TRIWAQ and SIMPAR, but we need additional geological data (hard, soft surfaces) for our simulation of the distribution of the oysters. The main problem remaining here is to simulate the life cycle of the oysters in order to determine the speed of population increase, reconstruction of the past development and prediction of future development, etc.

**6.3. Simulation of the life cycle of oysters.** In section 6.1 we have described the life cycle of an oyster. In order to obtain detailed information of the spread of the oysters we need to model the important factors which affect the development of oysters.

We may distinguish this life cycle into two major phases:

**Phase 1.** Each year in July there is a moment when oysters (older than 1 year) start to produce eggs and gametes. Then there will be billions of eggs and gametes or larvae floating in the water (initially above the spawning oyster colonies). This can be modelled as (multiple) sources in the transport model SIMPAR. The number of larvae produced by a single oyster depends on many factors: size of the oyster, environmental conditions, etc. (see the discussion in Section 3).

**Phase 2.** This is the period from the moment of settling down to developing into an adult oyster. During this period the oyster will stay in the same place except when they are very young (spat, juvenile oyster). In the latter case they can be

wiped out by strong currents. From juvenile oyster to adult oyster, we know that a large percentage will die before reaching adulthood.

In general, there are five biotic factors causing oyster death: disease, predation, competition, developmental complications, and energy depletion. And there are five abiotic factors: moisture depletion, temperature, salinity, water motion, and oxygen depletion. Disease and predation are said to have the greatest effect on oyster populations. However, as mentioned earlier, these two factors are not significantly present in the Eastern Scheldt. Which factors contribute predominantly to the death or survival of the Japanese oysters in this specific Eastern Scheldt remains to be investigated.

To conclude: many factors for the model of the life cycle of a Japanese oyster still need to be investigated, part of them can be determined by experiments, and part of them can be determined/tuned by comparing simulation results with the past observational data.

### A. Appendix: derivation of (14)

Consider equation (3) where  $u$  is given by (11). Let us rescale the space and time variables as

$$(40) \quad (X \quad Z \quad T) = (x/L \quad z/h \quad U_0 t/L),$$

where  $L$  is a characteristic scale in the  $x$ -direction. From the shallowness hypothesis, we assume that  $h = \epsilon L$ ,  $\epsilon \ll 1$ . Then, introducing the Peclet number  $P_e = \frac{U_0 h}{D}$ , equation (3) becomes

$$(41) \quad \epsilon P_e \left( \frac{\partial \mathcal{L}}{\partial T} + v(Z) \frac{\partial \mathcal{L}}{\partial X} \right) = \frac{\partial^2 \mathcal{L}}{\partial Z^2},$$

$$(42) \quad v(Z) = \frac{3}{2} \frac{(Z - Z_0)(2 - Z - Z_0)}{(1 - Z_0)^2},$$

with no-flux boundary conditions at  $Z = 0, 1$ :

$$(43) \quad \frac{\partial \mathcal{L}}{\partial Z} = 0, \quad Z = 0, 1.$$

The presence of the small number  $\epsilon P_e \ll 1$ , suggests to seek a solution in series of power of  $\epsilon P_e$  in the form

$$(44) \quad \mathcal{L} = \mathcal{L}_0(X, Z, T; \tau) + \epsilon P_e \mathcal{L}_1(X, Z, T; \tau) + \dots,$$

where  $\tau$  is a slow timescale defined by  $\epsilon P_e T$ . Inserting (44) into (41), we find at leading order that

$$(45) \quad \mathcal{L}_0(X, Z, T; \tau) = \mathcal{L}_0(\xi; \tau), \quad \xi = X - T,$$

while the solution at order  $(\epsilon P_e)^2$  is

$$(46) \quad \mathcal{L}_1(X, Z, T; \tau) = - \frac{[(-2 + Z)Z + 2Z_0(X) - Z_0^2(X)]^2}{8[-1 + Z_0(X)]^2} \frac{\partial \mathcal{L}_0}{\partial \xi}.$$



Finally at order  $(\epsilon P_e)^2$ , equation (41) reads

$$(47) \quad \frac{\partial^2 \mathcal{L}_2}{\partial Z^2} = F \left( \frac{\partial \mathcal{L}_0}{\partial \tau}, \frac{\partial \mathcal{L}_0}{\partial \xi}, \frac{\partial^2 \mathcal{L}_0}{\partial \xi^2}, Z, Z_0 \right).$$

The form of  $F$  is rather complicated and of no real interest so we do not explicitate it here. The key point is that  $\mathcal{L}_2$  has to satisfy the boundary conditions (43) and this yields the solvability condition

$$\int_0^1 F \left( \frac{\partial \mathcal{L}_0}{\partial \tau}, \frac{\partial \mathcal{L}_0}{\partial \xi}, \frac{\partial^2 \mathcal{L}_0}{\partial \xi^2}, Z, Z_0 \right) dZ = 0.$$

After integration, we obtain

$$(48) \quad \frac{\partial \mathcal{L}_0}{\partial \tau} = -\frac{24(1-Z_0)}{105} \frac{dZ_0}{dX} \frac{\partial \mathcal{L}_0}{\partial \xi} + \left( \frac{1}{P_e^2} + \frac{2(1-Z_0)^2}{105} \right) \frac{\partial^2 \mathcal{L}_0}{\partial \xi^2},$$

which, in terms of the original space and time variables  $x$  and  $t$ , is (5).

### Bibliography

- [1] J. R. Dew, A population dynamic model assessing options for managing eastern oysters (*Crassostrea virginica*) and triploid Suminoe oysters (*Crassostrea ariakensis*) in Chesapeake Bay, MS Thesis. Virginia Polytechnic Institute and State University, 2002.
- [2] M. Elorche, Vooronderzoek Particle-module in SIMONA (in Dutch). *Werkdocument RIKZ/OS-94.143x*, 1994.
- [3] A. Gangnery, C. Bacher, and D. Buestel, *Assessing the production and the impact of cultivated oysters in the Thau lagoon (Mediterranee, France) with a population dynamics model*, Can. J. Fish. Aquat. Sci. **58**, pp. 1012–1020, 2001.
- [4] P. Gouilletquer *et al*, *La reproduction naturelle et contrôlée des Bivalves cultivés en France*, IFREMER Rapport Interne DRV/RA/RST/97-11 RA /Brest, 1997.
- [5] H. Haas and T. Tosserams, *Balanceren tussen zoet en zout en Ruimte voor veerkracht en veiligheid in de Delta*, Rapporten RIKZ/2001.18 en RIZA/2001.014.
- [6] A. W. Heemink, Stochastic Modeling of dispersion in shallow water. *Stochastic Hydrol. Hydraul.* 4, pp. 161–174, 1971.
- [7] B.J. Kater, Japanse oesters in de Oosterschelde: ecologisch profiel, RIVO Report (in Dutch), April 2003.
- [8] M. Kobayashi, *et al*, *Aquaculture* **149**, pp. 285–321, 1997.
- [9] R. Mann, & D. A. Evans, Estimation of oyster, *Crassostrea virginica*, standing stock, larval production, and advective loss in relation to observed recruitment in the James River, Virginia, *J. Shellfish res.*, 17(1), pp. 239–253, 1998.
- [10] J. D. Murray, *Mathematical biology; an introduction*.
- [11] J. R. Ockendon, S. D. Howison, A. A. Lacey, and A. B. Movchan, *Applied Partial Differential Equations*, Oxford University Press 1999
- [12] D. B. Quayle, Pacific oyster culture in British Columbia, *Fish. Res. Board. Can. Bull.*, 169, pp 1–192, 1969.

- [13] J. W. Stijnen & H. X. Lin, The Modeling of Diffusion in Particle Models, Project Report to National Institute for Coastal and Marine Management (RIKZ), Contract RIKZ/OS 2000/06080, 14 p., September 2000.
- [14] J. W. Stijnen, A. W. Heemink & H. X. Lin, An Efficient 3D Particle Transport Model for Use in Stratified Flow *to be published*..
- [15] G. I. Taylor, *Dispersion of soluble matter in a solvent flowing slowly through a tube*, Proc. Roy. Soc. **A210**, pp. 186-203.
- [16] E. A. H. Vollebregt, Parallel Software Developemnt Techniques for Shallow Water Models, Ph.D. Thesis, Delft Unversity of Technology, 1997.
- [17] T. Yanagi, *A simple method for estimating ...*, see <http://data.ecology.su.se/MNODE/Methods/YanagiMixing/Yanagi.htm>.
- [18] M. Zijlema, TRIWAQ - three-dimensional incompressible shallow flow model, *Technical Documentation*, RIKZ/Rijkwaterstaat, 1997.

## CHAPTER 3

# Non-Newtonian Effects on Ink-Jet Droplet Formation: A Mathematical Model

Huaxiong Huang

Other participants: Frits Dijkman, Paul Duineveld, Fieke Dekkers,  
 Johan Dubbeldam, Yodi Gunawan, Simon Kronemeijer, Mikhail Tchesnokov,  
 Jan Bouwe van den Berg, Rein van de Hout, J.F. Williams

**ABSTRACT.** In this report we present a simple mathematical model for the formation and possible breakup of ink-jet droplets. The formation of the droplet is based on an energy argument. The evolution of the droplet tail is modelled by a one-dimensional axis-symmetric non-Newtonian liquid filament. A simple power-law formula is used for the elongational viscosity. Numerical simulations are presented and effect of the elongational viscosity on the tail evolution and breakup is investigated.

**KEYWORDS:** fluid mechanics, viscosity, rheology

## 1. Introduction

Light-emitting polymer display (LED) is a new technology. LED devices do not have viewing angle restriction, which gives them a major advantage over the existing liquid crystal displays (LCDs). The active material in the display is a very thin semi-conducting polymer layer of order 100 nanometer. To obtain these thin layers a small concentration of the polymer is dissolved in a suitable solvent. Different colors can be obtained with different polymers. To make a full color display, red, green and blue polymer solutions have to be applied in pixels of typically  $66 \times 200$  micron. The method with which the polymer solutions are applied is by means of ink-jet printing. Individual droplets are printed in the pixels and by evaporating the solvent the final polymer layer is obtained. The polymers have high molecular weights which cause the droplet formation to be quite different from an ordinary Newtonian liquid, i.e. a long filament may form. This can give rise to a decrease in the placement accuracy of the droplets on the substrate. Therefore, in predicting predicting the behaviour of a droplet in an ink-jet printer the material parameters of the liquid are very important.

The viscosity of an ink is an important parameter for the droplet formation in an ink-jet head. Standard in the ink-jet printing world is to measure the shear viscosity of the liquid. Most common inks are Newtonian liquids and the shear viscosity is a suitable characterization parameter. For inks that are solutions of a

high molecular weight polymer in small concentrations in a solvent, the situation is quite different. It has been observed in the laboratory that the droplet formation of these solutions is considerably different from the predictions based on the shear viscosity alone, i.e., assuming that the solution is a pure Newtonian liquid. For example, more energy is needed to eject the droplet and some droplets are formed with a filament, which can break up in satellite droplets. When these solutions are measured in a shear rheometer, the viscosity is shown to be a constant and is in the proper regime for ink-jet printing. This suggests that the filament formation during the ink-jet printing is caused by the different behaviour in the elongational viscosity. This is not surprising since it is well known (see for example [10] and references therein) that a small concentration of a high molecular weight polymer in a solvent can increase the elongational viscosity substantially.

Since the shear viscosity cannot be used to characterize the inks, we need to obtain the elongational viscosity in order to make reasonable predictions. The problem is that the elongational viscosity is not easy to measure, especially in the range of deformation rate of ink-jet printing, in contrast to the shear viscosity. On the other hand, the experimental results indicate that the tail behaviour is strongly affected by the properties of the polymer. This suggests that if a mathematical model of the non-Newtonian fluid can be constructed, then the model can be calibrated using the tail-length or other information of the droplets by varying the operational conditions. This approach can be viewed as an indirect mean to obtain information on the elongational viscosity. Obviously, the same model can also be used to study the dynamic process of droplet formation. In this report, we describe some preliminary results based on a one-dimensional model for the elongational viscosity and for the dynamics of droplet formation.

The rest of the report is organized as follows. In section 2, we describe a mathematical model for the dynamics of droplet formation. The model is one-dimensional in space but highly non-linear, therefore, numerical means are used to obtain the solution, which are given in section 3. In section 4, we present main numerical results which show that the length of the droplet tail is affected by a parameter in our elongational viscosity model. Finally, in section 5 we discuss the limitation of our model and possible ways of improving it.

## 2. Mathematical Model

The problem of droplet formation consists of two stages: the ejection of liquid from the nozzle and the breakup liquid filament. For an ink-jet printer, this is achieved using a large number of miniature valveless pumps. The pumps are actuated electronically, by heating elements of the liquid locally to high temperatures. Vapor bubbles are created and grow explosively at desirable locations inside pumps. As a result, they push a small amount of liquid at high velocity through nozzles. To completely understand the formation of an ink-jet droplet, detailed two (possibly three) dimensional hydrodynamics models are needed. However, the exit flow of general non-Newtonian fluid is still far from being understood and simple models are used instead. For example, the process has been investigated using an energy argument while the behaviour of the pumps is modeled as a Helmholtz resonator in [4]. Since our main purpose is to study the non-Newtonian effects on the filament

(tail) of the droplet, we will not investigate the exit flow in detail. Instead, we will follow the energy argument in [4] for the formation of droplets.

The breakup of a liquid jet in the context of Newtonian fluid has been well-studied by several researchers based on (one-dimensional) thin jet approximation. Such models predict the breakup of jets in finite time, described by a similarity solution [6, 7]. It is well-known that visco-elasticity and other non-Newtonian effects delay the breakup by suppress the surface tension [3, 8, 9, 11]. Nevertheless, pinch-off may still occur due to the surface tension force, which becomes dominant when the radius of the liquid jet is small. The problem being investigated here, however, has some distinct features. The ink-jet is driven by thermal expansion of the liquid (or a piezo-electric mechanism) with a period on the order of 20 microseconds [4]. Therefore, the breakup of the ink-jet may be caused by the highly oscillatory nature of the flow (at least initially). Our main objective is to set up a mathematical model and investigate numerically the combined effect of the highly frequency transient flow and the non-Newtonian nature of the polymer solution.

**2.1. Initial droplet formation.** In order to provide the initial condition for the evolution of the droplet tail, we use a simple energy argument to obtain the time  $t_0$  when the droplet forms. The basic argument is based on energy consideration, following that of [4]. We assume that before the droplet forms, the fluid simply flows out of the nozzle as a liquid cylinder with radius  $R_n$ . At  $t_0$ , the droplet forms and it moves with the fluid velocity at the nozzle,  $V(t_0)$ . Thus, we can estimate the total kinetic energy flowing out of the nozzle during the period  $0 \leq t \leq t_0$  as

$$T_{k.e.} = \int_0^{t_0} \frac{1}{2} \rho A V^3 dt$$

where  $A$  is the nozzle cross section area,  $\rho$  is the fluid density and  $V$  is the fluid velocity at the nozzle exit, assumed to be given. The total kinetic energy associated with the liquid cylinder at  $t_0$  is

$$T_{k.e.l.} = \int_0^{t_0} \frac{1}{2} \rho A V dt V^2(t_0).$$

The total surface energy associated with the liquid cylinder at  $t_0$  is

$$T_{s.e.l.} = \sigma 2\pi R_n \int_0^{t_0} V dt$$

where  $\sigma$  is the surface tension coefficient between the liquid and the air. We assume that the energy dissipation due to viscosity is negligible, then the drop forms at  $t_0$  when  $T_{k.e.l.} + T_{s.e.l.}$  is about to exceed  $T_{k.e.}$ , i.e., the total kinetic energy is not sufficient to maintain the growth of the liquid cylinder out of the nozzle beyond  $t_0$ .

In addition, we assume that after the droplet forms, it immediately takes a spherical shape and is connected with the nozzle by a tail, in the shape of a cylinder. The evolution and the eventual breakup of the tail is the subject of the rest of this report.

**2.2. Tail evolution.** The governing equations for the fluid inside the tail of the droplet are the conservation of momentum and incompressibility equations

$$\begin{aligned}\rho \left[ \frac{\partial \vec{v}}{\partial t} + (\vec{v} \cdot \nabla) \vec{v} \right] &= \nabla \cdot \mathbf{T} - \nabla p \\ \nabla \cdot \vec{v} &= 0,\end{aligned}$$

where  $\rho$  is the density,  $\vec{v}$  is the velocity,  $\mathbf{T}$  is the viscous stress tensor and  $p$  is the pressure. Here we have neglected the effect of the gravity.

To simplify the discussion, we make the following assumptions:

- (i) The flow is essentially one-dimensional since the tail radius is small compared to its length  $L$ . Furthermore, in general the variation of the tail radius is also small, except during the initial stage of droplet formation and when tail breaks up (pinch-off occurs).
- (ii) The formation of the initial droplet and subsequently its tail are determined by the flow at the exit of the nozzle. We assume that the flow-rate is a periodic function of time given by  $Q = A_n V$ ,  $V = V_n e^{-\beta t} \sin(\omega t)$ , where  $A_n$  is the nozzle cross section area,  $V$  is the velocity at the nozzle exit,  $V_n$  is the (undamped) mean velocity at the nozzle,  $\beta$  is the damping rate due to energy dissipation of the fluid before it reaches the exit and  $\omega$  is the frequency of the oscillation.
- (iii) We assume that the droplet forms at  $t_0$  with an initial tail length  $L_0$ . The droplet is assumed to be a sphere with a fixed radius  $R_d$ .
- (iv) The tail is of cylindrical shape with radius  $R$  which is both a function of time  $t$  and spatial coordinate  $x$ , measured from the nozzle. Upon exit from the nozzle, the radius  $R_c$  of the fluid volume contracts and its value is assumed to be known and equals to the tail radius.
- (v) Since  $\omega$  is large, and the break-up of the tail normally occurs within a short time period after the droplet forms, temperature effects on both the surface tension coefficient  $\sigma$  and fluid (elongational) viscosity  $\hat{\mu}$  are negligible.
- (vi) Finally, we assume that the fluid is incompressible with a constant density.

The governing equations for the evolution of the tail and the motion of the droplet can now be re-written as<sup>1</sup>

$$\begin{aligned}\frac{\partial A}{\partial t} + \frac{\partial Au}{\partial x} &= 0, \\ \rho \left( \frac{\partial Au}{\partial t} + \frac{\partial Au^2}{\partial x} \right) &= \frac{\partial}{\partial x} \left( A\tau + A \frac{\sigma}{R} \right),\end{aligned}$$

for  $0 \leq x \leq L(t)$ , where  $A = 2\pi R^2$  is the cross section area of the tail and  $\tau$  is the axial viscous stress. For Newtonian fluid, we have

$$\tau = 3\mu \frac{\partial u}{\partial x}$$

---

<sup>1</sup>The set of equations are based on the assumption that the problem is one-dimensional. It can also be derived formally using asymptotic argument, similar to the procedure outlined in [5].

where the constant  $\mu$  is the shear viscosity. For a non-Newtonian fluid, the viscous stress is no longer linearly proportional to the strain rate. In this study, we use a simple model in the form

$$\tau = \hat{\mu} \frac{\partial u}{\partial x} + kw, \quad \frac{\partial w}{\partial t} = \frac{\partial u}{\partial x}.$$

Here  $\hat{\mu}$  is the elongational viscosity,  $w$  are the velocity and relative displacement of the fluid in the axial direction and  $k$  is assumed to be a constant. The governing equations now become

$$\begin{aligned} \frac{\partial A}{\partial t} + \frac{\partial Au}{\partial x} &= 0, \\ \rho \left( \frac{\partial Au}{\partial t} + \frac{\partial Au^2}{\partial x} \right) &= \frac{\partial}{\partial x} \left( A\hat{\mu} \frac{\partial u}{\partial x} + Ak \int_{t_{ref}}^t \frac{\partial u}{\partial x} dt + A \frac{\sigma}{R} \right), \end{aligned}$$

where  $t_{ref}$  is some reference time.

The boundary conditions for the above equations are normally required at  $x = 0$  (exit) and  $x = L$  (when the tail joins the droplet). They are

$$\begin{aligned} x = 0 \quad : \quad R(0, t) &= R_c; \quad u(0, t) = \frac{Q}{2\pi R_c^2}, \\ x = L(t) \quad : \quad u(L(t), t) &= \frac{dL}{dt}, \quad A\hat{\mu} \frac{\partial u}{\partial x} + Ak \int_{t_{ref}}^t \frac{\partial u}{\partial x} dt + A \frac{\sigma}{R} = F. \end{aligned}$$

The motion of the droplet is given by the Newton second law of motion

$$F = -ma, \quad a = \frac{d^2 L}{dt^2}.$$

We note that the boundary condition on  $R$  is not required at  $x = L$  since the velocity there is positive and the characteristic goes out of the domain. However, the flow (velocity) at  $x = 0$  may reverse its direction and in that case the radius should not be given at  $x = 0$ . When the velocity is negative, it is likely that the contact line may move inside the nozzle and the flow pattern becomes quite complicated. Therefore, the artificial boundary condition for  $R$  at  $x = 0$  for negative velocity is unphysical. A possible remedy will be to allow the exit boundary to move with the fluid when the velocity is negative, without considering a two-dimensional model. However, this issue is not considered in the present study.

### 3. Numerical Method

We now describe the numerical method used for solving the model equations derived earlier.

#### 3.1. Coordinate transformation. We first define a one-dimensional map

$$\xi = \frac{x}{L(t)}, \quad t = t.$$

Under this transformation, we have

$$\begin{aligned}\frac{\partial}{\partial x} &= \frac{1}{L} \frac{\partial}{\partial \xi}, \\ \frac{\partial}{\partial t} &= \frac{\partial}{\partial t} - \frac{\xi \dot{L}}{L} \frac{\partial}{\partial \xi}.\end{aligned}$$

The governing equations become

$$\begin{aligned}\frac{\partial R}{\partial t} &= \frac{\xi \dot{L} - u}{L} \frac{\partial R}{\partial \xi} - \frac{R}{2L} \frac{\partial u}{\partial \xi}, \\ \frac{\partial u}{\partial t} &= \frac{\xi \dot{L}}{L} \frac{\partial u}{\partial \xi} + \frac{1}{\rho R^2 L} \frac{\partial}{\partial \xi} \left( \frac{\hat{\mu} R^2}{L} \frac{\partial u}{\partial \xi} + k R^2 w + \sigma R \right), \\ \frac{\partial w}{\partial t} &= \frac{1}{L} \frac{\partial u}{\partial \xi}\end{aligned}$$

where  $\dot{L} = \frac{dL}{dt}$ . The boundary and initial conditions are

$$\begin{aligned}\xi = 0 &: u(0, t) = V, \quad R(0, t) = R_c, \\ \xi = 1 &: u = \dot{L}, \quad \frac{\hat{\mu}}{L} \frac{\partial u}{\partial \xi} + kw + \frac{\sigma}{R} = -\frac{M}{\pi R^2} a\end{aligned}$$

where  $a = \frac{d^2 L}{dt^2}$  and  $M = \frac{4\pi R_c^3 \rho}{3}$  are the acceleration and the mass of the droplet, respectively.

**3.2. Finite difference discretization.** We set up a uniform grid  $\xi_j = jh$  for  $j = 0, 1, \dots, N$  with  $h = 1/N$  and define  $U_j(t)$  and  $W_j(t)$  as approximations of  $u(\xi_j, t)$  and  $w(\xi_j, t)$ . Applying the finite difference approximation in  $\xi$  and using short-hand notation for time differentiation yields

$$\begin{aligned}\dot{R}_j &= [1 + \text{sign}(\xi_j \dot{L} - U_j)] \frac{\xi_j \dot{L} - U_j}{2L} \frac{R_{j+1} - R_j}{h} \\ &\quad + [1 - \text{sign}(\xi_j \dot{L} - U_j)] \frac{\xi_j \dot{L} - U_j}{2L} \frac{R_j - R_{j-1}}{h} \\ &\quad - \frac{R_j}{2L} \frac{U_j - U_{j-1}}{h}, \\ \dot{U}_j &= [1 + \text{sign}(\dot{L})] \frac{\xi_j \dot{L}}{2L} \frac{U_{j+1} - U_j}{h} + [1 - \text{sign}(\dot{L})] \frac{\xi_j \dot{L}}{2L} \frac{U_j - U_{j-1}}{h} \\ &\quad + \frac{1}{\rho R_j^2 L h} \left( \frac{\hat{\mu}_{j+1/2} R_{j+1/2}^2}{L} \frac{U_{j+1} - U_j}{h} + k R_{j+1/2}^2 W_{j+1/2} + \sigma R_{j+1/2} \right. \\ &\quad \left. - \frac{\hat{\mu}_{j-1/2} R_{j-1/2}^2}{L} \frac{U_j - U_{j-1}}{h} - k R_{j-1/2}^2 W_{j-1/2} - \sigma R_{j-1/2} \right), \\ \dot{W}_{j+1/2} &= \frac{1}{L} \frac{U_{j+1} - U_j}{h}, \\ \dot{W}_{j-1/2} &= \frac{1}{L} \frac{U_j - U_{j-1}}{h}.\end{aligned}$$



where

$$\hat{\mu}_{j+1/2} = \hat{\mu} \left( \frac{U_{j+1} - U_j}{Lh} \right), \quad \hat{\mu}_{j-1/2} = \hat{\mu} \left( \frac{U_j - U_{j-1}}{Lh} \right), \quad R_{j\pm 1/2} = \frac{1}{2}(R_j + R_{j\pm 1})$$

for  $j = 1, \dots, N-1$ . Note that the upwind scheme is used for the convective terms in both continuity and momentum equations.

At the boundary, we have

$$\dot{U}_0 = V_n e^{-\beta t} [\omega \cos(\omega t) - \beta \sin(\omega t)]$$

and

$$\dot{U}_N = a \equiv -\frac{\pi R_N^2}{M} \left( \frac{\hat{\mu}_N}{L} \frac{U_N - U_{N-1}}{h} + \frac{\sigma}{R_N} \right), \quad \dot{L} = U_N.$$

**3.3. Time integration and initial conditions.** The semi-discretization using finite difference approximation results in a system of (non-linear) ordinary differential equations which is solved using a standard built-in Matlab time integrator `ode23s`, starting from  $t = t_0$ . The initial value of the tail length  $L_0$  is assumed to be an arbitrary small value. The velocity for the tail and the droplet  $u(x, t_0) = V(t_0)$  and the radius of the droplet  $R_s$  can be calculated once we know  $t_0$ , the time when the droplet forms.

The value of  $t_0$  is the non-trivial solution of the following equation

$$T_{k.e.} = T_{k.e.l} + T_{s.e.l}$$

which can be solved using Maple.

## 4. Results

The parameter values for the numerical simulations are:  $V_n = 5 \text{ m} \cdot \text{s}^{-1}$ ,  $\omega = 4\pi \times 10^4 \text{ s}^{-1}$ ,  $R_n = 2.5 \times 10^{-5} \text{ m}$ ,  $\rho = 10^3 \text{ kg} \cdot \text{m}^{-3}$ , and  $\sigma = 3 \times 10^{-2} \text{ N} \cdot \text{m}^{-1}$ . We have used a power-law model for the elongational viscosity  $\hat{\mu}$  based on the experimental data in [1]

$$\hat{\mu} = \begin{cases} 3 \times 10^{-3}, & |\dot{\epsilon}| \leq 10^3 \text{ s}^{-1} \\ 3 \times 10^{-3} \left( \frac{|\dot{\epsilon}|}{10^3} \right)^\gamma, & \text{otherwise} \end{cases}$$

in Pascal. Here  $\dot{\epsilon} = u_x$  is the axial strain-rate,  $\gamma$  is a free parameter and for a Newtonian fluid,  $\gamma = 0$ . Other free parameters are the elastic constant  $k$ , initial tail length  $L_0$  and tail radius near the nozzle exit  $R_c$ .

**4.1. Initial and boundary conditions.** The initial conditions are computed based on  $t_0$ , the time when the droplet forms, which is estimated based on energy arguments. With the velocity profile at the nozzle exit  $V$  without any damping ( $\beta = 0$ ), equation (3.3) becomes

$$2 \cos^2 \omega t_0 - \cos \omega t_0 - \frac{6\sigma C}{\rho A V_n^2} = 1$$

where  $C = 2\pi R_n$  is the circumference of the liquid cylinder. Based on the parameter values given here, the non-trivial solution can be obtained using Maple as  $\omega t_0 = 2.308$ . The total volume of the droplet and the tail can be computed as

$$Vol = \int_0^{t_0} A_n V dt = \frac{A_n V_n}{\omega} (1 - \cos \omega t_0) = 1.306 \times 10^{-13}$$

in  $m^3$ . In our computation, we assume that a certain portion of it forms the spherical droplet and the rest becomes the tail. Thus we should have the following constraint on the droplet radius

$$R_s \leq \left( \frac{3Vol}{4\pi} \right)^{\frac{1}{3}} = 3.148 \times 10^{-5}$$

in  $m$ . The initial velocity of the droplet and the tail is assumed to be equal to that near the nozzle exit with radius  $R_c$ , which is

$$U_j(0) = V_c(0) \equiv \frac{R_n^2}{R_c^2} V_n \sin(\omega t_0) = 3.702 \frac{R_n^2}{R_c^2}$$

in  $m \cdot s^{-1}$ . Therefore, it depends on the radius of the initial tail radius  $R_c \leq R_n$ , which is chosen as a free parameter here. The initial length of the tail  $L_0$ , in principle can be estimated based on the choice of  $R_s$  and  $R_c$ . However, for convenience, its value is chosen as a small but non-zero value of  $10^{-5} m$  for computational purposes. Finally, the boundary condition for the velocity is given by

$$U_0(t) = V_c(t) \equiv \frac{R_n^2}{R_c^2} V = \frac{R_s^2}{R_c^2} V_n \sin(\omega t).$$

**4.2. Power-law non-Newtonian fluid.** We now present some of the computational results based on the model and the parameter values listed earlier with no elastic effect, i.e.,  $k = 0$ . In Figure 1, the evolution of the droplet tail is shown for  $R_s = 1.25R_n$ ,  $R_c = R_n$ , and  $V_c = V$ . We have also included damping in  $V$  with the damping rate  $\beta = \frac{\log(2)}{2}\omega$ . The non-Newtonian power-law parameter is chosen as  $\gamma = 1.1$ . Figure 2 shows the tail evolution for the case when the radius of the tail near nozzle exit is reduced to half of the nozzle radius, i.e.  $R_c = 0.5R_n$ . This causes the tail to break up sooner, compared to the previous case with a large  $R_c$ . Next we have investigated the effect of parameter  $\gamma$ . In Figure 3, we have the same parameter values compared to Case 1, except that  $\gamma = 1.5$ . It can be seen that larger value of  $\gamma$  delays the breakup of the tail. Note that this is the case corresponds to a higher molecular weight polymer solution.

**4.3. Viscoelastic fluid.** The viscoelastic effect of the fluid on the tail is investigated numerically by giving non-zero values to the spring constant  $k$  and the results are presented in Figures 4-5. It can be seen from the figures that the elasticity has visible effects. For the small spring constant case ( $k = 10^{-5}$ ), the ‘satellite’ droplets appear but it is not clear whether this is physical or due to numerical instability. For the large  $k$  case, the elasticity is dominant and the droplet is pulled back towards the nozzle exit and the tail acts as a spring.

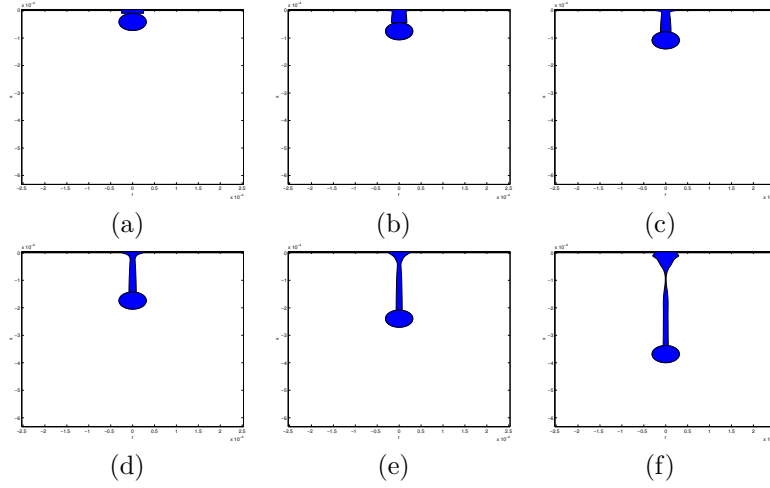


FIGURE 1. Case 1. Power-law elongational viscosity with  $\gamma = 1.1$ . Shape of the tail for  $R_s = 1.25R_n$  and  $R_c = R_n$  at (a)  $t = 0$ ; (b)  $t = 0.5t_0$ ; (c)  $t = t_0$ ; (d)  $t = 2t_0$ ; (e)  $t = 3t_0$ ; (f)  $t = 5t_0$ .

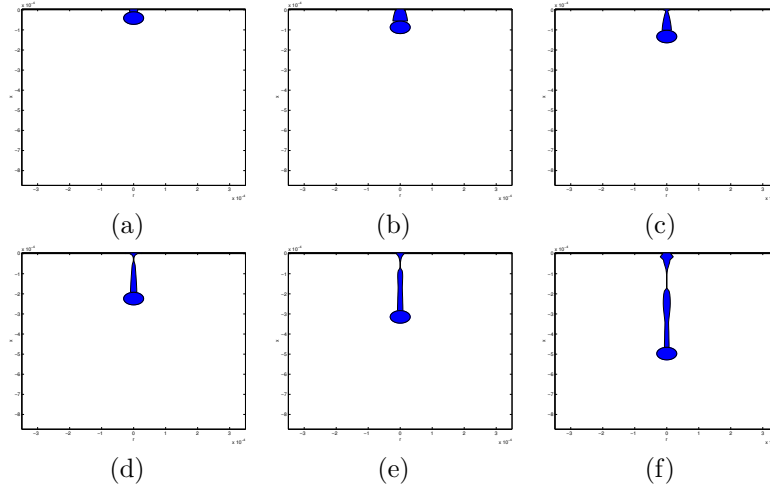


FIGURE 2. Case 2. Power-law elongational viscosity with  $\gamma = 1.1$ . Shape of the tail for  $R_s = 1.25R_n$  and  $R_c = 0.5R_n$  at (a)  $t = 0$ ; (b)  $t = 0.5t_0$ ; (c)  $t = t_0$ ; (d)  $t = 2t_0$ ; (e)  $t = 3t_0$ ; (f)  $t = 5t_0$ .

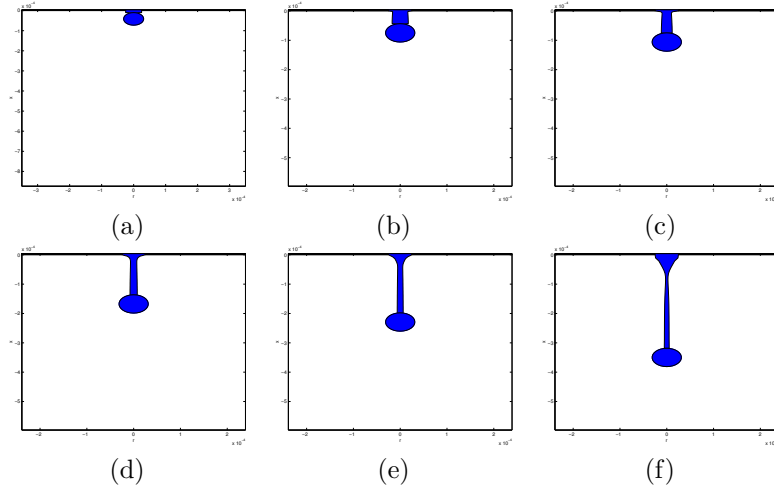


FIGURE 3. Case 3. Power-law elongational viscosity with  $\gamma = 1.5$ . Shape of the tail for  $R_s = 1.25R_n$  and  $R_c = R_n$  at (a)  $t = 0$ ; (b)  $t = 0.5t_0$ ; (c)  $t = t_0$ ; (d)  $t = 2t_0$ ; (e)  $t = 3t_0$ ; (f)  $t = 5t_0$ .

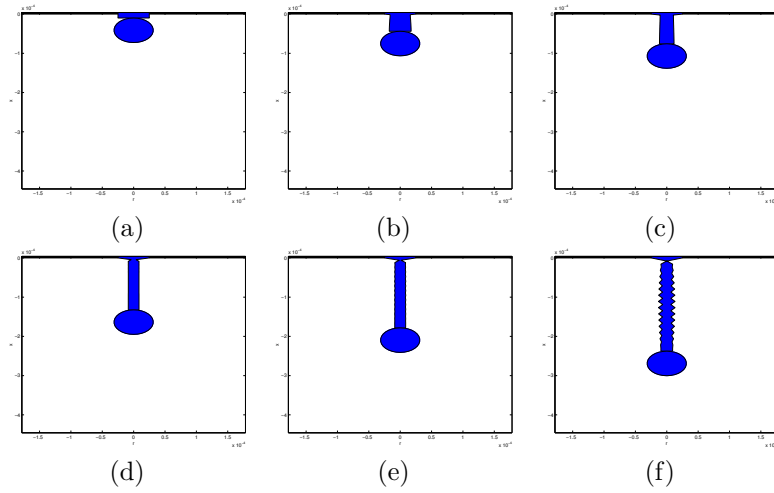


FIGURE 4. Case 4. Viscoelastic fluid with  $k = 10^{-5}$ . Shape of the tail at (a)  $t = 0$ ; (b)  $t = 0.5t_0$ ; (c)  $t = t_0$ ; (d)  $t = 2t_0$ ; (e)  $t = 3t_0$ ; (f)  $t = 5t_0$ .

## 5. Conclusion

In this report we have presented a one-dimensional model to predict the breakup of the liquid filament (tail) attached to an ink-jet droplet. The effect of a power-law

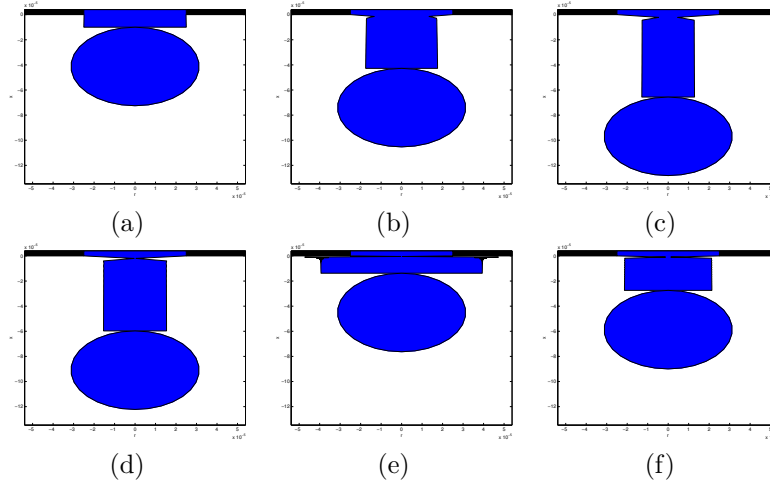


FIGURE 5. Case 5. Viscoelastic fluid with  $k = 10^{-4}$ . Shape of the tail at (a)  $t = 0$ ; (b)  $t = 0.5t_0$ ; (c)  $t = t_0$ ; (d)  $t = 2t_0$ ; (e)  $t = 3t_0$ ; (f)  $t = 5t_0$ .

non-Newtonian stress-strain relationship on the evolution of the tail is investigated by solving the model equations numerically. The results clearly show that the delay (and sometimes lack) of breakup of the droplet tail for higher power relationships. This is consistent with the experimental evidence for higher molecular polymer solutions. The elasticity effect on the behaviour of the tail was also investigated. However, it is not clear whether the appearance of ‘beads’ on the tail, resembling those satellite droplets observed experimentally, are physical or due to numerical instability.

The numerical results also suggest that the mechanism for the breakup of the tail is most likely not surface tension driven since the viscous stress is much larger than the surface tension based on the calculation. The oscillatory nature of the pump drives the flow at the nozzle exit and reverses the flow direction periodically. As a result, the flow inside the liquid filament (tail) near the exit may be ‘sucked’ back while the rest of the tail still moves forward with the main droplet. This causes a decrease of fluid mass locally near where the flow velocity changes direction and forms a ‘neck’. Surface tension may become a contributing factor after the formation of the neck and speed up the breakup process. It is possible that a theoretical model can be developed to predict the final breakup using a similarity solution. However, this was not pursued in the current study.

In summary, these simulation results seem to suggest that the one-dimensional model capture the essential features of the breakup of ink-jet droplet tails. A positive connection between the length of the tail and molecular weight is established from the fact that delay in breakups usually result in a longer tail. And the delay can be contributed to the higher power in the power-law elongational viscosity, which is one of the consequences of higher molecular weight in the polymer solution.

However, it must be noted that the current model has its limitations. First of all, we have ignored the dynamical evolution of the droplet and the flow between the tail and the droplet. Secondly, the flow near the exit is certainly not one-dimensional. Contrary to Newtonian flows where the exit liquid jets contract, the viscoelastic jets increase their diameters ('die swell') after the exit [2]. Therefore, a more careful study is needed to capture all the dynamics correctly and two or higher dimensional models will be needed.

**Acknowledgment.** The report is based on the discussions among the participants who worked on the problem during the workshop, and helpful comments by P. Duineveld and by P. Howell from Oxford. Numerical results were obtained by J.F. Williams during the workshop and are re-produced by HH. Financial support from the Natural Sciences and Engineering Research Council (NSERC) of Canada for HH is gratefully acknowledged.

### Bibliography

- [1] M. Stelter, G. Brenn, A.L. Yarin, R.P. Singh and D. Durst, Investigation of the elongational behavior of polymer solutions by means of an elongational rheometer, *J. Rheol.* **46**(2), 507-527 (2002).
- [2] R.B. Bird, R.C. Armstrong and O. Hassager, *Dynamics of Polymeric Liquids*. John Wiley & Sons, NY (1987).
- [3] D.W. Bousfield et al., Nonlinear analysis of the surface tension driven breakup of viscoelastic filaments, *J. Non-Newt. Fluid Mech.* **21**: 79-97 (1985).
- [4] J.F. Dijksman, Hydro-acoustics of piezoelectrically driven ink-jet print heads, *Turbulence and Combustion* **61**: 211-237 (1999).
- [5] A.D. Fitt et al, Modeling the fabrication of hollow fibers: capillary drawing, *J. Lightwave Tech.* **19**(12): 1924-1931 (2001).
- [6] J. Eggers, Theory of drop formation, *Phys. Fluids.* **7**: 941-953 (1995).
- [7] D.T. Papageorgiou, On the breakup of viscous liquid threads, *Phys. Fluids* **7**: 1529-1544 (1995).
- [8] M. Renardy, Some comments on the surface-tension driven break-up (or the lack of it) of viscoelastics jets, *J. Non-Newt. Fluid Mech.* **51**: 97-105 (1994).
- [9] M. Renardy, A numerical study of the asymptotic evolution and breakup of Newtonian and viscoelastic jets, *J. Non-Newt. Fluid Mech.* **59**: 267-282 (1995).
- [10] M. Stelter et al, Investigation of elongational behaviour of polymer solutions by means of an elongational rheometer, *J. Rheol.* **46**(2): 507-527 (2002).
- [11] A.L. Yarin, *Free Liquid Jets and Films: Hydrodynamics and Rheology*, Longman, Harlow (1993).

## CHAPTER 4

# Catch them ... if you can

Ludolf Meester, Jaap Molenaar, Misja Nuyens, Yves Rozenholc, Koos van Winden

**ABSTRACT.** An important part of forensic science is dedicated to the evaluation of physical traces left at the crime scene like fingerprints, bullets, toolmarks etc. These traces are compared with traces from a suspect. The evaluation of physical traces can be interpreted as the comparison of two noisy signals. We introduce an evaluation of the matching of two noisy signals at diverse scales and localisations in space. In a multi-resolution way a "probability" of matching is computed. Furthermore, a description is given to evaluate the complexity of a shoemark. A likelihood ratio approach is used for comparing two shoemark traces.

**KEYWORDS:** adaptive testing, complexity measure, likelihood ratio

### 1. Introduction

The Netherlands Forensic Institute (NFI) is part of the Dutch Ministry of Justice. This large centre of expertise (about 300 staff members) has forensic casework as core activity. To maintain a high scientific standard, a considerable amount of the budget (about 5 million euro in total) is spent on Research and Development in order to develop new methods and procedures. The NFI also participates in training of crime scene investigators and public prosecutors and the institute maintains various collections and databases. The projects carried out by NFI cover a wide variety of topics such as: DNA-recognition, finger print recognition, drugs analysis, traffic accident analysis, hand writing analysis, paint recognition, etc.

**1.1. Forensic science.** One of the major areas of interest in forensic science is to develop methods to evaluate physical evidence found at scenes of crime. The type of physical evidence found can vary significantly, i.e. traces of blood, hair, fingerprints, bullets, shoemarks, toolmarks etc. But also the state of development of the methods of evaluation differ across different types of evidence. For DNA evidence for instance, scientists have developed models for evaluation which are accepted by the majority of scientists and judicial systems. For other types of evidence the evaluation model is not yet as developed. Much of the evidence is still evaluated by experts in a system that clearly should be classified as "statement of opinion based on experience". These methods are well accepted in most judicial systems and are still common place in most areas of expertise, see for instance [14] and [2].

The problem with most systems of evaluation that are based on "statement of opinion based on experience" is that the experts are more or less operating like a "black box" and it is very hard for other individuals involved in the judicial process to find out whether in a specific case the "black box" is working properly. To solve this problem forensic science has in recent years started to develop models to assist the experts with the evaluation of the evidence and at the same time give the judicial system the option to verify, to a certain extent, their conclusions.

The NFI knows the special function *statisticus* and nearly all activities involve a lot of statistics. The problems to be considered here are from the division for 'toolmarks, footwear and tyre impressions'. They concern recognition of toolmarks and recognition of footwear. We briefly describe them separately.

**1.2. Toolmark recognition.** The perpetration of crimes often involves the use of tools like knives and screwdrivers. Such a tool may leave a mark on the crime scene, e.g. on a window or door frame. On each tool a characteristic pattern is present, which can be observed on a microscale. Such a pattern initially stems from the production process, but it usually gets modified through use of the tool. The pattern makes the tool unique. It is to be expected that a mark stemming from a certain tool contains the same pattern, establishing a clear relationship between tool and mark. In the forensic practice the procedure concerning toolmarks is as follows. If at the crime scene a mark is detected, its pattern is carefully conserved, e.g. via photographs and/or special substances. Corresponding tools possessed by suspected persons are seized. With such a seized tool marks are made under various circumstances. For example, in case of a screwdriver marks are made at various angles of inclination between screwdriver and surface. The patterns of the resulting marks are compared to the mark found at the crime scene. From the result of this comparison a conclusion whether the crime mark corresponds to the tool or not is drawn. For this, probabilistic models are used and the probability that the mark is caused by the tool is expressed according to some classification with levels like "highly probable, probable, less probable, probably not, certainly not".

**1.3. Shoe print comparison.** For shoe print analysis the procedure is similar to the toolmark procedure. Now the tool is a shoe sole or a bare foot and the mark is a print of it in sand or soil. The difference with toolmark comparison is that in shoe print comparison (usually) more complex shapes have to be analysed.

**1.4. Concepts and organization of the report.** The conceptual questions are the same for both kinds of marks. We mention the main items. The accuracy with which the marks are recorded is not infinitely high. Measurements have of course always some inherent inaccuracy, but it also often happens that the mark at the crime scene has been degraded in the time interval between the crime and the spot investigation by the forensic experts. The patterns/marks produced in the laboratory by the tool of the suspect may differ from the mark at the crime scene, even if it is absolutely certain that the tool is indeed used during the crime, simply because the circumstances always slightly differ. So, the statement that a mark stems from a tool will always have a probabilistic character and be based on statistical considerations. The present project is meant to develop reliable models



to be used for these purposes.

In section 2 we focus on toolmark recognition. There the question to be addressed is the following. Given two one-dimensional patterns, i.e. functions on an interval, one representing the mark found at the crime scene and the other produced in the laboratory by a tool, what is the probability that these two marks stem from the same tool? Similarity between two functions on a finite interval may be by chance. How to incorporate the possibility of similarity by fortune into the eventual conclusion? In section 2 a highly promising method is proposed which turned out to be very effective in several bench mark tests.

In sections 3 and 4 we deal with the shoe print problem. Here, a new type of question arises. If in two marks small two-dimensional subpatterns are found that agree to a high degree, what weight should this get in the whole procedure? Similarity in highly complex subpatterns must clearly be rated higher than when two simple shapes agree. This leads to the problem how to define the ‘complexity’ of geometric patterns in the plane. This issue is dealt with in section 3, while section 4 is devoted to the development of a statistical model to draw conclusions about similarity of complete shoe print patterns.

## 2. Toolmark recognition: testing nullity in a regression framework

In this section <sup>1</sup> we develop a model for the efficient comparison of striation marks as found when comparing for instance toolmarks or bullets. We here focus on toolmarks. The classical problem description is as follows: one mark - made by a tool - is found on the crime scene, the other is made in laboratory after having found a potential associated object by a suspect. The question then is "To what extent do these marks match?" and further "what is the evidential value of such a match?".

One way to evaluate this evidential value is to translate the difference between the probability of declaring a match if the two marks are made by the same tool and the probability of declaring the match when the marks are made by different tools into the *likelihood ratio*

$$LR = \frac{P(\text{match} \mid \text{marks made by the same tool})}{P(\text{match} \mid \text{marks are made by different tools})}.$$

The problem that rises immediately when one tries to operationalize this likelihood ratio approach is that in practice it is very hard to estimate the denominator. In section 4 we will calculate this likelihood ratio under simplifying assumptions.

In forensic science the physical marks are considered as "fingerprints" of the associated tool or gun. It is assumed that if the same tool makes several marks in a row, the marks will be very similar and distinguishable from marks made with another tool or gun. The assumption is that this holds for "good" or "perfect and complete" marks. However, in practice the marks are unfortunately not perfect. Sometimes marks found display only a partial mark, the crime scene mark and comparison mark are made under different circumstances, tools are used between the time of the crime and time of making the comparison mark etc. Therefore,

---

<sup>1</sup>This section is essentially an abbreviated version of [4]

noise is added to the "fingerprint". Another source of noise is due to the fact that the marks have to be recorded, photographed and normalized to make a comparison possible. In the method described below we assume that this preprocessing has already taken place. A good example of a possible normalization can be found in [3].

In this section these "fingerprints" are modeled as a noisy one-dimensional or two-dimensional signal depending on whether it is a striation comparison (reduced to gray levels along the perpendicular direction of the axis of striation) or a shape print comparison (reduced to gray levels on a surface). Hence, we have at our disposal two noisy signals

$$\check{Y} = \check{s} + \check{\varepsilon} \text{ and } \tilde{Y} = \tilde{s} + \tilde{\varepsilon},$$

where  $\check{s}$  and  $\tilde{s}$ , are two vectors in  $\mathbb{R}^n$  (for one-dimensional signals) or two  $n_1$  by  $n_2$  matrices (for two-dimensional signals) and where  $\check{\varepsilon}$  and  $\tilde{\varepsilon}$  are vectors (resp. matrices) of i.i.d. random errors of corresponding sizes.

To answer the question "Do these marks match?" and to evaluate the strength of the evidence, we propose to use the difference of the noisy signals  $Y = \tilde{Y} - \check{Y}$  and to test whether  $s = \tilde{s} - \check{s}$  equals 0 or not. This test is done not only globally but also locally: two "fingerprints" may not fit globally because additional use of the tool between the making of the two marks might have lead to a change of part of the "fingerprint". Instead of estimating the likelihood ratio of a match based on the probability of matching when the marks are made by the same tool and the probability of a "random match" when the marks are made by different tools, we use tests of the nullity of  $s$  in a specific piecewise way at different resolution levels to compute a *measure of matching*. To each piece (i.e. each test of nullity) is associated an elementary measure of matching. We compute the global measure of matching as a weighted sum of these elementary measures. The knowledge of the expert then should be used to convert this measure of matching into a real probability of matching.

**2.1. The statistical problem.** Let us consider the regression model

$$Y = s + \varepsilon,$$

where  $Y$  is a vector in  $\mathbb{R}^n$  of  $n$  observations,  $s$  is an unknown vector in  $\mathbb{R}^n$  and the  $\varepsilon$  is an unknown random centered vector with i.i.d. components. We consider a test of  $s = 0$  against  $s \neq 0$ . More precisely, we test the hypothesis

$\mathbf{H}_0$  : "Y is a random centered vector with i.i.d. components"  
against

$\mathbf{H}_s$  : "there exists a non-zero vector  $s$  in  $\mathbb{R}^n$  and a random centered vector with i.i.d. components  $\varepsilon$  such that  $Y = s + \varepsilon$ "

Here, various tests of nullity can be considered, associated to specific assumptions under the non zero hypothesis. Most are based on Gaussian assumptions on the vector of errors  $\varepsilon$  and/or an asymptotic construction for  $n$  going to infinity and/or the variance of the signal going to zero. Several tests of nullity in regression or related frameworks have been studied by many authors. We cite here the work of [1] in a classical non asymptotic Gaussian framework.

In other classical frameworks related to this problem, such as the Gaussian sequence model (Gaussian errors and a variance assumed to decrease like  $1/n$ ) or functional regression ( $s = (S(x_1), \dots, S(x_n))$  for  $S$  a real function on  $[0, 1]$  or  $\mathbb{R}$ ), the authors consider on one hand stronger assumptions on the noise structure such as Gaussian noise and/or suppose a known variance, on the other hand regularity assumptions on  $S$  (supposed to be in a Sobolev or a Besov ball for example) and prove (except for the previously cited work of Baraud & *al.*) asymptotic results with respect to the size  $n$ . We refer to the recent book [9] of Ingster and Suslina for a complete study on these asymptotic approaches in Gaussian model for nonparametric testing and point out the papers of [6], [7], [8], [11], [10], [15], [5].

Let us consider a sequence of subintervals of the integers in  $\{1, \dots, n\}$  denoted  $(I_m)_m$  where  $m$  represents an index belonging to a set  $\mathcal{M}_n$  which can vary with  $n$ . Later, we will state precisely the possible choices of this sequence of intervals. We denote by  $\ell_m$  the number of integers in the interval  $I_m$ . For each interval  $I_m$ , we consider the vector  $Y_m$  of  $\mathbb{R}^{\ell_m}$  which are those coordinates of  $Y$  which are indexed by  $I_m$ .

Let us consider a real number  $\alpha$  in  $[0, 1]$  and the sequence of levels

$$\alpha_m = \alpha \frac{[\ell_m]^c}{\#\mathcal{M}_n}, \text{ for } m \in \mathcal{M}_n$$

where  $c$  is a fixed parameter - used to control the significance of the tests with respect to the length of the intervals. For each  $m$  in  $\mathcal{M}_n$ , we test at the level  $\alpha_m$  whether the sub-vector  $Y_m$  is 0 or not. Let us note  $\Phi_m$  the result of this test:  $\Phi_m = 0$  if  $\mathbf{H}_0$  is not rejected and 1 otherwise.

One can remark that the global test of nullity  $\Phi$ , defined by  $\Phi = 0$  if all  $\Phi_m$  are null and 1 otherwise, is of level less or equal to  $\alpha \sum_m \ell_m^c / \#\mathcal{M}_n$ . Indeed

$$\mathbb{P}_{\mathbf{H}_0}(\Phi = 1) = \mathbb{P}_{\mathbf{H}_0}(\exists m \in \mathcal{M}_n, \Phi_m = 1) \leq \sum_{m \in \mathcal{M}_n} \mathbb{P}_{\mathbf{H}_0}(\Phi_m = 1) = \frac{\alpha}{\#\mathcal{M}_n} \sum_{m \in \mathcal{M}_n} \ell_m^c.$$

Under Gaussian assumptions and following [1], it is possible to derive the power of the test  $\Phi$  for our family of sub-tests  $(\Phi_m)_m$ .

**2.2. A model for evidential testing.** In order to construct a model for evidential testing, one has to keep in mind the two characteristics of this problem. The time interval between acquiring the two noisy signals to be compared can be large. Then, it is natural to imagine that the signals will always be globally different and a global test of nullity will always reject the hypothesis  $\mathbf{H}_0$ , hence such a test is not relevant in practice. For example, the shoe of a criminal may have been used after a crime and the shape of its surface may have changed. It is important to understand that the experts usually look only for small features as evidence as they are usually considered as strongly individualizing. In one-dimensional comparison

for tool-marks, there are experts that consider five matching lines in a row on two toolmarks as sufficient evidence. In our framework, this means that  $s$  has five co-ordinates in a row equal to 0.

The model introduced here is based on a dyadic multi-resolution evaluation of the *measure of matching*, or *measure of similarity*, using a dyadic sequence of intervals. The measure of matching  $\mu$  has to be all the greater so that the signals match (i.e.  $s = 0$ ) on larger interval. It has to be additive with respect to disjoint intervals, i.e.

$$\mu(\text{match on } A \cup B) = \mu(\text{match on } A) + \mu(\text{match on } B),$$

for any  $A, B$  s.t.  $A \cap B = \emptyset$ . Finally it has to be additive with respect to the resolution, i.e. for any  $B \subset A$

$$\mu(\text{match on } A, B) = \mu(\text{match on } A) + \mu(\text{match on } B).$$

We make precise our construction for one-dimensional signals (tool-mark evidence). Similar constructions can be made for higher dimensions. Let us assume that  $n = 2^J$ . We consider the set of index  $\mathcal{M}_n = \{m = (j, k), j = 0, \dots, J-1, k = 1, \dots, 2^j\}$  and the integer intervals of the form

$$I_{j,k} = \left\{ i \in \mathbb{N}, \frac{k-1}{2^j} < \frac{i}{n} \leq \frac{k}{2^j} \right\}.$$

The tests of nullity associated with these intervals are denoted by  $\Phi_{j,k}$ . Let  $(\pi_j, j \in \mathbb{N})$  denote a sequence of non negative reals such that  $\sum_{j=0}^{\infty} \pi_j = 1$ . The measure of matching associated to the interval  $I_{j,k}$  is  $2^{-j}\pi_j(1 - \Phi_{j,k})$ . The global measure of matching is equal to

$$\mu(\text{matching}) = \sum_{j=0}^{J-1} 2^{-j}\pi_j \sum_{k=1}^{2^j} (1 - \Phi_{j,k}) = 1 - \sum_{j=0}^{J-1} 2^{-j}\pi_j \sum_{k=1}^{2^j} \Phi_{j,k}.$$

Hence,  $2^{-j} \sum_{k=1}^{2^j} (1 - \Phi_{j,k})$  appears as the measure of matching at resolution level  $j$ . Our definition of the measure of matching can be interpreted using conditional expectations as the expectation of matching at any levels using a prior distribution  $\pi$  on the resolution levels. More simply, if all tests of the level  $j$  are equal to 0 ( $s = 0$  on  $I_{j,k}$  for all  $k = 1, \dots, 2^j$ ) the contribution to the measure of matching is equal to  $\pi_j$  and the previous formula defines a number between 0 and  $1 - \sum_{j=J}^{\infty} \pi_j$ .

To ensure that the measure of matching is invariant by translation a natural sequence of  $\pi_j$  is the uniform distribution by level:  $\pi_j = 1/(2^j \log n)$  for  $j = 0, \dots, J-1$  and 0 otherwise.

By construction, on each interval  $I$ , the test associated with  $I$  makes an error of first kind (i.e. does not reject  $\mathbf{H}_0$  when  $s$  is non zero) if  $s$  is too close to 0 on  $I$  - the precise sense of "close" depend on the choice of the test of nullity considered and is not discussed here as this has to be studied in practice using real data. Using a uniform distribution by level ( $\pi_j = 1/(2^j \log n)$ ), we give to the errors of first kind a weight which is only proportional to the level of each elementary test and hence which is only proportional to the length of the interval up to the power constant  $c$ .

An other way to compute a measure of matching is to consider the  $p$ -value of each test  $\Phi_{j,k}$  denoted  $\rho_{j,k}$  and to define

$$\mu(\text{matching}) = \sum_{j=0}^{J-1} 2^{-j} \pi_j \sum_{k=1}^{2^j} \rho_{j,k}.$$

In order to keep translation invariance here again a natural distribution for the  $\pi_j$  is the uniform by level previously defined.

### 3. Shoe print comparison: complexity of patterns

Shoe print patterns usually contain a lot of details, but for shoe print comparison only those subpatterns are relevant that are not general and found on all shoes of a specific type. In the forensic practice one is interested in subpatterns that are supposed to be characteristic for the shoe under consideration. When comparing the patterns of two shoe prints it may happen that the prints match at some subpatterns but differ at others. Matching of subpatterns means that both prints have similar subpatterns at similar positions. The matching is never perfect due to circumstances. In Figure 3 examples of realistic subpatterns are shown.



FIGURE 1. Some realistic subpatterns.

The probability that the two prints stem from the same shoe has to be determined via some probabilistic model. Such a model will be proposed in the next section. A necessary ingredient of such a model is the rate of complexity of a subpattern. If two prints match with respect to a simply shaped pattern, this agreement has to be weighed differently from a match at a highly intricate pattern, simply because a shape with high complexity characterizes the shoe to a higher degree. The underlying idea for this is: the more complex a feature, the more rare it is. The question here is how to define the complexity of a pattern. We shall deal separately with one and two-dimensional features. In fact it could suffice to deal only with two-dimensional patterns, since the one-dimensional case is a special case then. However, we pay rather extensive attention to the one-dimensional case, since it helps in developing and clarifying ideas. Moreover, in the forensic practice one-dimensional features on shoe prints are treated as a special class. For this class tables are available in which the complexity of these patterns is specified according to common sense and expertise. Comparison of the numbers of these tables and the present results is one of the intriguing aspects of this project. This comparison can not yet be reported here, since this is part of future research.

The complexity of shapes is hard to define in general. It is to be expected that

any definition involves a relative measure with which two shapes can be compared. The absolute value of the complexity of shapes is in practice irrelevant. A particular nice starting point is the following definition of complexity, which is from [12] and [13].

DEFINITION 3.1 (Complexity of an object). The complexity of an object is the length of its description in some language.

Of course, one has to choose the language suitable, depending on the nature of the object. To illustrate this definition, consider the following examples.

*What is the complexity of a natural number?*

The numbers 0, 1, 2, ..., 9 are described with one digit, the numbers 10, 11, 12, ..., 99 are described with two digits, etcetera. In general, the numbers  $10^n, \dots, 10^{n+1} - 1$  are described with  $n$  digits. In this simple case it is clear that the parameter  $n$  is a good measure of the complexity  $c$ . So for a natural number  $N$ :

$$(1) \quad c(N) \sim n \sim \log(N) .$$

Here, the symbol  $\sim$  stands for 'scales with'; it implies that possible multiplicative and/or additive constants are left out.

*What is the complexity of the repeating sequence of finite length*

$$S = 104104104104104?$$

Since the sequence is repeating, it can be described by mentioning the periodic part (104) and the number of times it is repeated (5). So, the parameters are period and repetition rate. As seen above, the length of a natural number grows with the log of its value. It is not directly clear how to combine the complexities of period and repetition rate. At first we tend to add them. For short series this is appropriate. However, in the limit of a very long series, the contribution from the repetition rate will then strongly dominate the contribution from the period and the information in the period gets lost. For long series we prefer to multiply both contributions, so

$$(2) \quad c(S) \sim c(\text{period}) * c(\text{repetition rate}) .$$

*What is the complexity of an arbitrary real number?*

An arbitrary real number needs infinitely many digits to describe it, so it has complexity  $c = \infty$ . Note that the real number  $\pi$  has also infinitely many digits. However, it can be described alternatively by a few words, i.e. as the ratio of circumference and diameter of a sphere. The number  $\pi$  is certainly not 'arbitrary'.

**3.1. The complexity of one-dimensional patterns.** Here, we restrict ourselves to shapes that consist of one-dimensional components such as line and arc segments.

*Line segment  $L$* 

From the very general definition above we may deduce a definition for a line segment. Such a segment can be considered as a sequence of identical pixels, so with a simple internal structure, and with a certain length  $D$ . We assign to a single pixel the complexity one. Formula (2) then yields

$$(3) \quad c(L) = \log(D/D_0) .$$

Here,  $D_0$  is the minimal length that is relevant for the shoe print under consideration. This is dependent on the signal to noise level of the print. A typical value is  $D_0 \approx 1.5$  mm. For  $D = D_0$  we have  $c(L) = 0$ .

*Two line segments  $L_{2,V}$  forming a 'V'-shape.*

In practice these shapes are considered a relevant separate class. The total size of this pattern is measured by the diameter  $D$  of the smallest circumscribing sphere. The complexity of the internal pattern is measured by counting the number of parameters characterizing it. We find two lengths and an angle. Only the ratio of the lengths is relevant, since the pattern may be freely scaled: its size is accounted for via  $D$ . So, we have two dimensionless parameters and this leads to

$$(4) \quad c(L_{2,V}) = 2 \log(D/D_0) .$$

*Two crossing line segments  $L_{2,C}$ .*

We proceed in an analogous way. The internal structure of  $L_{2,C}$  has 4 lengths and one angle. So we have 4 dimensionless parameters and

$$(5) \quad c(L_{2,C}) = 4 \log(D/D_0) .$$

*Triangle  $L_{3,T}$ .*

The internal structure of  $L_{3,T}$  is characterized by 3 lengths. So we have 2 dimensionless parameters and

$$(6) \quad c(L_{3,T}) = 2 \log(D/D_0) .$$

*Three crossing line segments  $L_{3,C}$ .*

The internal structure of  $L_{3,C}$  is characterized by 7 lengths and two angles. So we have 8 dimensionless parameters and

$$(7) \quad c(L_{3,C}) = 8 \log(D/D_0) .$$

*Arc (circle segment)  $A$ .*

The internal structure of  $A$  is characterized by its length, the radius of the circle, and the angle. We remark that the length of the segment is different from the diameter  $D$  of the circumscribing sphere. For a full circle,  $D$  equals the diameter of the circle, but the length of  $A$  equals  $2\pi D$ . We have 2 dimensionless parameters and

$$(8) \quad c(A) = 2 \log(D/D_0) .$$

*Arbitrarily curved line segment*

An arbitrarily curved line segment can in good order be approximated by series of arcs of different radii. This introduces an error and the expert has to specify which error is still acceptable. The complexity of the curve is then the sum of the complexities of the circle segments. So, if  $N$  segments are used, the internal

structure of the curve is  $2N$ . An alternative approach is to construct a polynomial that fits the curve reasonably well. The number of coefficients of the polynomial then yields the internal complexity of the curve. So, if the polynomial is of order  $N$ , the complexity is  $N + 1$ .

**3.2. The complexity of two-dimensional features.** Figure 3.2 gives some idea how two-dimensional features look like in the shoe print practice: from a completely filled circle to a more or less randomly distributed pattern. Here, black and

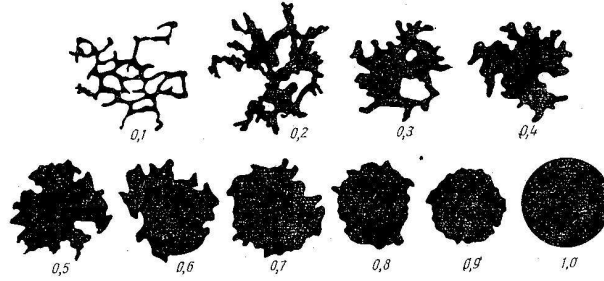


FIGURE 2. Some two-dimensional shapes.

white shapes are studied that are specified via pixels on a grid. We start from the intuitive observation that the complexity of the shape is low if it contains a lot of symmetry and is high when we can not discover any regular pattern in it. It is clear that if the pattern is completely random we have to specify it pixel by pixel. In the latter case we expect that the white and black pixels are distributed more or less uniformly over the shape. This suggests to introduce for a two-dimensional feature the concept of entropy. We define it in this context as follows.

- a. Put a square grid with boxes over the object. Take the grid size equal to  $n$  pixels. The resulting number of boxes is  $N_n$ , say.
- b. The number of pixels in each box is  $n^2$ . In box  $i \in \{1, \dots, N_n\}$  we count the number  $n_b$  of black pixels; so box  $i$  has  $n^2 - n_b$  white pixels.

The entropy in box  $i$  is then defined as:

DEFINITION 3.2 (Entropy of box  $i$ ).

$$s_i = \frac{1}{a} \log\left(\frac{n^2}{n_b}\right),$$

with  $\binom{N}{M}$  for arbitrary  $N$  and  $M$  defined as usual by

$$\binom{N}{M} = \frac{N!}{M!(N-M)!}.$$

The normalization factor  $a$  is defined as  $a = \log\left(\frac{n^2}{n^2/2}\right)$ . This factor is chosen such that the entropy in box  $i$  is maximal and equal to 1 if there are as many black pixels as white pixels, i.e.  $n_b = \frac{1}{2}n^2$ .



Summing the entropies over all the boxes we get the total entropy  $s_n$  of the grid under consideration:

$$(9) \quad s_n = \sum_{i=1}^{N_n} s_i.$$

The entropy of course depends on the grid size  $n$ . By varying  $n$ , it is measured on which length scale the object has most structure. The complexity is therefore defined as the maximum entropy under variations of the grid size. So

DEFINITION 3.3 (Complexity of two-dimensional objects (not normalized)).

$$\bar{c} = \max_n s_n.$$

Here,  $n$  is varied continuously. In practice, however, one usually will halve the grid size again and again. It is clear that boxes with no variation in colour have vanishing entropy. So, if an object consists of subpatterns of one colour, the inner part of such a subpattern will not contribute as soon as the grid size becomes smaller than the size of the pattern. Only boxes containing parts of the boundaries have variation in colour and contribute to the entropy. We conclude that this complexity  $\bar{c}$  determines the total length of the boundaries in and around the object. This agrees with the intuitive notion of complexity for objects of this kind. The important point here is that the definition also applies to objects that do not consist of well-defined subshapes of one colour, but have a more or less fractionated appearance or even a fractal dimension.

The definition of  $\bar{c}$  above has the disadvantage that it does not yet reduce to the definition for complexity in one dimension in case of a line segment. This is because  $\bar{c}$  is not yet normalized by the size of the object as a whole.

#### Example: reduction to a line segment

As an example we take a line segment consisting of 8 horizontal pixels. According to the procedure described above, it has to be covered by an  $8 \times 8$  grid. If we take  $n = 8$ , this implies that the segment is covered with only one box of size  $8 \times 8$ . Applying the formula we find  $s_8 \approx 0.5$ . Halving the grid size and thus taking  $n = 4$ , we obtain that  $s_4 \approx 1.5$ . Two grid boxes have vanishing entropy since they are uniform in colour. The other two contribute equally. Halving the grid size again we find that now 4 grid boxes contribute and  $s_2 \approx 4.0$ . Of course,  $s_1 = 0$ , since grid boxes consisting of one pixels have vanishing entropy. The maximum is thus found for  $n = 2$  and we denote this  $n$ -value by  $n_{max}$ .

To manage that the two-dimensional complexity definition reduces to the one-dimensional one if applied to a line segment, we adjust the  $\bar{c}$  a bit. Let, as usual,  $D$  be the diameter of the smallest circle around the object. Then, the following definition has the required properties:

DEFINITION 3.4 (Complexity of two-dimensional objects (normalized)).

$$c = \bar{c} \frac{n_{max}}{D} \log(D/D_0).$$

This complexity could be conveniently used as weight function when two shoe print patterns agree on one or more shapes.

#### 4. Shoe prints comparison: likelihood ratios

We are given two prints. One is a trace found on the crime scene. The second is a print of the shoe of the (a) suspect. What can we say, on basis of these two prints, about the two statements '*The shoe of the suspect made the print on the crime scene*' and '*The shoe of the suspect did not make the print on the crime scene*'? We will measure the strength of the evidence by a *likelihood ratio*.

The characteristics of a shoe (print) are the marks (damages) on it. Each mark is given by a vector  $(x, y, s, \alpha)$ , where  $(x, y)$  are the coordinates of the mark on the shoe,  $s$  is the complexity of the shapes of the mark and  $\alpha$  is the orientation of the mark. For reasons of simplicity, we describe the shape of the mark by its complexity and not shape itself. Of course this results in the loss of some information.

Suppose the shoe of the suspect contains  $K$  marks and the crime scene print contains  $K'$  marks. We then define

DEFINITION 4.1. The *suspect evidence*  $E$  is given by the matrix

$$E = \{(x_i, y_i, s_i, \alpha_i), i = 1..K\}.$$

The *crime scene evidence*  $E'$  is given by

$$E' = \{(x'_i, y'_i, s'_i, \alpha'_i), i = 1..K'\}.$$

Note that the order of the rows of  $E$  is not important. Furthermore we define  $(x, y) = (x_i, y_i)_{i=1..K}$ ,  $s = (s_i)_{i=1..K}$  and  $\alpha = (\alpha_i)_{i=1..K}$  and likewise for the columns of the crime scene evidence matrix.

There are two possible worlds. We call them  $H_p$  and  $H_d$ , where the  $p$  stands for prosecutor and the  $d$  for the *defender*. In the world  $H_p$ , the shoe of the suspect made the print. The probability measure in this world is  $P_p$ . In  $H_d$ , the shoe of the suspect did not make the print. In this second world the probability measure is  $P_d$ .

**4.1. A likelihood ratio.** A measure to indicate the strength of the evidence supplied by the two marks is the so-called *likelihood ratio*. The likelihood ratio of an event  $A$  in the two worlds  $H_p$  and  $H_d$  is defined by the ratio

$$(10) \quad LR(A) = \frac{P_p(A)}{P_d(A)}$$

if the two probabilities are well-defined and the denominator is non-zero. In case the event  $A$  is defined in terms of continuous random variables, the likelihood ratio is defined by the quotient of the densities  $f_p$  and  $f_d$  i.e.  $LR(A) = f_p(A)/f_d(A)$ . In the following we will encounter events defined in terms of discrete and continuous random variables. For reasons of simplicity, we will, with abuse of notation, use the notation of (10) for the continuous case as well.

The event  $\{E, E'\}$  consists of those outcomes in the probability space that result in the suspect evidence  $E$  and crime evidence  $E'$ . To measure the strength of the evidence supplied by the two prints, we calculate the likelihood ratio  $LR$  of  $\{E, E'\}$  in the two worlds  $H_p$  and  $H_d$ , i.e.

$$LR = \frac{P_p(E, E')}{P_d(E, E')}.$$

The numerator of  $LR$  is the probability that in the world where the shoe of the suspect is the shoe that made the crime scene mark, the event  $\{E, E'\}$  happens. The denominator is the probability that  $\{E, E'\}$  happens in the world where the suspect is innocent. The quotient of these two terms can be used to measure how likely it is that the suspect is guilty: a high value of  $LR$  pleads against him, a low value strengthens his defense.

In the following we assume that the position of the mark(s), their shape and their orientation are independent. This assumption is not clearly untrue, but whether this independence is valid should be checked with (the) data. Conditioning on  $K$  and  $K'$  then yields

$$(11) \quad LR = \frac{P_p(K, K') P_p((x, y), (x', y') \mid K, K') P_p(s, s' \mid K, K') P_p(\alpha, \alpha' \mid K, K')}{P_d(K, K') P_d((x, y), (x', y') \mid K, K') P_d(s, s' \mid K, K') P_d(\alpha, \alpha' \mid K, K')} \\ = LR_{number} \times LR_{position} \times LR_{shape} \times LR_{orientation}.$$

**4.2. One mark.** As an indication how to obtain (11), we now calculate each of the four factors of the likelihood ratio  $LR$  in the case that both the crime scene print and the shoe of the suspect show one feature, i.e.  $K = K' = 1$ .

The calculation of the four likelihood ratios in (11) follows the same pattern. In the numerator we calculate the probability under the assumption that the crime scene print was made by the suspect's shoe. Hence there is in fact one shoe! The probability is then a product of two factors: the probability (density) that an arbitrary shoe has the given property and the probability (density) that two different prints from the same shoe differ a certain (given) way.

For the denominator, the assumption is that the two prints are made by two different shoes. The two shoes have nothing to do with each other and are independent. The probability (density) that they both contain certain features is then the product of the probability (density) that an arbitrary shoe contains certain feature(s) and the probability that a print from an arbitrary shoe contains certain (other) feature(s). Note that these two probability (densities) are not (necessarily) the same: in the shoe print there is less detail and much more noise.

•  $LR_{position}$ . Let us first look at the numerator. In the world  $H_p$  the shoe of the suspect made the crime scene print. The positions of the two marks however do not have to coincide: both in the making of the print by the suspect and in the measuring of the positions  $(x, y)$  and  $(x', y')$  (little) errors are made. These errors depend on the surface on which the crime scene print is found as well. It seems reasonable to assume that  $(x - x', y - y')$  has approximately a bivariate normal distribution with mean zero, covariance matrix  $\Sigma_{surface}$ , and density  $f_{surface}$ . Let the

surface area of the shoe be  $B$ . Assuming that the position of the mark is uniformly distributed over the surface of the shoe, we have

$$P_p((x, y), (x', y') \mid K = 1, K' = 1) = B^{-1} f_{surface}(x - x', y - y').$$

For the denominator we move to the world  $H_d$ , where the shoe of the suspect did not make the print. Hence the numerator is the probability (density) that the marks on two different shoes are situated in positions  $(x, y)$  and  $(x', y')$  respectively. By the uniformity assumption above and the independence, this density is  $1/B^2$ . Hence

$$\begin{aligned} LR_{position} &= \frac{P_p((x, y), (x', y') \mid K = 1, K' = 1)}{P_d((x, y), (x', y') \mid K = 1, K' = 1)} \\ (12) \quad &= B^{-1} f_{surface}(x - x', y - y') / B^{-2} = B f_{surface}(x - x', y - y'). \end{aligned}$$

Note that (12) does not depend on the unit in which  $x, y$  and  $B$  are measured. Indeed, changing the unit of measurement by a factor  $c$  results in  $c^{-2} B c^2 f_{surface}$  and  $c$  nicely drops out.

•  $LR_{angle}$ . Assuming the distribution of the angle of the mark is uniform, again, the numerator consists of a 'uniform' term  $1/(2\pi)$  and Gaussian 'error' term  $g_{surface}(\alpha - \alpha')$ , again depending on the surface where the crime scene print was found. In the denominator the uniform term  $1/(2\pi)$  is squared. Then  $LR_{angle}$  is given by

$$LR_{angle} = 2\pi g_{surface}(\alpha - \alpha').$$

•  $LR_{shape}$ . For calculating the likelihood ratio of (complexity of) the shapes, we use the theory of section 3. The complexity of a mark is an element of a countable set. Its discrete distribution  $P_{compl}$  should be obtained from the data, possibly in combination with the theory of section 23. Let  $P_{err}(s, s')$  be the probability that a shoe with a mark of complexity  $s$  produces a print with a mark of complexity  $s'$ . As before, this is term caused by a 'reproduction error'. Then

$$(13) \quad LR_{shape} = \frac{P_{compl}(s) P_{err}(s, s')}{P_{compl}(s) P_{compl}(s')} = \frac{P_{err}(s, s')}{P_{compl}(s')}.$$

Note that  $P_{err}(s, s')$  is not the same as the probability that a print with a mark of complexity  $s'$  was produced by a shoe with a mark of complexity  $s$ , nor is it equal to  $P_{err}(s', s)$ .

•  $LR_{number}$ . In the numerator we have  $P_p(K, K') = P(K' = 1) P_p(K' = K \mid K' = 1)$ . The probability that the shoe of the suspect contains one mark given that the crime scene print (made by the same shoe!) contains one mark,  $P_p(K' = 1 = K \mid K' = 1)$ , depends on a number of things:

- the time the shoe has been used between the crime time and the moment the shoe was secured
- the probability that a mark on a shoe actually is found in a print the shoe leaves behind - this in its turn depends on the surface
- the probability that a mark in the crime scene print is caused by something temporarily attached to the shoe, for instance a leaf, a twig, etcetera.

Therefore this probability is perhaps the most complex of all the probabilities in this section. Further study (of data) could reveal its nature. Let us conclude by calculating the denominator:  $P_d(K, K') = P(K = 1)P(K' = 1)$ . The likelihood ratio then satisfies

$$LR_{number} = \frac{P_p(K = 1 | K' = 1)}{P(K = 1)}.$$

If the marks on the suspect's shoe are similar to those on the crime scene, the likelihoods described above will in general be larger than one. Hence, when it is impossible to compute all likelihoods, we can find a lower bound for  $LR$  by computing some of them. Such a lower bound might already be powerful enough for a clear conclusion about strength of the evidence.

**Acknowledgements** We thank the members of the NFI that were part of our study group. Huub Hardy, Andre Hoogstrate, Ies Keereweer and Marjan Sjerps were almost continuously present, usually arrived even before the mathematicians did and created a very pleasant atmosphere. Some of the ideas brought forward by them have found their way into this report. We thank Sergei Anisov and Dinard van der Laan for interesting discussions and for developing some of the ideas written down in these proceedings. Finally, we thank the NFI for organizing a symposium in May 2003 where we have explained our ideas and have had the opportunity to 'match' them with the views of the NFI experts.

### Bibliography

- [1] Y. Baraud, S. Huet, and B. Laurent. Adaptive tests of linear hypotheses by model selection. *Annals of Statistics*, 31(1), 2003.
- [2] A. P. A. Broeders. Daubert en Saks: het einde van de klassieke identificatiecriminalistiek? *Nederlands Juristenblad*, 36:1786–98, October 2002.
- [3] M. Heizman. Strategies for the automated recognition of marks in forensic science. In *Proceedings SPIE 4709*, April 2002.
- [4] A. Hoogstrate and Y. Rozenholc. Testing nullity in a regression framework to evaluate forensic evidence. *To appear*.
- [5] Joel L. Horowitz and Vladimir G. Spokoiny. An adaptive, rate-optimal test of linearity for median regression models. *J. Amer. Statist. Assoc.*, 97(459):822–835, 2002.
- [6] Yu. I. Ingster. Asymptotically minimax hypothesis testing for nonparametric alternatives. I. *Math. Methods Statist.*, 2(2):85–114, 1993.
- [7] Yu. I. Ingster. Asymptotically minimax hypothesis testing for nonparametric alternatives. II. *Math. Methods Statist.*, 2(3):171–189, 1993.
- [8] Yu. I. Ingster. Asymptotically minimax hypothesis testing for nonparametric alternatives. III. *Math. Methods Statist.*, 2(4):249–268, 1993.
- [9] Yu. I. Ingster and Irina A. Suslina. *Nonparametric goodness-of-fit testing under Gaussian models*, volume 169. Lecture notes in statistics (Springer-Verlag), 2003.
- [10] O. V. Lepski and A. B. Tsybakov. Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point. *Probab. Theory Related Fields*, 117(1):17–48, 2000.

- [11] Oleg V. Lepski and Vladimir G. Spokoiny. Minimax nonparametric hypothesis testing: the case of an inhomogeneous alternative. *Bernoulli*, 5(2):333–358, 1999.
- [12] Li, Ming & Vitányi, Paul. An introduction to Kolmogorov complexity and its applications. Second edition. Graduate Texts in Computer Science. Springer-Verlag, 1997.
- [13] Li, Ming & Vitányi, Paul. An introduction to Kolmogorov complexity and its applications. Texts and Monographs in Computer Science. Springer-Verlag, 1993.
- [14] M. J. Saks. Merlin and Solomon: Lessons from the laws formative encounters with forensic identification science. *Hastings Law Journal*, 49:1069–1141, 1998.
- [15] V. Spokoiny. Data-driven testing the fit of linear models. *Math. Methods Statist.*, 10(4), 2002.

## CHAPTER 5

# Hanging a Carillon in a Broek-system

Chris Budd, Robbert Fokkink, Geertje Hek, Peter van der Kamp,  
Derk Pik, Vivi Rottschäfer

**ABSTRACT.** A carillon is a musical instrument consisting of a (fairly large) set of bells that can be tolled by playing a keyboard, which is usually located one storey below the bells in a tower. Wires connect the keyboard to the clappers of the bells, forming an intricate web that is hinged to the walls. The web may vibrate, rub and tangle during play and some of the keys may require more pressure than others. The paper presents some methods to prevent these problems.

**KEYWORDS:** geometric optimization, carillon, equilibration of musical instruments.

### 1. Introduction

A carillon is a musical instrument. It is a set of bells in a church tower that can be played by finger keys. The bells are hung high up in the tower. The keyboard, which is one metre long, is located one or more floors below. It is connected to the bells by an intricate web of wires, illustrated by figure 1.

There are two common ways to connect a key to a bell, the tumble system and the broek-system. This paper considers the broek-system only. As the tumble system is easier to design, it is nowadays the most common way to hang a carillon. However, for both playing and maintenance reasons, the broek-system is preferred. A sketch of the broek connection is shown in figure 2. In this particular example the wire has two segments that are connected by a ring, a so-called ‘broekring’, that is connected to the wall by a wire. By depressing the key the wire is pulled down and the clapper strikes the bell. Using this kind of wire construction *each and every* bell of the carillon needs to be connected to the keyboard. Figure 1 illustrates that this is not an easy task, especially not since there are quite a few requirements that an ideal broek-system has to meet. For one, the wires should not be too close to each other because they will swing a little during play.

The analysis of the system comprises the following stages.

- 1 We firstly study the statics of a single wire-bell connection. In particular we identify the force and displacement that needs to be applied to the keyboard wire to ring the bell, and see how this force depends upon the geometry of the wire configuration. A key feature of this calculation is identifying a suitable geometry of the broekring system which leads to an even force for the keyboard player when

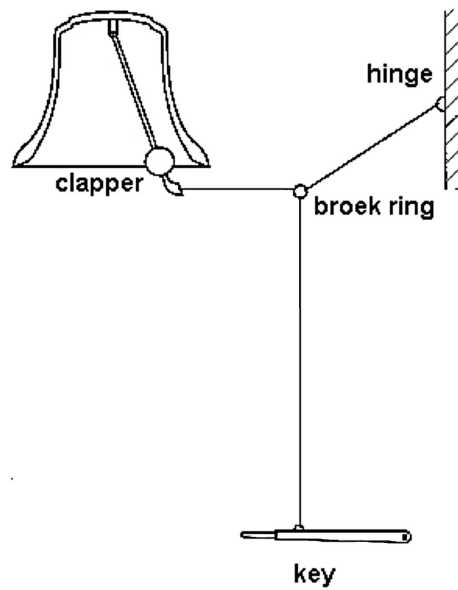


FIGURE 1. Schematic broekring connection

playing each bell.

- 2 We then use Lagrangian mechanics to study the (periodic) dynamics of a single wire-bell connection with one broekring. The oscillation of the system causes two problems. Firstly it makes it difficult to press the the same key too often. Secondly, the oscillations can potentially bring the wires for different bells into contact. Calculating the (frequency and) amplitude of these vibrations allow us to determine how far apart the wires should be.
- 3 Finally we apply the information determined in stages 1 and 2 above to a larger system with several bells. Each wire system can be hung in a certain geometry identified by 1 and the wires must be kept separate by 2. Applying each of these rules as constraints in the full network, we can identify an 'optimal' configuration by using a (partly stochastic) *greedy algorithm* to determine where the bells should be placed in the tower.
- 4 Throughout the text the reader finds four rules of thumb for the design of a carillon, that follow from our calculations.



## 2. Figures and facts

In this section we identify some key constraints of the geometry of the bells and of the wires connecting them to the keyboard.

**2.1. Notation.** The following symbols are used throughout the article.

$N$	=	number of bells
$T$	=	top of bell, the clapper is attached here.
$C$	=	end of the clapper, the ball is here
$R$	=	broekring
$K$	=	key
$H$	=	hinge where the broekring is attached to the wall
$\sigma$	=	angle of the clapper with the vertical
$\gamma$	=	angle between the clapper and the clapper-wire
$\alpha, \beta$	=	angles at the broekring
$l_0$	=	length of the clapper
$l_i$	=	lengths of the other wires
$m_c$	=	mass of the clapper (varies from 0.3-30 kg (!))
$m_i$	=	masses of the wires
$dC$	=	distance the clapper moves (approx 2–5 cm)
$dK$	=	displacement of the key (approx 5 cm)

**2.2. The general geometry.** A carillon comprises  $N$  bells connected to an equal number of keys on the keyboard.  $N$  can be quite large and towers with 40 bells exist. In the tower we examined, the bells could be hung at one of two levels at a set of fixed points around the tower circumference. The position of the keyboard is fixed, as is the order of the keys, however there is freedom in the position of the bells around the towers. In Section 5 we will consider the problem of the optimal positioning of the *bells*. In this section we will assume that the bell is fixed and look at the resulting configuration of the *wires*. In the simplest case the keys are attached to the bells via a single broekring system, but in small towers containing many bells it is often necessary to use many broekrings. Whilst giving additional flexibility to the possible geometry of the wires, the use of more than one broekring is undesirable as it leads to significant extra friction in the system. We will look mostly at problems with one broekring, but will give a general formula for problems with many broekrings.

**2.3. Playing the keyboard.** The player sits at a horizontal, and straight, keyboard in which each key is depressed by  $dK$ , which is preferably the same for all keys. The wire attaching the key to the bell is initially vertical, with a broekring directly above the key. The motion of the wire causes the clapper of the bell to move a distance  $dC$  and then to strike the surface of the bell itself. When depressing a key to play the bell the player prefers the key to ‘feel’ the same when played, regardless of the bell to which it is attached. For all keys the player prefers a similar displacement  $dK$  and a similar force  $F$  needed to move the key. These preferences conflict, since  $FdK$  is equal to the gain of potential energy of the clapper and the clappers have varying weights. We come back to this in section 4.9.

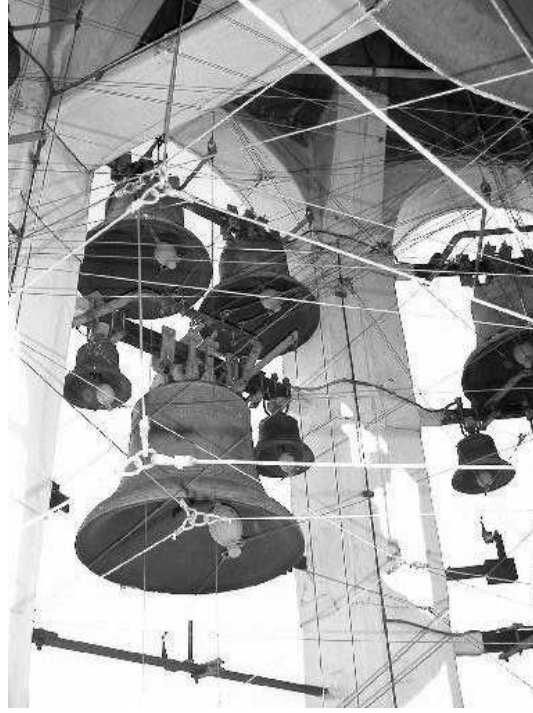


FIGURE 2. The carillon of the Zuiderkerk in Amsterdam

The displacement  $dK$  is roughly equal to 5 cm. The displacement  $dC$  obviously depends upon the size of the bell, and varies from 2-5 cm. Both distances are small compared to the length of the wires. The key is blocked from above to prevent it from moving upward since without this blocking the heavy clapper would move to its preferred, vertical position. If the clapper is *light* then a spring on the key returns the key to its rest position after it is depressed. The springs help somewhat to make the keys 'feel' the same while keeping  $dK$  fixed. In general, the only way to have similar feeling keys while keeping  $dK$  the same for all of them would be a system of springs at all keys, which would be hard to perfectly design.

**2.4. The configuration of the wires.** In a system with *one* broekring, the wire attaching the broekring to the clapper is called the *clapper-wire*, and we denote this as *wire 2* with length  $l_2$  and mass  $m_2$ . The wire from the fixed point on the wall to the broekring is called the *broek-wire*, this we denote by *wire 3* with a corresponding length  $l_3$  and mass  $m_3$ . Wire 1, the *key-wire*, goes down vertically from the broekring to the key and has length  $l_1$  and mass  $m_1$ . In the usual configuration all of the wires are in the *same plane* by the rigidity of the wires. Only if there is some physical obstruction (e.g. a beam or a bell) this will not be the case.

The wire configuration is best described by the angles between the wires. At the broekring there are three angles  $\angle(l_1, l_3)$ ,  $\angle(l_3, l_2)$ ,  $\angle(l_2, l_1)$  of which the third is the complement of the other two. We denoted  $\alpha = \angle(l_1, l_3)$  and  $\beta = \angle(l_3, l_2)$ . The

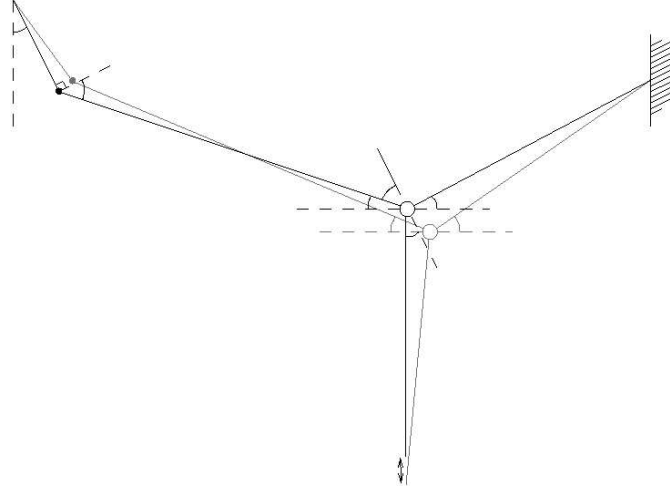


FIGURE 3. The one-broeking configuration in rest and after a vertical displacement  $dK$ . See figure 4 for the names of the angles between the wires.

angle between the clapper and the clapper wire is denoted by  $\gamma = \angle(l_2, l_0)$ . The angle between the clapper and the vertical at  $T$  is denoted by  $\sigma$ . If the key-wire is vertical then  $\alpha + \beta + \gamma = 2\pi + \sigma$ .

The order of which key is connected to which bell is (obviously) fixed, although the actual location of the bell is not. However, to work out the statics and dynamics of the above system we will assume that in this configuration the position of the bell is known, therefore, point  $T$  is known. Similarly, in a particular configuration we can fix the position of the keyboard and hence of the key  $K$  (to leading order). The angle of the clapper  $\sigma$  is known (to leading order) since we know the distance from the clapper to the bell and the shape of the bell. Given these constraints we need to determine the *unknown* position of the hinge  $H$  and of the broeking  $R$ . As the wires are in a plane, and we assume that the location of the bell and the key are known, this leads to a system with *two degrees of freedom* given by the heights of the hinge and the broeking.

### 3. The geometry of a single wire system

In this section we look further at the precise geometry of the one broeking system described above. We find simple equations for the changes in the angles of the wires when the key is depressed. We compute the ratio  $|dC|:|dK|$ , where  $dC$  represents the displacement of the clapper and  $dK$  represents the displacement of the key. We use the prefix  $d$  to emphasize that the equations are linearized, as discussed above in section 2.3. The ratio  $|dC|:|dK|$  depends on the angles at the broeking  $R$ . By manipulating these angles the displacement and the 'feel' of the key can be adjusted. Again consider the one broeking system as in figure 2.

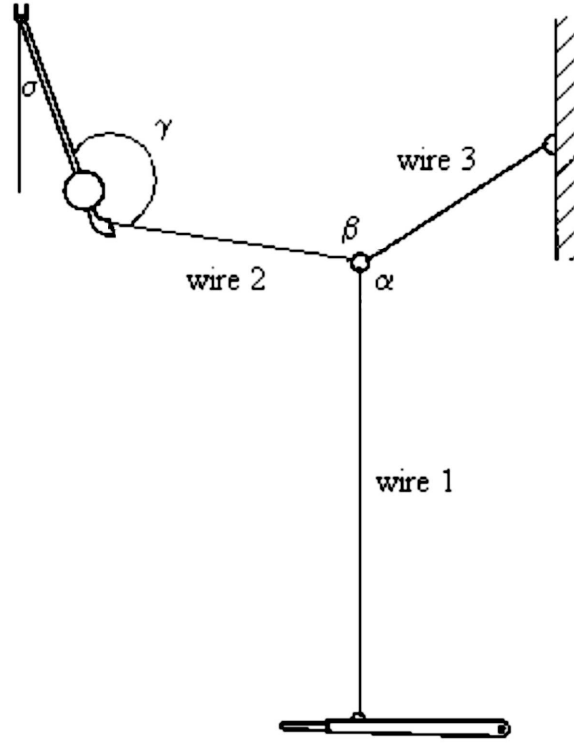


FIGURE 4. The angles in the wire system: the lengths of the wire 1, 2, and 3 are denoted by  $l_1$ ,  $l_2$ , and  $l_3$ , respectively, and the length of the clapper is denoted by  $l_0$ .

The positions of  $H$  and  $T$  are fixed. The positions of  $R$  and  $C$  change when  $K$  is depressed. In the end, what we want to know is how the depression of  $K$  affects the position of  $C$ . As the broekring is small, we can neglect the sliding of the wire through the broekring  $R$ , so the lengths  $l_i$  of the wires are constant. This gives the four equations

$$(1) \quad \begin{aligned} |T - C| &= l_0 \\ |R - K| &= l_1 \\ |C - R| &= l_2 \\ |R - H| &= l_3 \end{aligned}$$

The unknowns are the positions of  $R$  and  $C$ . The depression of a key is a relatively minor movement compared to the length of the wire, so we may linearize the equations about the equilibrium position. The linearization of a distance  $|A - B|$  is obtained from the inner product  $\langle A - B, A - B \rangle = |A - B|^2$ . We get four linear

equations

$$(2) \quad \begin{aligned} \langle T - C, dC \rangle &= 0 \\ \langle R - K, dR - dK \rangle &= 0 \\ \langle C - R, dC - dR \rangle &= 0 \\ \langle R - H, dR \rangle &= 0 \end{aligned}$$

If we assume that all points remain in a plane, then these are 4 linear equations with 4 unknowns, being the coordinates of  $R$  and  $C$ . The first equation and the last equation give the directions of  $dC$  and  $dR$ . The second and the third equation give the magnitudes of  $dC$  and  $dR$ . In particular,  $dC$  has an angle  $\gamma \pm \pi/2$  with the clapper wire, while  $dR$  has an angle  $\alpha \mp \pi/2$  with the key wire and an angle  $\beta \pm \pi/2$  with the clapper wire. The third equation implies that  $|dC|:|dR| = \sin \beta : \sin \gamma$ . The depression of  $K$  is vertical, directed along  $R - K$ , so the second equation implies that  $|dR|:|dK| = 1 : \sin \alpha$  and we find

$$(3) \quad \frac{|dC|}{|dK|} = \frac{\sin \beta}{\sin \gamma \sin \alpha}$$

In general the wire connection will not be in a plane and the coordinates of  $R$  and  $C$  comprise 6 unknowns. One can still compute  $R$  and  $C$  by geometric means, but it turns out that it is easier to use a force balance and that equation (3) remains valid even if the wire system is not in a plane. This we shall see in the section below.

In equation (3) we can influence the ratio  $\frac{\sin \beta}{\sin \alpha}$  by moving  $H$  up or down. If  $H$  moves up then  $\alpha$  increases and  $\beta$  decreases. Realistically  $\pi/2 < \alpha, \beta < \pi$ . So if  $H$  moves up the ratio  $\frac{\sin \beta}{\sin \alpha}$  increases while it decreases if  $H$  moves down.

**Rule of thumb.** *If you want to increase the ratio  $|dC|:|dK|$  then raise the hinge keeping the angle  $\gamma$  fixed or lower the position of the broekring while keeping the hinge fixed.*

Below we shall calculate the potential and kinetic energy of the wire system. For the potential energy the vertical components of the displacements are important, denoted by  $dR_2$  and  $dC_2$  (the displacement of the key is always vertical). The equations above imply that

$$(4) \quad |dR_2| = |dK| \text{ and } |dC_2| = \frac{\sin \sigma \sin \beta}{\sin \gamma \sin \alpha} |dK|$$

Note that this is only a linear approximation. In the real situation,  $dR_2 \approx dK$ , but it is slightly less by higher order ( $O(dK^2)$ ) corrections. We will need second order approximations once we consider the vibrations of the system.

#### 4. The periodic dynamics of a single wire system

As explained in Section 2, a carillon player prefers to play the carillon using approximately the same force for every key. The clapper weights vary from 0.3 to 30 kg, so there is a bit of a problem, even if we allow playing basses to be somewhat heavier than playing trebles. To understand which forces play an important role in the system, we first calculate the forces in a static system with massless wires. This can simply be done using Newton's laws. Incorporating the mass of the wires and the dynamics is done in a more general Lagrangian setting. For the computation

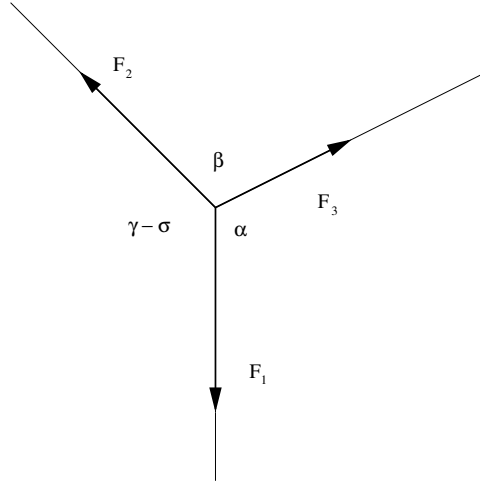


FIGURE 5. The force balance

of the forces and the energy we assume that the wires are rods, which is correct up to first order. For the computation of the vibrations, which is of second order, this assumption cannot be maintained.

**4.1. Static forces for massless wires.** We first calculate the force on the key if the clapper is in rest. This calculation generalizes equation (3) to wire systems with many broekrings. We assume that the mass  $m_c$  of the clapper is contained in  $C$  and that the wires are massless rods of constant length. A wire pulls at both its ends, being  $K, B, C$  or  $T$ , with forces that have equal magnitude but opposite sign by Newton's third law. We denote the force in the  $i$ -th wire by  $F_i$ .

At the broekring, the forces balance by Newton's first law as shown in figure 4.1, so

$$(5) \quad \frac{|F_1|}{|F_3|} = \frac{\sin \beta}{\sin \alpha}$$

At the clapper the balance of forces gives

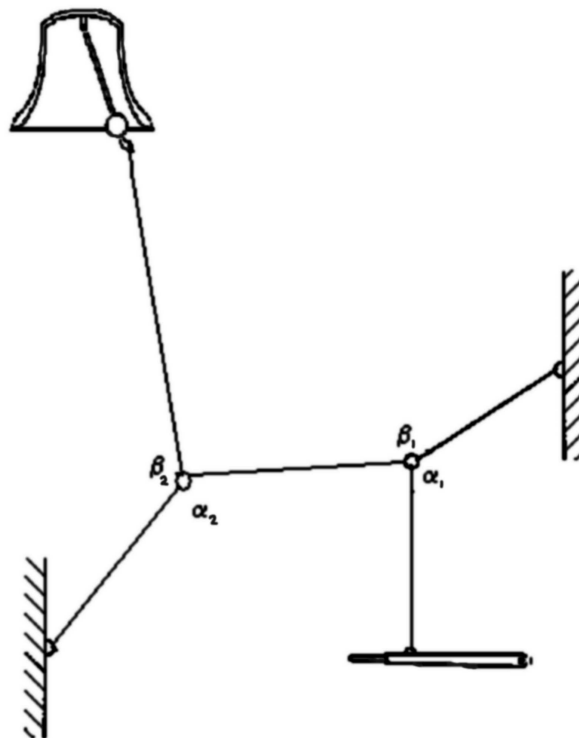
$$(6) \quad \frac{|F_3|}{|m_c g|} = \frac{\sin \sigma}{\sin \gamma}$$

Since  $|F_1|$  is the force on the key, we find that the clapper exerts a force on the key of

$$(7) \quad m_c g \frac{\sin \beta \sin \sigma}{\sin \alpha \sin \gamma}.$$

This force becomes singular in the cases  $\sin \alpha = 0$  or  $\sin \gamma = 0$ . Exactly the same singular case occurs when the wires do have mass, see also figure 7 below.

So far we carried out the calculations for wire connections that have one broekring only. It is possible to extend the calculations to systems that have many broekrings, as in figure 6.



**4.2. Configuration with more broekrings.** In a more general configuration, the wire connects the key to the bell by  $k$  broekrings  $R_i$ ,  $i = 1, \dots, k$ . Following exactly the same procedure as above, we obtain  $2(k+1)$  conditions from linearising the fixed distances  $|T - C|$ ,  $|C - R_k|$ ,  $|R_i - H_i|$  and  $|R_i - R_{i-1}|$ , where  $R_0 = K$  and the positions of  $C$  and the broekrings  $R_i$  are unknown. If we assume that, as before, all the points remain in one plane there are  $2(k+1)$  equations with  $2(k+1)$  unknowns (the coordinates of  $C$  and the  $R_i$ ). In this case,  $dR_i$  can be computed from  $dR_{i-1}$ . Let  $\alpha_i$  be the angle between the wire connecting  $R_i$  to  $R_{i-1}$  and the wire connecting  $R_i$  to  $H_i$ . And, let  $\beta_i$  be the angle between the wire connecting  $R_i$  to  $R_{i+1}$  and the wire connecting  $R_i$  to  $H_i$ . The angles  $\alpha_i$  and  $\beta_i$  are shown in figure 6 for two broekrings. Then the relation between the distances  $dR_i$  and  $dR_{i-1}$  over which the broekrings move, can be obtained from the above equations in a similar way as for the one broeking system, and it is given by  $|dR_i|/|dR_{i-1}| = \sin(\beta_{i-1})/\sin(\alpha_i)$ . We introduce the notation

$$(8) \quad w_i = \frac{\sin(\beta_{i-1})}{\sin(\alpha_i)}$$

and call this the *index* at the  $i$ -th broekring. For a system with  $k$  broekrings, the ratio  $|dC|:|dK|$  is the product over all the indices with an additional contribution of the angles at the clapper, according to the following formula

$$\begin{aligned} \frac{|dC|}{|dK|} &= \frac{|dC|}{|dR_k|} \cdot \frac{|dR_k|}{|dR_{k-1}|} \cdots \frac{|dR_2|}{|dR_1|} \cdot \frac{|dR_1|}{|dK|} \\ (9) \qquad &= \frac{1}{\sin(\alpha_1)} \prod_{i=1}^k w_i \end{aligned}$$

where we assume that the displacement of the key  $dK = dR_0$  is vertical. In the simple configuration with one broekring this again reduces to

$$(10) \qquad \frac{|dC|}{|dK|} = \frac{\sin(\beta)}{\sin(\gamma) \sin(\alpha)}.$$

As mentioned in Section 2, the fraction  $|dC|:|dK|$  is important for the design of the carillon and again we can influence it by moving the hinges

**Rule of thumb.** *If you want to increase the ratio  $|dC|:|dK|$  then raise the hinge of the first broekring keeping the angle  $\gamma$  fixed or lower the position of the first broekring keeping the hinge fixed.*

**4.3. The Lagrangian of the system with massive wires.** To find an expression for the force that is needed to play a key in case the masses of the wires are relatively large, we calculate the total energy for a single bell system, and the change in potential and kinetic energy under a small vertical displacement  $dK$  of the key.

The natural way to describe this mechanical system is by computing its Lagrangian. In order to obtain the total energy, we first determine the potential energy and the kinetic energy in every single point in the system and then integrate over the wires. We shall need second order approximations.

**4.4. Potential energy.** For the computation of the potential energy of the whole system, it is important to know where the gravitational force on each wire and on the clapper works, so where the center of mass is located. As the clapper is a rather thin pole with a large massive ball at the hanging end, we assume that all the mass of the clapper is located in this end. The wires are more or less uniform, so we can take their centers of mass in the middle of each wire. Suppose that the coordinates of the end points of a uniform wire are  $\underline{a} = (a_1, a_2)$  and  $\underline{b} = (b_1, b_2)$ . Then the potential energy of the wire is  $mg(a_2 + b_2)/2$ , where  $m$  denotes its mass. This applies to wires 2 and 3. We treat wire 1 in a different way, since both the weight of this wire and the force on the key approximately work along the same line in the vertical direction. The force on the key is always present: in rest, the key is blocked from above to prevent it from moving upward since without this blocking the rather heavy clapper would move to its preferred, vertical position. (Recall that the distance of the clapper to the bell in rest should be only 2 - 5 cm., so  $\sigma > 0$ .) When a bell is played, this force should first be overcome and then some additional force should be applied to move the clapper. We simply put this force and the gravitational force on the wire into one ‘tension’  $\tau$ , which in fact consists of



the tension in rest (directed upwards and canceled by the blocking force) and the (downward) tension added by the player.

The potential energy depends on the end points  $K, R, C, H$  of the wires. Denote the coordinates of  $K, R, H, C$  in the initial position by  $(K_1, K_2), (R_1, R_2), (H_1, H_2)$  and  $(C_1, C_2)$  respectively. Under the assumption of uniform mass distribution, the potential energy of the system in equilibrium is given by

$$V_0 = C_2(m_c + \frac{m_2}{2})g + R_2 \frac{(m_2 + m_3)g}{2} + H_2 \frac{m_3g}{2} + K_2\tau$$

Let  $V(dK)$  be the potential energy of the system, with the key in vertical position  $K_2 + dK$ , such that  $V(0) = 0$ . (This means that  $V(dK)$  is the relative potential energy with respect to equilibrium.) Under the above assumptions on the centers of mass the potential energy of the system is given by

$$V(dK) = dC_2(m_c + \frac{m_2}{2})g + dR_2 \frac{(m_2 + m_3)g}{2} + dK\tau$$

In equation (4) we find the linear dependence of  $dC_2$  and  $dR_2$  on  $dK$ . The linear part of the Lagrangian determines the equilibrium. We also need the quadratic part once we consider the vibrations of the system. Introducing the exact angles  $\alpha + d\alpha$ ,  $\beta + d\beta$ ,  $\gamma + d\gamma$  and  $\sigma + d\sigma$  after displacement, this quadratic part can be computed using geometrical constraints and Taylor expansions. We found that

$$\begin{aligned} V(dK) &= dK \left( (m_c + \frac{m_2}{2})g \frac{\sin \sigma \sin \beta}{\sin \gamma \sin \alpha} + \frac{(m_2 + m_3)g}{2} + \tau \right) \\ &+ \frac{dK^2}{2 \sin^2 \alpha} \cdot \left( m_c g \frac{\cos \sigma \sin^2 \beta}{l_0 \sin^2 \gamma} + m_2 g \frac{\cos(\sigma - \gamma) \sin^2(\sigma - \alpha)}{2l_2 \sin^2 \gamma} - \frac{m_3 g \cos \alpha}{2l_3} \right) \\ &+ O(dK^3) \end{aligned}$$

Note that, since RH is attached to the wall, the potential energy involving  $m_3$  must equal the potential energy for a simple rod hinged at  $H$ . This is indeed the case.

**4.5. Kinetic energy.** Consider a rod of uniformly distributed mass  $m$  that is moving. Its movement can be described by the coordinates of its end points  $\underline{a}, \underline{b}$  depending on time. If we denote the derivative with respect to time of the coordinates  $a$  and  $b$  by  $\dot{a}, \dot{b}$ , then the kinetic energy of the rod is

$$\frac{1}{2}m \int_{t=0}^1 |t\dot{\underline{a}} + (1-t)\dot{\underline{b}}|^2 dt = \frac{m(|\dot{\underline{a}}|^2 + \langle \dot{\underline{a}}, \dot{\underline{b}} \rangle + |\dot{\underline{b}}|^2)}{6}.$$

So the kinetic energy of the broek-system is given by

$$\frac{m_1(|\dot{K}|^2 + \langle \dot{K}, \dot{R} \rangle + |\dot{R}|^2)}{6} + \frac{m_2(|\dot{R}|^2 + \langle \dot{R}, \dot{C} \rangle + |\dot{C}|^2)}{6} + \frac{m_3|\dot{R}|^2}{6} + \frac{m_c|\dot{C}|^2}{2}.$$

In section 3 we computed the relations between  $dK, dR$  and  $dC$  and we found that

$$\begin{aligned} |dR| &= |dK|/\sin \alpha, & \angle(dR, dK) &= \alpha - \frac{\pi}{2} \\ |dC| &= |dK| \frac{\sin \beta}{\sin \alpha \sin \gamma}, & \angle(dR, dC) &= \beta + \gamma \end{aligned}$$

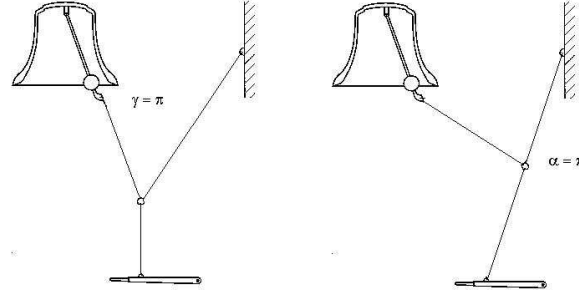


FIGURE 7. Two singular cases. When the key-wire is vertical, the brook-wire and key-wire hang along the wall in the right hand case

Denote the ratio  $|dR|/|dK|$  by  $\rho_R$  and the ratio  $|dC|/|dK|$  by  $\rho_C$ . Then using relations as  $\dot{R} = \dot{K}dR/dK$ , the kinetic energy in terms of  $\dot{K}$  is

$$(11) \quad T(K) = \frac{1}{6}|\dot{K}|^2 (m_1(1 + \rho_R \sin \alpha + \rho_R^2) + m_2(\rho_R^2 + \cos(\beta + \gamma)\rho_R\rho_C + \rho_C^2) + m_3\rho_R^2 + 3m_c\rho_C^2).$$

**4.6. Lagrangian motion.** With the obtained  $T$  and  $V$  we can write down the Lagrangian  $\mathcal{L} = T - V$  of the system, and hence the equations of motion

$$\frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{K}} - \frac{\partial \mathcal{L}}{\partial K} = 0.$$

Here  $\frac{\partial \mathcal{L}}{\partial K}$  can be seen as the force in the system, which derives in this case from a potential. In equilibrium, the force acting on the system should be equal to zero, meaning that  $\frac{\partial \mathcal{L}}{\partial K} = -\frac{dV}{dK} = 0$  in equilibrium. This immediately gives a relation for  $\tau$ , since it means that the coefficient of the linear part of  $V$  should be zero. The  $\mathcal{O}(dK^2)$  terms in  $V$  have to balance with  $T$ . From this we can derive natural frequencies of the system, that can give a clue about the amount of vibration one can expect in the wires.

We will do these calculations in a system with massless wires. In this case the system should more or less satisfy the ‘classical’ intuition; therefore it is a good check for our calculations.

**4.7. Dynamics of massless wires.** In case of massless wires only terms concerning the clapper remain. From  $\frac{dV}{dK} = 0$  we derive in this case

$$(12) \quad \tau = m_c g \frac{\sin \sigma \sin \beta}{\sin \gamma \sin \alpha}$$

which is exactly the same expression as (7). This becomes singular in the cases  $\sin \alpha = 0$  or  $\sin \gamma = 0$ . Exactly the same singular cases appear in the full expression for  $\tau$  when the wires do have mass. From the sketches 7 it is immediately clear that

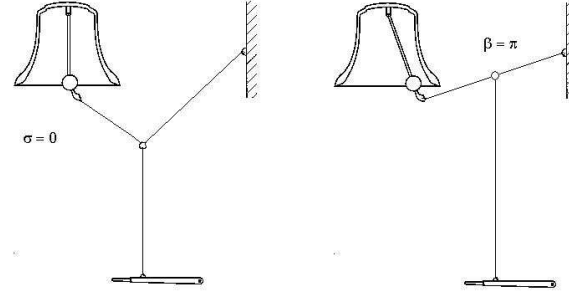


FIGURE 8. A massless system that does not exert a force on the key

it should indeed cost an infinite amount of work to move the key or clapper when the configuration satisfies either case.

In this massless case it is also easy to understand how the system should behave when there is no tension at all, so when the key is not blocked or played. The system then just hangs, and one would expect that the clapper would hang in its preferred, vertical position or would somehow not be able to reach that position. This indeed follows from the assumption  $\tau = 0$ : the solutions are  $\sin \sigma = 0$  and  $\sin \beta = 0$ , leading to the configurations in figure 8.

**4.8. Swinging and vibration.** In reality a wire is not a rod but it is an elastic string. The key strokes induce displacements of the broekrings, causing vibrations. This vibration should not be too much, since the wires are, in an ideal situation, not allowed to touch each other. To get an idea about the possible vibration, we derive the natural frequencies of the system from the Hamiltonian  $H = T + V$ . These natural frequencies tell us what happens if you *twang* a wire.

For massless wires, putting  $H = 0$  and comparing equations 4.4 and 11 yields the differential equation

$$\dot{K}^2 = \frac{g \cos(\sigma)}{l_0} dK^2.$$

Recall that the coefficient of the linear term in  $V$  is zero. If we put  $x = e^{i\omega t}$ , this results in

$$\omega = 2\pi f = \sqrt{\frac{g \cos(\sigma)}{l_0}}.$$

Here  $f$  is the frequency. For typical values of the parameters, say  $l_0 = 0.5$ ,  $\cos \sigma = 0.8$ , this results with  $g = 10$  in  $\omega = 4$ , or a frequency of about 0.6 Hz. From our observations in the church we estimated a frequency of 0.2 Hz.

The same kind of calculations can be done for wires with mass, for which one ends up with more complicated expression that we do not display here.

When playing the carillon, the vertical wire will always move a bit. But since the vertical wires are all attached to the keyboard, they are very close to each

other, even when they are in rest. This means that a small sideways movement can already cause touching wires, which we should prevent.

It is difficult to give a precise bound on the amplitude, since this will depend on the geometry, the tension and the number of key strokes during play. However, we observed that the maximal amplitude of the vibration occurs at the first broekring  $R_1$  at the end of the vertical first wire segment. Its displacement twangs the wire, as represented by the horizontal component of  $dR_1$ . When the vertical movement of the wire is  $dK$ , the sideways swing of the wire at the position  $R_1$  is approximately  $dR_1 = dK / \tan \alpha$ , which means that the configuration should optimally satisfy  $\alpha = \pi/2$ .

**Rule of thumb.** *The vibration of the first broekring can be damped by putting the hinge at the same level as the ring.*

**4.9. Playing the carillon.** We have already addressed some rules of thumb for the angles in the system, that follow from conditions on the swing when playing the carillon. However, we did not mention the requirement that is maybe the most important one: the keyboard player roughly wants every key to ‘feel’ the same. The basses may need some more force to play than the trebles, but the amount of force that is needed to play neighbouring tones should not differ much. We suppose that the travel  $dK$  is constant for all keys, i.e., that the difference between the upper and the lower position is the same for each key. The question now is: under what conditions is the tension that the player should add to play a bell (about) the same for each key?

A first idea can again be obtained from the system with massless wires. We assume that there is a weight  $W$  that ‘helps’ the player to move the key down (which is in our former calculations incorporated in  $\tau$ ). This weight may contain the mass of the wires, that we assume to be concentrated at the key, if not negligible. Since the total energy should be remained, this means that the following should be satisfied:

$$\text{Energy input by player} = \text{potential energy gain of clapper} - \text{energy lost by help.}$$

Note, that the clapper moves upward when the key is played, while the helping weight moves downward. This means that the clapper gains potential energy, while the weight loses potential energy. In a formula we have

$$(13) \quad \tau_{\text{player}} dK = m_c g l_0 \sin \sigma \cdot d\sigma - W dK.$$

If each key should need the same amount of added tension, this means that  $\tau_{\text{player}} dK$  should be constant. In other words, for  $W = 0$  we obtain the restriction

$$(14) \quad m_c l_0 \sin \sigma \cdot d\sigma = C,$$

where  $C$  is the same constant for all bells!

However, detailed information about the St. John’s tower in Gouda tells us that this requirement is impossible to satisfy. There the specification of the bells gives:

large bell  $l_0 \sin \sigma \approx 50 \text{ cm}$ ,  $m_c \approx 30 \text{ kg}$   
 small bell  $l_0 \sin \sigma \approx 10 \text{ cm}$ ,  $m_c \approx 0.3 \text{ kg}$ .

To impose (14) for this tower, the angles of swing in the smallest and the largest bells should satisfy

$$d\sigma_{\text{large}} = 50 \cdot d\sigma_{\text{small}},$$

which is impossible (it would for instance mean that a small bell would move by 1 degree and a large one by 50 degrees). This actually tells us, that the helping weight  $W$  is really needed at the larger bells to obtain an evenly played keyboard.

These calculations nicely correspond to our observations. In a real carillon system the wires are always very long and quite thick, and have a total mass of up to 1 kg. This means that the mass is not negligible for small bells with clapper weights of only 0.3 kg. The mass of the wires that we put in  $W$  may even be so large compared to  $m_c$ , that the right hand side of (13) is negative. This would mean that help is needed to move the clapper back to its original position, which is indeed the case in real carillons. We observed that some of the smaller clappers had a spring to move them back after they had been played.

On the other hand, we observed springs that helped to move the keys of larger bells, which could account for the extra weight  $W$  needed in order to obtain an evenly played keyboard.

## 5. The 3D configuration of the broek-system

There are two interesting optimization problems related to the placement of bells in a tower. In the first problem we view the tower from above and we try, mainly to get an idea of the situation, to find a configuration such that the intersection points (of the projections) of the wires are as far apart from each other as possible. Of course there is no guarantee that this configuration is the best in the three dimensional case.

In the second optimization problem we try to approximate the three-dimensional real situation in a belfry. Then we would like to maximize the smallest distance between all the wires so that they're as far apart from each other as possible to avoid that they will touch when the carillon is played.

**5.1. The two-dimensional problem.** Given the possible positions of the bells ( $\{b_i \in \mathbb{R}^2 \mid i = 1, 2, \dots, n\}$ ) on the circumference of the tower and the position of the attachment points to the keys ( $\{k_i \in \mathbb{R}^2 \mid i = 1, 2, \dots, n\}$ ) in the keyboard, the place of each wire, being a straight line through these two points, is uniquely determined. If we start with  $n$  keys and  $n$  bells, then the wires will intersect in (at most)  $\frac{1}{2}(n-1)n$  points. Our aim is to design the configuration of the bells in such a way that these intersection points lie as far apart as possible. Therefore, we first look at the distances between all these points. There are

$$\binom{\frac{1}{2}(n-1)n}{2}$$

distances between the intersection points which need to be computed. Among these distances we have to find the smallest distance and then the optimization problem is to decide which positioning of the bells maximizes this minimal distance. A

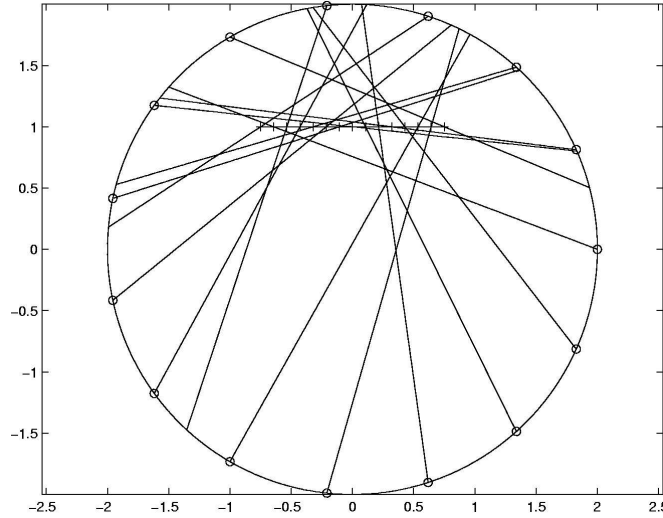


FIGURE 9. A belfry with a radius of 2 m with 15 bells seen from above. The distance between the intersections is never less than 1.31 cm.

computer program which tries different permutations of the bells randomly can approximate the maximum. However, this is a quite cumbersome task since for  $n$  bells there are  $n!$  configurations that need to be checked. Thus for one level in a belfry with 15 bells, it will take  $0.714 \cdot 10^{16}$  computations, and for a belfry with 45 bells the number of computations will be  $0.586 \cdot 10^{62}$ .

Computations with Matlab yield for a belfry with 15 bells and a radius of 2 m after a 698 tries, a configuration where each distance between two intersections is greater than 1.3 cm (see figure 9).

Notice that there are two more general interesting geometrical questions connected to this problem. The first question, important to obtain a good estimate in the optimization problem, is the following. How should  $n$  points be placed in a unit disk, such that the smallest distance between two points is maximal? For  $n = 2, 3, 4, 5, 6$ , this distance is  $2, \sqrt{3}, \sqrt{2}, \sqrt{(5 - \sqrt{5})/2}, 1$ , respectively, and the points have to be placed on the boundary. For  $n = 7$  the maximal distance is also 1, but one point has to be placed in the interior of the disk. But what about higher  $n$ ?

The second question is the following: consider two sets of  $n$  colored points in the real plane,  $\{b_i \in \mathbb{R}^2 \mid i = 1, 2, \dots, n\}$  (blue, the keyboard) and  $\{k_i \in \mathbb{R}^2 \mid i = 1, 2, \dots, n\}$  (black, the bells). How can we connect each of the blue points with a different black point with a straight line such that the distance between the intersection points is as large as possible.

**5.2. The three-dimensional problem.** Let us now consider the belfry in three dimensions. So, assume we can place bells in an equidistant way on the

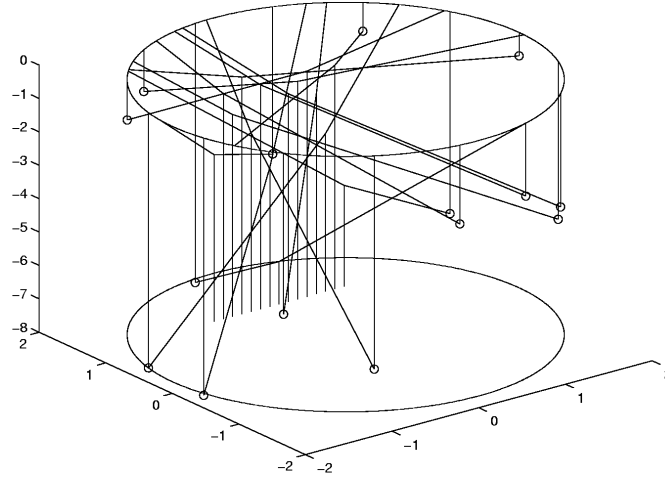


FIGURE 10. A belfry with a radius of 2 m with 15 bells. The distance between the wires is greater than 2.95 cm.

circumference of the belfry, where the height is free to choose. Also, we assume that the wires which connect the broekring to the wall are all attached to the wall at a fixed height on the circumference of the belfry. Then we find that once the horizontal position of the bell is chosen, the vertical position of the bell is completely determined by the formulas

$$\begin{cases} \frac{\sin \beta}{\sin \gamma \sin \alpha} = \frac{|dC|}{|dK|} \\ \alpha + \beta + \gamma - 2\pi = \sigma \end{cases}$$

because  $|dC|/|dK|$  and  $\sigma$  are fixed (see figure 4). Starting with an arbitrary permutation of the bells, we compute the height of each bell by the above formulas, and then we compute all the distances between the pieces of wire. We would again like the smallest of these distances to be as large as possible. It is astonishing that it is hard to find good configurations by a random permutation of the bells. After 1112 permutations of 15 bells in a belfry with a radius of 2 meters we find that for the optimal configuration the smallest distance is 2.95 cm (see figure 10). The geometrical character of this problem is similar to the two-dimensional case.

### Acknowledgments

A large part of this work was done during the Study Group with Industry in February 2003 in Leiden, in cooperation with the group members Miguel Frasson, Nick Ovenden, Nicole Rommelse and Floske Spiekstra.

**Bibliography**

- [1] André Lehr *Some mathematical background of the broek-system* (in Dutch), Nationaal Beiaardmuseum, 2003.