Living near an Airport,

Risky or just Annoying?

Spatial Variability of the Risk of Ground Fatalities due to Crashing Aircraft in the United States

R. Frank Rabouw

Harvard Center for Risk Analysis

Delft University of Technology

Living near an Airport,

Risky or just Annoying?

Spatial variability of the risk of ground fatalities due to crashing aircraft in the United States.

R. Frank Rabouw

October 2000

Section Control, Risk, Optimization, Stochastics and Systems
Faculty of Information Technology and Systems
Delft University of Technology

Harvard Center for Risk Analysis

Delft University of Technology

Preface

This thesis concludes the Master's program of Technical Mathematics at Delft University of Technology. It is the result of research conducted at the Harvard Center for Risk Analysis in Boston, United States. The Center is part of the Harvard School of Public Health and its mission is to promote reasoned public responses to health, safety and environmental hazards.

This report is interesting for people who are active in the field of risk communication or risk management and readers who are concerned with safety around airports.

Several people helped me to complete this report and all of them deserve my sincere thanks. In the first place, I would like to thank Kimberly Thompson for her great support in all dimensions and for giving me the opportunity to work at the Center for 10 months. Her inspiration, ideas and comments were crucial for completing this thesis. I also would like to thank Roger Cooke for all his comments and ideas and for having enough faith in me to send me to Boston. Further, I would like to thank all my colleagues at the Center for Risk Analysis for all the support I got and all the fun we had during my 10-month stay. Finally, I would like to thank the following organizations for their financial support: Fundatie van de Vrijvrouwe van Renswoude, Koninklijk Instituut van Ingenieurs, Universiteitsfonds Delft, Faculty of Information Technology and Systems and Delft University of Technology.

Delft, October 6th, 2000
Frank Rabouw

Abstract

Recent airplane crashes that caused ground fatalities, like the Concorde crash in Paris, France on July 25th of the year 2000 and a DC-9 crash in Tamaulipas, Mexico on October 7th, 2000, confirm that people on the ground are exposed to some risk of dying from a crashing airplane. In 1992, Goldstein et al quantified this risk for Americans as an annual risk of 6 in hundred million and a 70-year lifetime risk of 4.2 in a million. They noted that the risk exceeds the commonly used risk management threshold of 1 in a million lifetime risk and suggested to use the risk to ‘groundlings’ as a risk communication tool because of its characteristics of being an uncontrollable, technological risk. Subsequently, their estimate was used for risk communication purposes many times. Goldstein et al presented the risk as completely random and ignored variability of the risk in any dimension.

This study reviewed recent aviation accident data and updated the annual risk of ground fatalities to 1.3 in a hundred million or a 1 in a million 75-year lifetime risk. Further, in this analysis the spatial variability of the risk associated with the dimension distance to an airport was quantified. A Geographical Information System was built to relate the American population to airports and crash locations were reviewed to analyze where ground fatalities occur. A probabilistic approach combined both aspects and showed that the exposure to the ground fatality risk is approximately a factor 100 higher in the vicinity of a busy airport. An uncertainty analysis was performed and determined that uncertainty was a less important factor than variability.

This study showed that estimates of the current risk of ground fatalities due to crashing airplanes does not exceed the threshold for lifetime risks of 1 in a million anymore. Further, the variability of the risk in the dimension distance to an airport was quantified as approximately a factor 100 and the variability mainly applies to the first two miles around an airport. The results of this data-based approach were compared to more scenario-based models (RAND, 1993) and estimates provided by this analysis are significantly lower.

Table of Contents

List of Definitions

Accident (NTSB)	An occurrence associated with the operation of an aircraft that takes place between the time any person boards the aircraft with the intention of flight and all such persons have disembarked, and in which any person suffers death or serious injury, or in which the aircraft receives substantial damage.
Airport-related accident	An accident is related to an airport if the airport is the origin or the final destination of the flight and the accident occurred within 10 miles of the airport.
Airport-unrelated accident	Accident that is not airport-related
Air carrier	An aircraft with a seating capacity of more than 60 seats or a maximum payload capacity of more than 18,000 pounds carrying passengers or cargo for hire or compensation. This includes US and foreign flag carriers.
Air taxi	An aircraft designed to have a maximum seating capacity of 60 seats or less or a maximum payload capacity of 18,000 pounds or less carrying passengers or cargo for hire or compensation.
General aviation	All civil aircraft, except those classified as air carriers or air taxis.
Groundling accident	Accident that caused at least one ground fatality who had little or no control over the aviation activity that caused his or her death.
Ground fatality due to uncontrollable aviation accident	Ground fatality who had no or little control over the aviation activity that caused his or her death.
Group risk	See population risk
Third party risk	Individual with little or no control over their exposure are those at third party risk.
Population risk	The expected number of fatalities in a certain population in a certain time period.

List of Acronyms

FAA	Federal Aviation Administration
FAR	Federal Aviation Regulations
GEOCORR	Geographic Correspondence Engine
MABLE	Master Area Block Level Equivalency
NTSB	National Transportation Safety Board
TAF system	Terminal Area Forecasting system

1 Introduction

1.1 Background

People are exposed to numerous risks with different magnitudes and characteristics. Acceptability of a risk depends on both its magnitude and character. For example, involuntary risks are regarded as worse by the public than voluntary risks. The risk of ground fatalities associated with crashing airplanes has the characteristics that it is an involuntary, uncontrollable, technological risk and public tolerance versus these risks is relatively low. The European Transportation Safety Council (ETSC) spoke out their concern about the risk to residents in the vicinity of major airports in a recent briefing[1]. The council concluded that at the same time air traffic is rapidly increasing and public tolerance for environmental effects of air traffic such as noise, air pollution and risk of ground fatalities appears to be decreasing.

On December 5th, 1997 an airplane crashed into a department store in Irkutsk, Russia, killing 45 people on the ground. More recently, all 109 passengers and 4 people on the ground were killed when a Concorde collided with a Hotel in Paris, France, on July 25th of the year 2000. And on October 6th of 2000, 6 ground fatalities occurred in Tamaulipas, Mexico when a DC-9 made an overshoot landing and hit several homes. These recent airplane crashes serve as reminders of the fact that people on the ground are exposed to some risk of being killed by a crashing airplane. Goldstein et al (1992) were the first to quantify this risk for Americans.

In 1992, Goldstein et al[2] estimated the individual 70-year lifetime risk of ground fatalities due to a crashing airplane in the United States. They performed a risk assessment using data provided by the National Transportation Safety Board which recorded 150 ground fatalities due to airplane accidents in the 11-year period from 1975 to 1985 (resulting in an average of 13.6 ground fatalities per year). Goldstein et al, 1992, used the U.S. resident population of 1980 (227,757,000) to estimate the individual risk of ground fatalities:

Annual risk:

$$\frac{13.6}{227,757,000} = 6.0\,10^{-8}$$

Corresponding 70-year lifetime risk (assuming that the risk remains constant):

$$6.0\,10^{-8} \times 70 = 4.2\,10^{-6}\,.$$

Goldstein et al suggested using this estimate as a risk communication tool, because dying due to a crashing plane as a 'groundling' is not controllable, voluntary or natural, and passing airplanes do not provide direct benefits to those who are at risk on the ground. Further, they noted that with respect to this

---

[1] European Safety Transportation Council, *Safety In and Around Airports,* Brussels, 1999.

[2] Goldstein, B.D., Demak, M., Northridge, M., Wartenberg, D., *Risk to Groundlings of Death Due to Airplane Accidents: A Risk Communication Tool,* Risk Analysis, Vol 12, No. 3, 1992.

risk, regulatory action is warranted, such as airplane control and traffic, but that a significant change of behavior to reduce the risk, such as living in the basement, does not make sense considering the magnitude of the risk.

Goldstein et al argued that the risk of death for groundlings due to plane crashes is a catastrophic and unexpected event and they noted that the estimate of the 70-year lifetime risk exceeds one in a million, a commonly used threshold in regulatory risk management. Subsequently, their estimate for the risk of ground fatalities due to a crashing airplane has been used for risk comparison purposes many times [3,4].

Goldstein et al based their estimate on data from 1975-1985, but accident and fatality rates have been significantly decreasing due to the introduction of new technologies. On the other hand, the air traffic has increased and the airspace has become more congested, especially around airports. This raises the question of whether their estimate is still applicable. An analysis of recent accident data is needed to update the estimate.

Another interesting issue concerns the variability of the risk to groundlings. A review of crash data published by the Boeing Company[5] showed that the 70% of commercial jet crashes occurred in the airport-related phases of the flight suggesting that people who live near an airport are at higher risk. Goldstein et al qualitatively addressed the issue of variability but their analysis did not consider or quantify the spatial variability of the risk for groundlings. However, high variability of the risk can undermine its use for risk communication purposes.

There are models that evaluate the risks around a single airport and quantify the spatial variability of the risk associated with this airport. On October 4th, 1992 an El Al freight carrier crashed into an apartment building in Amsterdam, Netherlands, killing 43 people on the ground. As a result of this crash the Dutch authorities asked RAND to undertake an evaluation of the external risks of Schiphol airport. The RAND company developed a model[6] that evaluated the risk of ground fatalities in the vicinity of Schiphol airport and concluded that Schiphol was a safe airport. However, they mentioned that airport external risk assessment is not a well-developed science and that further research and better data collection are needed to reduce uncertainty of the risk estimates.

Estimates of the risk of ground fatalities in the United States are outdated and do not consider any variability. This study updates the risk of ground fatalities in the United States and relates the risk to airports. At the same time, it compares the results with existing models that evaluate the risk of ground fatalities around a single airport.

---

[3] Wilson, R., *Regulating Environmental Hazards*, Regulation, Vol 23, 2000.

[4] Harvard Center for Risk Analysis, *Cellular phones and driving, weighting the risks and benefits,* Risk in Perspective, July 2000.

[5] Boeing, *Statistical Summary of Commercial Jet Airplane Accidents, Worldwide Operations 1959-1999*, 1999.

[6] Hillestad, R., et al, *Airport Growth and Safety, a Study of the External Risks of Schiphol Airport and Possible Safety-Enhancement Measures.*, RAND, 1993.

1.2 Objectives

The objectives of this analysis are to:

a) estimate the current risk of ground fatalities due to aviation accidents in the United States based on recent accident data,

b) quantify the spatial variability and uncertainty of this risk,

c) discus the consequences of the variability and uncertainty of the risk for its use as a risk communication tool.

1.3 Risk Measurement

Risk is commonly defined as the product of the probability of occurrence of an event and the consequences of that event. The consequences can be measured in terms of fatalities, injuries with varying degrees of severity, loss of life years or loss of Quality Adjusted Life Years (QALY`s). Consistent with previous studies[2,6,7,8], this analysis applies fatality as the measure of risk.

Several terms have been used to name the risk of ground fatalities around airports. The three studies that were performed to evaluate the risk of ground fatalities around Schiphol airport talk about *external risk* and RAND introduced the term *third-party* risk. Only ground fatalities with a special character are considered to be fatalities of external or third party risk. RAND used the following definition for their analysis:

Individuals with little or no control over their exposure are those at third-party risk.

This analysis used the same definition and the following criterion is applied to distinguish ground fatalities of different nature:

Ground fatality criterion:
Ground fatalities who had little or no control over the aviation activity that caused their death are fatalities due to third-party risk

Only ground fatalities that meet this criterion are considered when talking about the risk for groundlings of death due to crashing airplanes. These ground fatalities are denoted in this thesis as:

Ground fatalities due to uncontrollable aviation accidents.

Note that with respect to the risk at stake controllability and voluntariness are basically equivalent characteristics.

---

[7] Piers, M.A. et al, *Study of the external risk around Schiphol airport*, National Aerospace Laboratory of The Netherlands, 1993.
[8] Smith, E., Spouge, J. *Risk Analysis of Aircraft Impact at Schiphol Airport*, Technica Consulting, May 1990.

This survey considers both individual and group risk. Group risk or population risk is defined as the expected number of fatalities in a specific group or population in a certain time period. Goldstein et al estimated the annual population risk due to aviation accidents as 13.6 ground fatalities in the United States. They estimated the individual risk as 6 in a hundred million per year. Note that the occurrence of ground fatalities due to aviation accidents is a stochastic process and the annual number of ground fatalities fluctuates.

Goldstein et al used the individual 70-year lifetime risk as a metric. However, risks related to aviation have been changing due to the introductions of new technologies, increase of aviation activity and improving safety procedures. Consequently, it is not reasonable to assume that the groundling risk will remain constant for the next 70 years. This raises the question whether the 70-year lifetime risk is a legitimate metric for the risk to groundlings (Note that it is almost impossible to generate reasonable projections of the risk to groundlings for a 70-year time horizon). This survey focuses on the annual risk and sometimes speaks about the corresponding 70-lifetime risk to compare the results with Goldstein's analysis. The corresponding 70-year lifetime risk assumes that the risk will remain constant for the next 70 years.

1.4 Risk of ground fatalities as a Risk Communication tool

When people try to evaluate or estimate uncertainties or probabilities, like risks, they use a set of heuristics. These intuitive rules might do well in some circumstances, they sometimes can lead to a biased perception[9,10]. In evaluating the frequency of a certain event, people often rely on the heuristic procedure availability. This means that their judgment of the probability or risk is driven by the availability of previous occurrences of the considered event. Although catastrophic air crashes might be rare, all accidents are widely broadcasted by the media. On the other hand, deaths from cancer are common but people only observe these casualties when it concerns celebrities, friends or family. Lichtenstein et al conducted an experiment that confirmed that use of the heuristic availability can lead to a biased perception, people in general overestimate the risks of rare events and underestimate risks of more frequently occurring events.

Besides that the public misjudges the real risk due to the use of heuristics as availability, the public response is not consistent with numerical levels of risks because some types of risk are regarded as worse than others. For example, involuntary risks are regarded as worse than voluntary risks and risks that are not a function of our own skills are more frightening than risks we can control. Consequently, regulatory agencies should be aware of the nature of the risk at stake when they define thresholds.

Goldstein et al noted that the public manifests concern well out of proportion to risks associated with catastrophic and unexpected events, like an airplane crashing into one's house, compared to levels of risk due to more familiar causes (see Table 1.1). They suggested the following criteria for plausible risk comparison for these types of risk:

1. The risk is low- so low that we generally do not consider ourselves with risk at this level during our daily lives.

---

[9] Morgan, M.G., Henrion, M., *Uncertainty, a Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*", 1990.
[10] Cooke, R.M., *Experts in Uncertainty*, 1991.

2. Though we are not in control of the event that leads to the risk, we can reduce the exposure by adjusting our lifestyles. However, the magnitude of the risk does not warrant the change in lifestyle and it seems ridiculous to do so.

3. We do not derive direct benefit of the activity that is responsible for the risk.

4. The risk concerns no act of nature.

According to Goldstein et al, the risk to groundlings of death due to a crashing airplane meets these criteria. They suggest that the risk can be used as an effective risk communication tool for comparison with chemical and nuclear hazards. However, their estimate of the risk exceeds the commonly used threshold for these types of risk, one in a million 70-year lifetime risk, and no action by regulatory agencies is taken to control the ground fatality risk. Consequently, using the risk of ground fatalities due to crashing airplanes as a comparison tool would imply inconsistency with the one in a million threshold for risks associated with chemical factories and other risks that meet the four criteria.

Criterion 3 is a little confusing, because it is not clear whether it does apply to the ground fatality risk. People on the ground seem to derive no direct benefits of the air transportation, but on the other hand if you travel by air, and a significant part of the U.S. population does travel by air sometimes, you do benefit in some way. Note that criterion 2 is equivalent with the ground fatality criterion that is used in this analysis (see Section 1.3).

Action or type of exposure	75-year lifetime risk
All cancers	1 in 4
Cigarette smoking	1 in 3
Motor vehicle accident (motorist or passenger)	1 in 80
Motor vehicle accident (pedestrian)	1 in 400
Home accidents	1 in 120
Electrocution	1 in 3,000
Being hit by a meteorite	1 in 25,000
Being hit by a falling aircraft	1 in 200,000

Table 1.1 Some commonplace Public Risks (Table obtained from Wilson R., 2000)

1.5 A data-based approach

The studies that were performed to quantify the external risk of Schiphol airport used an analytical approach. These studies developed rather descriptive models that focused on the local circumstances of the airport, like airport activity and population densities in surrounding areas, but on the other hand used general accident rates and crash distributions. These general accident rates need to be used when a single airport is evaluated, because fortunately the number of crashes that occurred at Schiphol airport is too small to generate reasonable estimates of accident rates. The model that evaluates the risk to groundlings in the vicinity of Schiphol airport is discussed in Section 5.3.

However, this analysis focuses on the risk to groundlings in the whole United States and therefore more accident data are available. A significant effort was put in reviewing U.S. aviation accident data to derive groundling accident rates and ground fatality rates, and crash distributions. On the other hand local circumstances at single airports were ignored and only the average groundling risk associated with a group of airports was evaluated.

The analysis involved the following steps:

1. The derivation of a mathematical formulation of the risk and the variability of the risk (Chapter 2).

2. A study of historical aviation accidents in which ground fatalities occurred was performed to derive (Chapter 3):
    a) groundling accident and fatality rates,
    b) crash distribution functions.

3. A Geographical Information System model was built to relate the U.S. residence population to a group of airports (Chapter 4).

4. Fatality rates, crash distribution functions and the GIS model were combined to quantify the variability and uncertainty of the risk for groundlings (Chapter 5).

This analysis only focused on civil aviation accidents, risks of ground fatalities due to crashes of military aircraft are excluded.

2 Mathematical Formulation of the Variability of the Risk

This Chapter discusses the mathematical background of the problem. In Section 2.1 and 2.2 an expression that defines the spatial variability of the risk in the dimension distance to an airport is derived using basic probability theory and probabilistic reasoning. Section 2.3 discusses how this expression is used in the following chapters to determine the variability.

2.1 Bayes' theorem

The U.S. residence population on July 1st, 2000 is approximately 275 million. Each resident is exposed to numerous risks of different magnitude. This analysis focuses on the risk of ground fatalities due to aviation accidents (only ground fatalities of people who had little or no control over the aviation activity that caused their death are taken into account). Consider a random U.S. resident X and define the following event:

$B_{[t,t+\Delta t)}(X)$: {X dies due to an uncontrollable[1] aviation accident in [t, t+Δt)}

The probability $P(B_{[t,t+\Delta t)}(X))$ represents the risk of death for individual X due to an uncontrollable aviation accident in the time interval [t, t+Δt). As mentioned in the introduction, airplanes are more likely to crash in the vicinity of an airport and consequently, individuals who spend most of their time close to an airport are at higher risk. This analysis quantifies the variability of the risk associated with the dimension distance to an airport. The following stochastic quantity defines the behavior of individual X with respect to this dimension:

D(X,t) : Distance between individual X and the nearest airport at time t.

D(X,t) is a stochastic process parameterized by time.
The quantity $P(B_{[t,t+\Delta t)}(X)\,|\,D(X,\tilde{t}) : \tilde{t} \in [t, t+\Delta t))$ represents the risk for individual X in [t, t+Δt) given its distance to the nearest airport in that time interval. Consider a time interval Δt and distance interval Δd and define the following event:

$A^{[d,d+\Delta d)}_{[t,t+\Delta t)}(X)$: $\{D(X,\tilde{t}) \in [d, d+\Delta d) : \tilde{t} \in [t, t+\Delta t)\}$

Then, $P(B_{[t,t+\Delta t)}(X)\,|\,A^{[d,d+\Delta d)}_{[t,t+\Delta t)}(X))$ represents the risk of death for individual X in [t, t+Δt) due to an uncontrollable aviation accident given that X stays within [d, d+Δd) for that time interval. Applying Bayes' theorem to the conditional probability results in:

---

[1] X had no or little control over the aviation activity that caused his or her death, see Section 1.3 for definition of the risk.

$$P(B_{[t,t+\Delta t)}(X) \mid A^{[d,d+\Delta d)}_{[t,t+\Delta t)}(X)) = \frac{P(A^{[d,d+\Delta d)}_{[t,t+\Delta t)}(X) \mid B_{[t,t+\Delta t)}(X))}{P(A^{[d,d+\Delta d)}_{[t,t+\Delta t)}(X))} P(B_{[t,t+\Delta t)}(X)) \qquad (2.1)$$

The left hand side of (2.1) represents the quantity of interest with respect to the spatial variability of the risk in the dimensions time and distance to an airport. In this analysis the variability of the ground fatality risk is determined by estimating the probabilities on the right hand side of formula (2.1). The probabilities on the right hand side can be interpreted as follows:

$P(A^{[d,d+\Delta d)}_{[t,t+\Delta t)}(X) \mid B_{[t,t+\Delta t)}(X))$: The probability that X stayed within [d, d+Δd) for [t, t+Δt) given that X dies due to an uncontrollable aviation accident in [t, t+Δt).

$P(A^{[d,d+\Delta d)}_{[t,t+\Delta t)}(X))$: The probability that X stays within [d, d+Δd) for [t, t+Δt).

$P(B_{[t,t+\Delta t)}(X))$: The risk that X dies due to an uncontrollable aviation accident in [t, t+Δt).

2.2 *Measure of the variability of the risk*

The variability of this risk in the dimensions time and distance to an airport is represented by $P(B_{[t,t+\Delta t)}(X) \mid A^{[d,d+\Delta d)}_{[t,t+\Delta t)}(X))$. Define the conditional mortality rate as:

$$h_{X,d}(t) = \lim_{\Delta d \to 0} \left(\lim_{\Delta t \to 0} \frac{P\left(B_{[t,t+\Delta t)}(X) \mid A^{[d,d+\Delta d)}_{[t,t+\Delta t)}(X)\right)}{\Delta t} \right) \qquad (2.2)$$

and

$$\lambda_X(t) = \lim_{\Delta t \to 0} \frac{P\left(B_{[t,t+\Delta t)}(X)\right)}{\Delta t} \qquad (2.3)$$

Define the following distribution functions:

$$G_{X,t}(d) = \lim_{\Delta t \to 0} P\left(D(X,\tilde{t}) \le d : \tilde{t} \in [t, t+\Delta t) \mid B_{[t,t+\Delta t)}(X)\right) = P\left(D(X,t) \le d \mid B_t(X)\right) \qquad (2.4)$$

$$F_{X,t}(d) = \lim_{\Delta t \to 0} P\left(D(X,\tilde{t}) \le d : \tilde{t} \in [t, t+\Delta t)\right) = P\left(D(X,t) \le d\right) \qquad (2.5)$$

$G_{X,t}(d)$ and $F_{X,t}(d)$ are cumulative distribution functions in d and are defined for all t. Further, the related density functions are defined as follows:

$$g_{X,t}(d) = \frac{\partial}{\partial d} G_{X,t}(d), \tag{2.6}$$

$$f_{X,t}(d) = \frac{\partial}{\partial d} F_{X,t}(d) \tag{2.7}$$

Although these functions are defined as a limit, the interpretations of the distributions are rather clear:

$G_{X,t}(d)$	:	The probability that X is within distance d of an airport at time t given that X dies at time t.
$F_{X,t}(d)$	:	The probability that X is within distance d of an airport at time t.

The functions $G_{X,t}(d)$, $F_{X,t}(d)$, $g_{X,t}(d)$ and $f_{X,t}(d)$ can, under appropriate assumptions, be estimated from data. As $\Delta t \to 0$, the fraction of the right hand side of (2.1) will converge to:

$$\lim_{\Delta t \to 0} \frac{P(A^{[d,d+\Delta d)}_{[t,t+\Delta t)}(X) \mid B_{[t,t+\Delta t)}(X))}{P(A^{[d,d+\Delta d)}_{[t,t+\Delta t)}(X))} = \frac{G_{X,t}(d+\Delta d) - G_{X,t}(d)}{F_{X,t}(d+\Delta d) - F_{X,t}(d)} \tag{2.8}$$

Expression (2.8) will converge as $\Delta d \to 0$:

$$\begin{aligned} \lim_{\Delta d \to 0} \frac{G_{X,t}(d+\Delta d) - G_{X,t}(d)}{F_{X,t}(d+\Delta d) - F_{X,t}(d)} &= \lim_{\Delta d \to 0} \frac{G_{X,t}(d+\Delta d) - G_{X,t}(d)}{\Delta d} \times \frac{\Delta d}{F_{X,t}(d+\Delta d) - F_{X,t}(d)} = \\ &= \frac{g_{X,t}(d)}{f_{X,t}(d)} \end{aligned} \tag{2.9}$$

Combining (2.1), (2.2), (2.3), (2.8) and (2.9) results in:

$$h_{X,d}(t) = \frac{g_{X,t}(d)}{f_{X,t}(d)} \lambda_X(t) \tag{2.10}$$

This analysis focuses on the spatial variability of the risk associated with the dimension distance to an airport and the risks are measured per year. Therefore the following hypothetical risk is considered:

The risk of death for an individual due to an uncontrollable aviation accident in 2000 given that this individual stays at distance d of the nearest airport for the whole year.
Notation: $P(B_{2000}(X) \mid A^{d}_{2000}(X))$.

The risk is hypothetical because nobody stays at exactly distance d of the nearest airport for a year, but it is the most reasonable measure to quantify the spatial variability of the risk. The risk can be calculated by integrating the hazard rate $h_{X,d}(t)$ over the year 2000:

$$P(B_{2000}(X) \mid A^{d}_{2000}(X)) = \int_{2000}^{2001} h_{X,d}(t)\,dt = \int_{2000}^{2001} \frac{g_{X,t}(d)}{f_{X,t}(d)} \lambda_X(t)dt\,, \tag{2.11}$$

This expression defines the variability of the risk associated with the dimension distance to an airport in 2000 and will be used to represent the variability.

2.3 Approach

The quantity of interest $P(B_{2000}(X) \mid A^{d}_{2000}(X))$ is represented by expression (2.11). Some simplifying assumptions were made to keep the problem feasible. This analysis assumes that the mortality $h_{X,d}(t)$ and each of the quantities on the right hand side of (2.10) are constant within the year 2000 and consequently:

$$P\left(B_{2000}(X) \mid A^{d}_{2000}(X)\right) = \int_{2000}^{2001} h_{X,d}(t)\,dt = \frac{g_{X,2000}(d)}{f_{X,2000}(d)} P\left(B_{2000}(X)\right) \tag{2.12}$$

Expression (2.12) is the basis of this analysis. The quantities on the right hand side are separately estimated in Chapter 3 and 4, and the results are combined in Chapter 5.

The probability $P(B_{2000}(X))$ is estimated using the following formula:

$$P(B_{2000}(X)) = \frac{\begin{array}{c}\text{expected number of ground fatalities due to uncontrollable}\\ \text{aviation accidents in 2000 in the U.S.}\end{array}}{\text{U.S. residence population in 2000}} \tag{2.13}$$

In Chapter 3 the numerator is estimated using historical data and a time series analysis. Estimates of the U.S. residence population of the U.S. census bureau are used to approximate the denominator of (2.13).

Chapter 3 also discusses the method that is applied to approximate $G_{X,2000}(d)$ and $g_{X,2000}(d)$. Historical accident data provided 34 locations of accidents that caused ground fatalities over the 12-year period 1978-1999. The accident locations were related to airports resulting in realizations $d_1, d_2, \ldots, d_{34}$ (distances between related-airports and accident locations) and these distances are used to estimate $G_{X,2000}(d)$. Consequently, this analysis assumes that $G_{X,t}(d)$ has not changed since 1978.

Chapter 4 focuses on the question "Where do Americans live?" U.S. census data from 1990 are used to estimate $F_{X,1990}(d)$ and $f_{X,1990}(d)$. Because the census data of 2000 is not available yet, it is assumed that $F_{X,2000}(d)$ and $f_{X,2000}(d)$ can be approximated by $F_{X,1990}(d)$ and $f_{X,1990}(d)$. Note that estimates of $F_{X,1990}(d)$ are based on where U.S. residents lived in 1990 according to the census data. However, people are mobile and tend to travel, go to work or go to school. The impact of these assumptions is discussed in Chapter 4.

In Chapter 5 results are generated by applying (2.12) and using the results from Chapter 3 (estimates of $P(B_{2000}(X)$ and $G_{X,2000}(d)$) and Chapter 4 (estimate of $F_{X,2000}(d)$).

3 Data Analysis

In order to make proper estimates of the risk of ground fatalities due to aviation accidents in the United States, it is important to accurately analyze the available data and data sources. Section 3.1 describes how the accident data are collected, what sources are used, and discusses the quality of the data. Section 3.2 presents a time series analysis to explore whether the ground fatality rate changed over time. Section 3.3 provides estimates of the expected number of ground fatalities due to uncontrollable aviation accidents in 2000 in the U.S. In Section 3.4 a fatality distribution is derived using the locations of 34 accidents that caused ground fatalities.

3.1 Accident Data

3.1.1 Data acquisition

This section gives an overview of the process that led to the collection of the accident data and relates the data collection process to the objectives of the project. The data collection process included three steps:

1. The first step was to generate a complete list of aviation accidents that caused at least one fatality on the ground. This list was obtained from the National Transportation Safety Board (NTSB) Accident Database and Section 3.1.2 describes how these accidents were extracted.
2. Second, each accident was reviewed and the nature of the ground fatalities involved was determined. In Section 3.1.3, all the ground fatalities who had significant control over the aviation activity causing their death were filtered out, because they do not meet the criterion for third party risk defined in Section 1.3. Newspaper articles and the NTSB accident abstracts were used to determine the nature of the ground fatalities.
3. The third step was to determine the distance between the accident locations and the related airports. Accident locations and their proximities to airports are required to meet the second objective of this analysis (the spatial variability of the risk). Because the risk was expected to be higher in the vicinity of an airport, the distance between the accident location and the related airport was determined for all accidents that caused ground fatalities. Section 3.1.4 shows how the accident locations were identified and how the distances to the related airports were measured. NTSB accident reports, newspaper articles and several Internet sites provided the information needed to pinpoint the accident locations and related airports.

The accident data obtained from the NTSB accident database contained no military accidents.

3.1.2 NTSB Accident abstracts

The first step in the data collection process was to generate a complete list of aviation accidents that caused at least one ground fatality. The National Transportation Safety Board (NTSB) is an independent Federal agency that investigates every civil aviation (non-military) accident in the United States. The NTSB defines an accident as[1]:

An occurrence associated with the operation of an aircraft that takes place between the time any person boards the aircraft with the intention of flight and all such persons have disembarked, and in which any person suffers death or serious injury, or in which the aircraft receives substantial damage.

With this broad mandate, NTSB investigations cover all the civil aviation accidents including those that caused one or more ground fatalities. The NTSB is officially charged with investigating aviation accidents and to manage its investigations and information, the NTSB created the Accident Database. This database contains records of all the civil aviation accidents reported (believed to be nearly 100%). Although unlikely, the NTSB accident database may include mistakes (type mismatches, failure caused by converting files to databases, etc) and be incomplete due to underreporting of accidents. However, the NTSB is highly reliable and errors in the database are expected to be small or non-existent given the agency's high profile mission. This analysis assumes that extracting the data from the NTSB accident database yields a complete set of civil aviation accidents in which at least one groundling was killed.

The NTSB policies concerning preservation of accident reports and the evolution in the NTSB's information technology led to different formats for the accident information available over time:

- Before 1964: No accident information is available.
- 1964-1977: Only very short accident abstracts are available, the NTSB destroyed the original accident reports.
- 1978-1999: Accident abstracts and the original reports are available at the NTSB office in Washington D.C.
- Recent reports: Only preliminary abstracts and reports are available.

The NTSB and the Federal Aviation Administration (FAA) provide online access to the accident data collected since 1983 in a searchable database[2]. A search of the NTSB database produced a list of civil aviation accidents (1964-1999) in which at least one ground fatality occurred (shown in Appendix A).

---

[1] http://www.ntsb.gov/aviation/classify/classification.htm
[2] http://nasdac.faa.gov/asp/fw_ntsb.asp

3.1.3 Nature of the Ground Fatalities

The initial accident list contained an enormous variety of accidents. The following examples of accidents demonstrate the variety:

1. An airplane crashed into a house killing a resident.
2. An airplane hit someone who was walking on the runway.
3. An agricultural airplane hit a flagman during chemical spreading.
4. A mechanic was killed during maintenance.
5. A ground crewman walked into a propeller.
6. An airplane made an emergency landing on a highway, struck an automobile, and killed the driver.
7. An airplane hit an individual who was taking pictures at the end of the runway.

This thesis focuses on the individuals who have little or no control over the exposure of the risk (Section 1.3). Casualties among aviation employees working at the time of the accident are considered occupational and in control. Non-employees can also be killed by an aircraft on the ground due to their decisions. An individual who chooses to walk on a runway risks being struck by an airplane. In contrast, a person who is killed while in a residence, car, or other non-airport location had little or no control over the activity that caused her or his death. This analysis focuses only on those who have little or no control over the exposure of the risks associated with aviation activities and consequently, only ground fatalities who had little or no control were included (i.e., only fatalities 1 and 6 above).

Although the seven above examples might suggest that the determination of the nature of the ground fatality is trivial, some research was necessary to appropriately filter the data. Most of the NTSB accident abstracts contain short narratives that do not give a definitive answer on the degree of controllability of the ground fatalities. Consequently, in several cases a search for newspaper articles was performed to get a better description of the accident. If no relevant articles were found or they were insufficient, the NTSB accident reports (if available) were consulted. Unfortunately, for accidents that occurred more than 20 years ago it is hard to find any newspaper articles and no NTSB reports exist. Consequently, for 14 accidents the question whether the ground fatalities had significant control over the aviation activities that caused their death could not be answered.

After filtering, a list of 88 accidents remained (plus 14 accidents for which the controllability question was not answered). These accidents accounted for 205 confirmed ground fatalities who had little or no control (plus 23 unconfirmed). The number of accidents (in which at least one groundling died who had no or little control) per year is shown in Figure 3.1 for the period 1964-1999. Figure 3.2 shows the number of ground fatalities per year. Although the interpretation of little control is rather vague, in most cases it was rather clear how to classify the fatalities.

As mentioned in Section 1.1, Goldstein et al assumed that 150 groundlings died in the period 1975-1985 due to airplane accidents. After carefully reviewing the data, for that same period this analysis found 75 ground fatalities that had little or no control over the aviation activity that caused their death and 2 additional fatalities for whom the degree of control remained unknown.

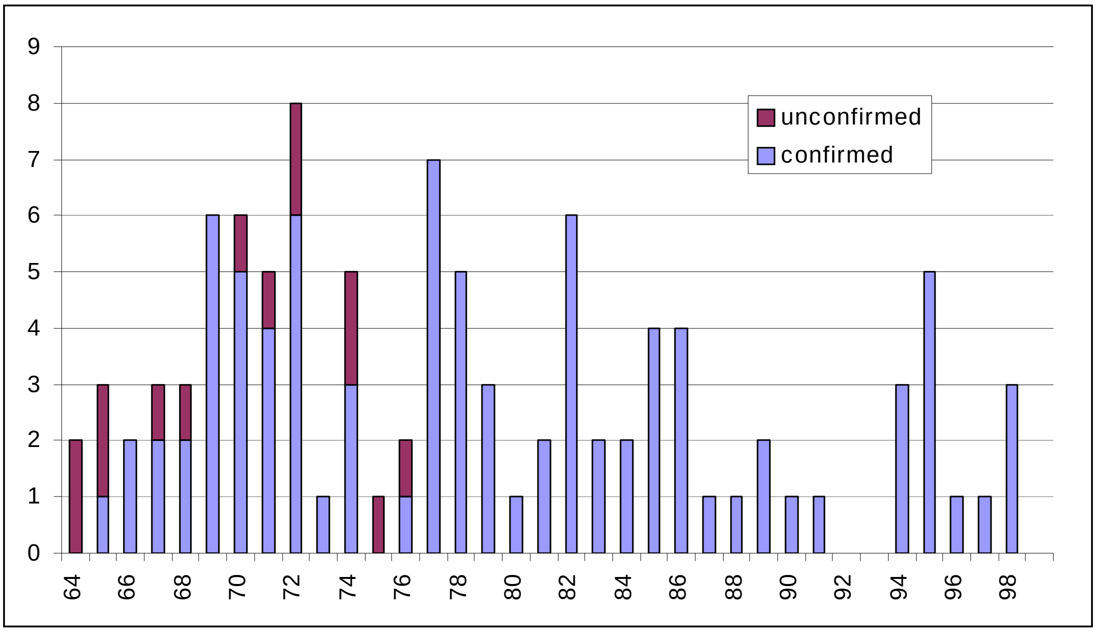

Figure 3.1 The number of aviation accidents that caused at least one ground fatality having little or no control of the aviation activity that caused her or his death (unconfirmed accidents are accidents in which the nature of the ground fatalities, little or no control versus significant control, remained unknown).

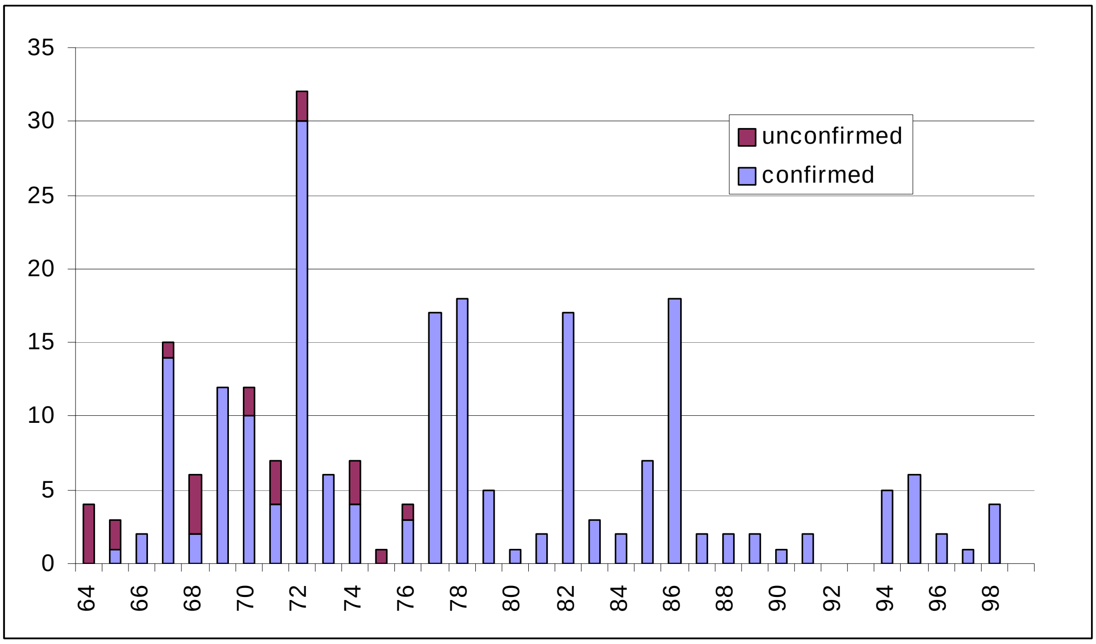

Figure 3.2 The number of ground fatalities having little or no control over the aviation activity that caused death (an unconfirmed ground fatality is a fatality for whom the nature, little or no control versus significant control, remained unknown).

3.1.4 Accident locations

The objective of this thesis includes determining the spatial variability of the risk and consequently, the exact accident locations must be identified. Once the locations are known, the next step is to estimate the distance to the related airport. In most cases the accident abstracts did not give precise information about the accident location. Newspaper articles, and (if necessary and available) the NTSB accident reports, were consulted to pinpoint the locations where the accidents occurred. For example, the following quotation from the *Los Angeles Times* of September 2nd, 1986 helped to determine the accident location of the crash on August 31st that killed 15 groundlings:

"The collision broke off the horizontal stabilizer of the Aeromexico aircraft, a portion of the tail gear crucial to controlling the plane, plunging it into a residential neighborhood near **Carmenita Road and 183rd Street** in Cerritos." [bold added]

The crossroad is assumed to be a reasonably accurate approximation of the actual accident location. The Yahoo! Map service[3] on the Internet was used to find the crossroad on a map (see Figure 3.3).

The accident locations of 47 accidents were found similarly. For all of the accidents that occurred before 1978, locations could not be found. In the period 1978-1982, for one accident the NTSB accident report was not available and therefore this accident location is missing (only one ground fatality).

The groundling risk was expected to be higher in the vicinity of an airport, because most accidents occurred during landings or takeoffs. In order to determine the spatial variability, information about the distance to the related airport was sought. The following definition for an airport-related accident was used:

An accident is related to an airport if the airport is the origin or the final destination of the flight and the accident occurred within 10 miles of the airport.

Accidents that occurred outside 10 miles of both the origin and final destination (or if the flight plan is not known) are classified as airport-unrelated accidents. In some cases, the destination of the aircraft changed during the flight due to an emergency situation and therefore the definition says final destination.

A review of the flight plans of the aircraft involved revealed that 13 of 47 groundling accidents were not related to an airport. The distances from the accident location to the fence of the related airport were measured for the 34 airport-related accidents and the results are presented by a histogram shown in Figure 3.4. This figure strongly supports the expectation that most accidents occurred in the vicinity of an airport and is the basis of the crash distribution function estimated in Section 3.4. The data are represented in tables in Appendix A.3.

---

[3] http://maps.yahoo.com/py/maps.py

Figure 3.3 Accident location of the collision on August 31st, 1986 in Cerritos, California.

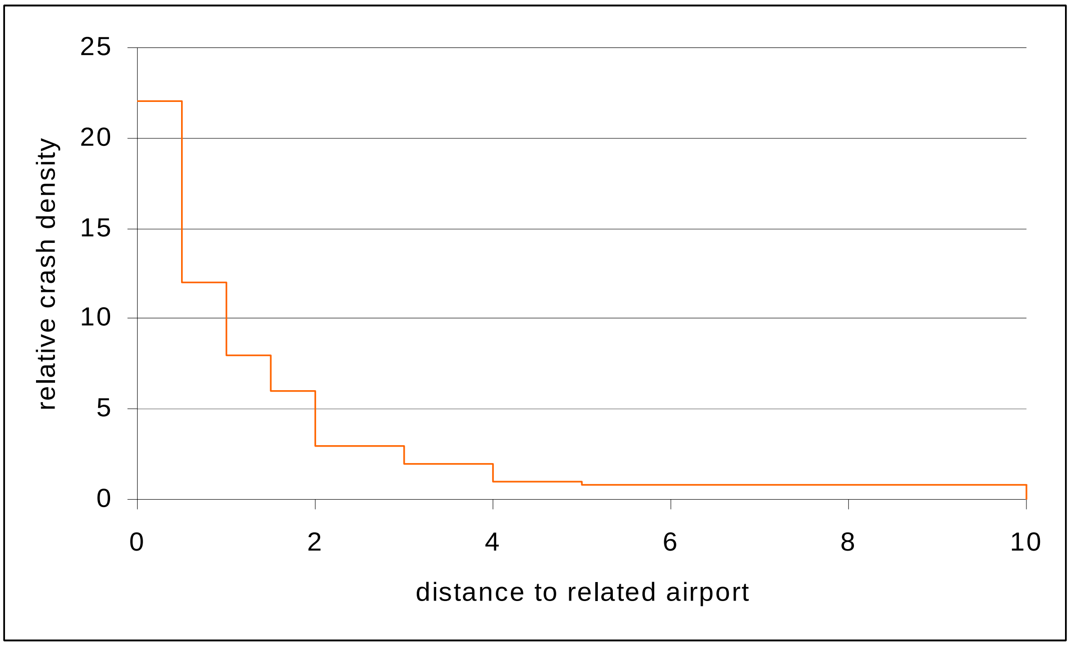

Figure 3.4 Histogram of the distances between groundling accidents and related-airports (miles).

3.2 Time Series of accident rates

3.2.1 Expected number of ground fatalities in 2000?

The first objective of this analysis is to determine the current average risk per year for groundlings $P(B_{2000}(X))$ and this quantity is estimated using formula (2.13) in Section 2.3. This formula requires an estimate of the expected number of ground fatalities in 2000. However, Figure 3.2 shows that the annual number of ground fatalities may have followed a decreasing trend over the last 30 years. Using ground fatality data from 20 years ago will probably result in overestimating the expected number of ground fatalities in 2000. In this section a time series analysis is performed to better understand the applicability of the historical accident data with respect to estimating current ground fatality rates and risks.

3.2.2 Statistical Significance of the Decline of the Annual Number of Groundling Accidents.

Fortunately, accidents that cause ground fatalities are rather rare events, but that makes it more difficult to observe trends over time. A close examination of Figure 3.1 suggests that the number of groundling accidents per year may have decreased over the last 30 years. Consequently, better understanding of the decrease is needed to determine whether the annual number of groundling accidents is really decreasing or this observation can be explained due to a statistical fluctuation. A Kolmogorov-Smirnov test is applied to determine the statistical significance of the decline.

The significance of the decline of the annual number of groundling accidents is determined by testing the following hypothesis:

The groundling accident rate per time unit was constant in 1964-1999.

The Kolmogorov-Smirnov test compares the expected distribution function with the empirical distribution function. Assuming a random variable has a distribution function F and the empirical distribution function is F_n, the Kolmogorov-Smirnov statistic D_n is given by

$$D_n = \sup_t | F(t) - F_n(t) |, \tag{3.1}$$

where n is the number of observations. The hypothesis (the random variable has a distribution function F) will be rejected when D_n is above a critical value.

The applicable time interval in this case is [0, 36), where t=0 corresponds to January 1st, 1964 and t=36 corresponds to December 31st, 1999. The data analysis provided 88 accident times $t_1, t_2, .., t_{88}$ and these times were used to determine the empirical distribution function:

$$F_{88}(t) = \frac{\#\{t_i : t_i \le t\}}{88}$$

The hypothesis is equivalent with the following distribution function:

$$F(t) = \tfrac{1}{36}\,t, \quad 0 \le t \le 36.$$

The Kolmogorov-Smirnov test focuses on the largest distance between the empirical distribution function and the expected distribution (both are plotted in Figure 3.5). A disadvantage of the KS test is that it ignores the difference between the two distribution functions in the rest of the interval and therefore in some cases the KS test is insufficiently powerful. Figure 3.5 shows that the largest distance between these two distribution functions represents the significance of the decline and consequently the KS test should suffice.

Both F(t) and $F_n(t)$ are plotted in Figure 3.5 and the KS-statistic D_{88} was calculated to be 0.149 (equation (3.1)). This is just above the critical value 0.144 that corresponds to a significance level of 0.025 and consequently the hypothesis was rejected with confidence. Note that in constructing the empirical distribution function, the 14 unidentified accidents were ignored. The impact of their omission led to a lower KS-statistic in the above analysis than would have been obtained otherwise, because these 14 accidents all happened before 1978. Consequently, H_0 would have been rejected with even more confidence (i.e. at a lower significance level) when the unidentified accidents had been included.

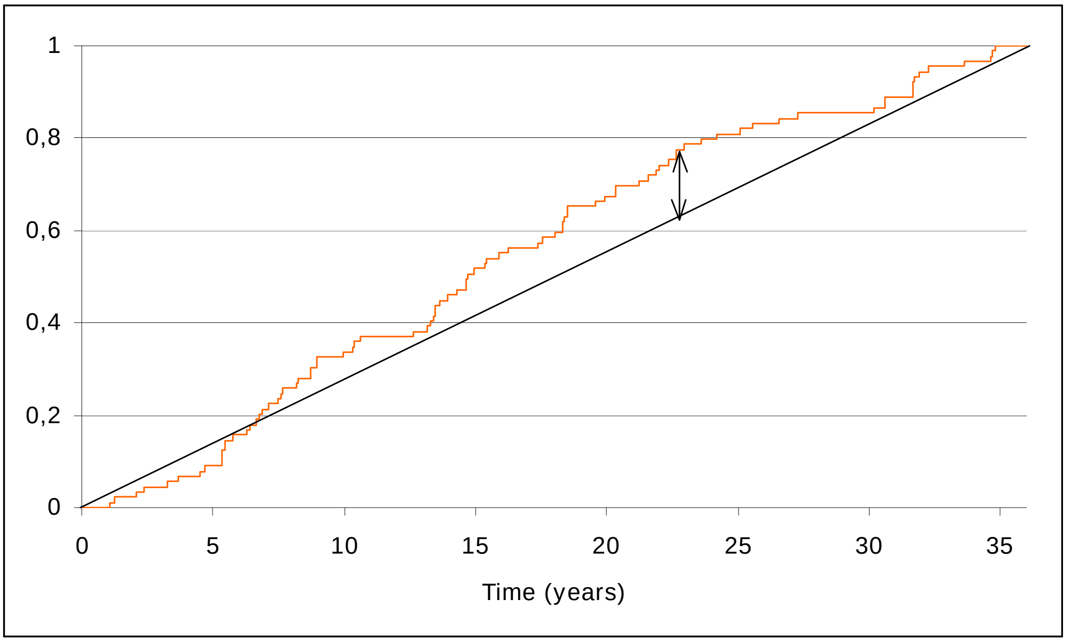

Figure 3.5 The empirical distribution function $F_n(t)$ of ground fatality accidents.

3.2.3 Trend of the groundling accident rate per airport operation.

The KS-test revealed that the decrease in groundling accidents per time unit since 1964 has been significant, but Figure 3.1 and Figure 3.5 show that the decrease has slowed down since the late 1980s. Explaining these trends is difficult, but the groundling accident trends probably are related with other aviation accident trends.

A commonly used accident rate in the aviation industry is the accident rate per operation[4]. The worldwide accident rate per million departures for commercial jets has stabilized since 1980, after a significant decrease in the sixties and seventies[5]. One would expect that the groundling accident rate per operation has followed a similar trend. Figure 3.6 presents the average groundling accident rate per million operations for each year since 1970 and a 5-year moving average. The figure shows that the groundling accident rate per operation has decreased since 1970 and that the trend has flattened in the last 10 to 15 years.

The fact that the groundling accident rate per operation seems to follow a horizontal trend over the last 15 years makes accident data from this period useful for estimating the current rate. In Section 3.3 the groundling accident rate per operation is estimated and this parameter plays an important role in the procedure that is applied to estimate the risk of ground fatalities.

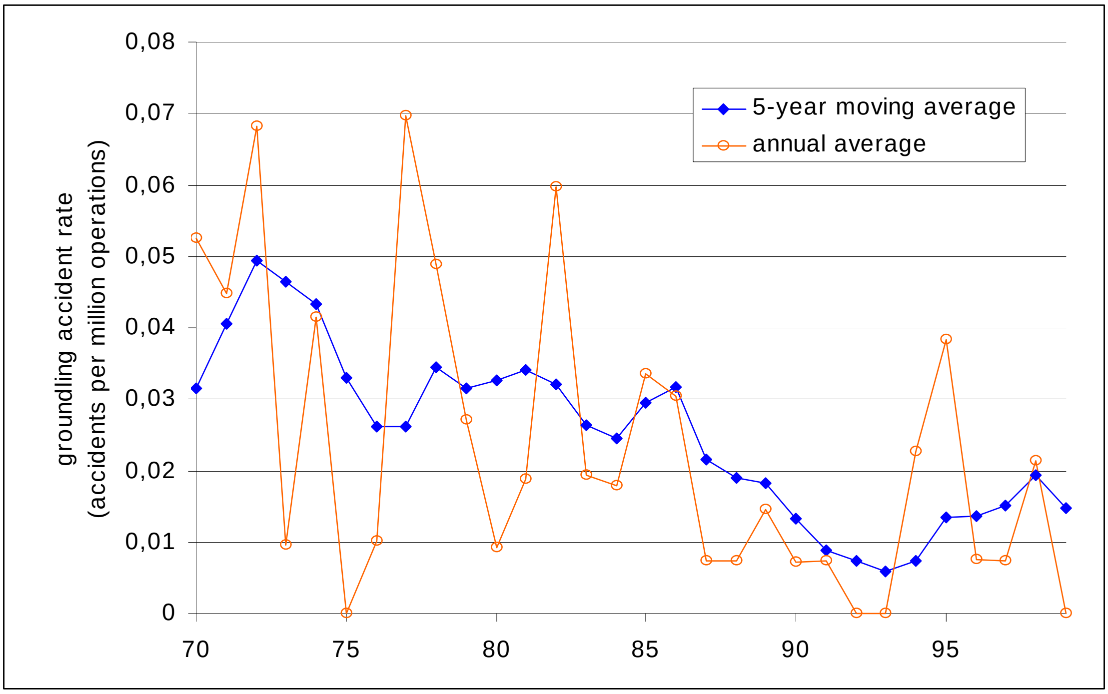

Figure 3.6 Groundling accident rate per million operations, annual average and 5-year moving average for 1978-1999.

---

[4] Landing or takeoff.

[5] *Statistical Summary of Commercial Jet Airplane Accidents, Worldwide Operations 1959-1999*, Boeing, 1999.

3.3 Estimates of the current ground fatality rates

3.3.1 Approach

One of the objectives of this analysis is to estimate the current risk of ground fatalities due to uncontrollable aviation accidents. This requires an estimate of the expected number of ground fatalities in 2000. The easiest approach to estimate the expected number of ground fatalities due to uncontrollable aviation accidents is dividing the number of ground fatalities in 1964-1999 (205 fatalities) by the duration of the period (26 years). This simple model results in an estimate of 7.9 ground fatalities per year. Figure 3.2 and Section 3.2 showed that the annual number of ground fatalities followed a decreasing trend since 1964 and thus using ground fatality data from 1964-1999 to estimate the expected number of ground fatalities in 2000 will overestimate this number.

Section 3.2.2 argues that the annual number of groundling accidents has remained reasonably constant since the late eighties. In the period 1987-1999, 27 ground fatalities having little or no control occurred due to aviation accidents, suggesting an average of 2.1 ground fatalities per year. From a groundling perspective, no large accidents occurred in the United States in 1987-1999 (the largest accident caused 2 ground fatalities), but this does not guarantee that these will not happen in the future (they did occur before 1987). Ignoring the rare, large accidents results in underestimating the expected number of ground fatalities in 2000, because their contribution to the risk can be significant even though their probability of occurrence is low. Another important issue is that the number of groundling accidents per year probably depends on the number of landings and takeoffs that occurred during that year. An increase in aircraft movements obviously increases the expected number of groundling accidents and ground fatalities.

In order to take into account the large accidents (i.e. accidents with a lot of ground fatalities) and the airport activity, the following approach was used to estimate the current ground fatality rate:

- First, the current groundling accident rate per airport operation[6] was estimated based on accident and airport activity data since 1987.
- Second, the expected number of ground fatalities per groundling accident was estimated using data since 1964. This estimate is significantly influenced by large accidents, because several large groundling accidents occurred in 1964-1986 (Figure 3.1, Figure 3.2).
- Third, the current ground fatality rate per operation was estimated by multiplying the groundling accident rate per operation and the expected number of ground fatalities per accident.
- Finally, multiplying the ground fatality rate per operation and the estimated number of operations in 2000 resulted in an estimate of the expected number of ground fatalities in 2000.

This approach is separately applied to three aviation categories, air carriers, air taxis and general aviation, because of their different accident rates and the different impact of their crashes. This methodology resulted in an estimate of 3.5 ground fatalities in 2000 (air carriers and air taxis each account for 20% of the expected number of fatalities, general aviation is responsible for 60%). The above approach assumes that the groundling accident rate per operation has remained constant since 1987 and the expected number of ground fatalities per groundling accident has not

---

[6] An airport operation is either a landing or a takeoff.

changed since 1964 for all three aviation categories. In Section 3.3.2 and 3.3.3 these assumptions are argued, referring to the time series in the previous section, and the groundling accident rate per operation and the expected number of ground fatalities are estimated. In Section 3.3.4 estimates for the ground fatality rates per million operations are presented and the expected number of ground fatalities in 2000 is estimated based on FAA-predictions of the airport activity in 2000.

3.3.2 Groundling accident rates per million airport operations

Accident or mortality rates can be measured by different units of exposure depending on the activity and the availability of data. For the groundling accident rate 3 potential units of exposure were available; per year, per million airport operations, and per million hours flown.

In this analysis, accident rates are measured per airport operation for the following reasons:

- About 70% of all aviation accidents concerning commercial jets[7] and 60% of the accidents involving general aviation aircraft[8] occur during the landing, final approach, takeoff or initial climb. Further, results presented in Section 3.4.1 show that 87% of air carrier accidents and 63% of general aviation crashes that caused at least one ground fatality were airport-related. The airport-related operations are the most risky phases of a flight and it seems fair to measure the risk per airport operation.
- Applying an accident rate per operation makes it easier to compare risks at busy and non-busy airports.
- The groundling accident rate has not fluctuated extremely since the late eighties and therefore the current rate per operation can be reasonably estimated by using data from the last decade.

Because the groundling accident rate decreased until the late eighties, only accident and airport activity data for the period 1987-1999 were used to get an estimate of the present groundling accident rate. Groundling accident rates were separately estimated for air carriers, air taxies and general aviation, because of the different impact of their crashes and their different accident rates. A crashing Boeing 747 is more likely to kill a groundling than a small Piper.

Estimates of the current accident rate per million operations are shown in Table 3.1. Two sources provided the information to estimate the number of airport operations at U.S. airports:

- The FAA Terminal Area Forecasting System (TAF).
  The TAF system provides the airport activity data at the 525 FAA-airports[9] (highly accurate numbers) and estimates of the airport activity at another 2,900 airports. The historical data for the FAA-airports has been reported by FAA air traffic and contracted controller staff report. The FAA provides online access to the airport activity data in the TAF system[10] and the data go back to 1976.

---

[7] *Statistical Summary of Commercial Jet Airplane Accidents, Worldwide Operations 1959-1999,* Boeing, 1999.

[8] *Annual review of accident data – U.S. General Aviation – 1996,* ARG-99/01, National Transportation Safety Board, 1999.

[9] airports with FAA or Federal contract air traffic control services, i.e. towered airports.

[10] http://www.apo.data.faa.gov/faatafall.HTM

- FAA Form 5010.
  Form 5010 reports aviation activity at airports as estimated by FAA inspectors and State planning agencies using information provided by airport managers and others[11]. This Form is used to get estimates of the number of operations at the airports that are not in the TAF system. An electronic version of the 5010 record was downloaded[12] and provided estimates of the airport activity in 1998 at an additional 10,500 registered airports and seaplane bases. It is expected that the airport activity at these additional airports followed the same trend as at the 3,500 airports in the TAF system and this assumption is made to generate estimates of airport activity in 1987-1997 at the non-TAF airports.

The FAA-airports account for more than 99% of the airport operations for air carriers, 70% for air taxis and 35% for the general aviation and all the TAF-airports account for at least 99% of the air carrier operations, 95% of the air taxi operations and 80% of the operations for general aviation. Consequently, the estimates of the number of airport operations (at registered airports) for air carriers are nearly exact, for air taxis the estimates are good and for general aviation some uncertainty remains. Furthermore, the TAF system provides predictions for the airport activity until 2015. The estimates and predictions of the annual number of airport operations at registered airports are presented in Table 3.2.

	Air carrier	Air taxi	General aviation
Number of groundling accidents in 1987-1999[13]	2	4	14
Number of airport operations in millions 1987-1999	174.5	174.3	1,413.4
Groundling accident[14] rate per million operations	0.011	0.023	0.0099

Table 3.1 Estimates of the current groundling accident rates per million operations.

---

[11] *Terminal Area Forecasts – Fiscal years 1998-2015*, U.S. Department of Transportation, Federal Aviation Administration, FAA-APO-98-10, October 1998.

[12] http://www.gcr1.com/aims/5010_fed.htm

[13] collisions between aircraft of different categories are double counted, total number of accidents in 1987-1999 is 19.

[14] Groundling accident is an accident that caused at least one ground fatality who had little or no control over the aviation accident that caused his or her death.

Year	Air Carrier	Air Taxi	General Aviation
1987	13.14	11.33	110.68
1988	12.88	12.23	110.90
1989	12.66	12.27	111.70
1990	13.02	12.76	113.45
1991	12.68	12.87	110.23
1992	12.59	13.35	108.60
1993	12.74	13.60	106.26
1994	13.30	14.08	104.70
1995	13.79	14.08	102.83
1996	13.99	14.47	104.59
1997	14.23	14.41	105.75
1998	14.35	14.56	111.40
1999	15.13	14.34	112.29
Total 87-99	174.49	174.34	1,413.37
2000	15.48	14.58	113.12
2005	17.45	15.84	117.48
2010	19.66	17.07	121.99
2015	22.00	18.31	126.69

Table 3.2 Number of airport operations in millions since 1987 and predictions for 2000, 2005, 2010 and 2015 for air carriers, air taxis and general aviation.

3.3.3 Number of ground fatalities per accident

From a groundling perspective, no large aviation accidents occurred between 1987 and 1999, but this does not mean that no large accidents will occur in the future. Using only the accident data in 1987-1999 to estimate the ground fatality rate will result in underestimating the rate because of the absence of large groundling accidents in this period. Although these large accidents are rare, their contribution to the risk can be significant. Therefore, accident data since 1964 are used to estimate the expected number of ground fatalities per accident, *given* that at least one ground fatality occurred. These estimates also reflect the rare large accidents, because several large accidents occurred in this period.

The statistics for the number of ground fatalities per groundling accident are shown in Table 3.3 and Table 3.4 presents an overview of the fatality data. As expected, an air carrier crash on average accounts for more ground fatalities than a general aviation crash. The air taxi category includes commuter air carriers and therefore their average number of ground fatalities per groundling accident is higher than the average for general aviation.

The current expected number of ground fatalities per groundling accident is estimated using 1964-1999 data, consequently this analysis assumes that this number remained constant since 1964. Basically, this assumption means that the expected number of ground fatalities, given at least one ground fatality, per accident remained unchanged for 27 years. Although the US population increased by 43% since 1964, this assumption appears to be reasonable (an expected increase due to the increase in population is not supported by the data).

Category	Air Carrier	Air Taxi	General Aviation
Number of groundling accidents[15] in 1964-1999	14	11	66
Number of ground fatalities[16] in 1964-1999	60	25	120
Expected number of ground fatalities per groundling accident[17]	4.3	2.3	1.8
Variance	13.6	4.8	7.6
Max. number of ground fatalities per accident	15	7	22

Table 3.3 Ground fatality statistics.

	Number of Accidents[18]		
Number of Ground Fatalities	Air Carrier	Air Taxi	General Aviation
1	5	5	45
2	3	1.5	13.5
3	-	1	2
4	1	-	1
5	-	-	1
6	1	1	-
7	0.5	1	0.5
8	2	-	-
9	-	-	-
≥10	1.5	-	1.5

Table 3.4 The number of ground fatalities per accident.

---

[15] Collisions between aircraft of different categories are double counted, total number of accidents since 1964 is 88.

[16] Ground fatalities that resulted from a collision between aircraft of different categories were evenly divided.

[17] Expected number of ground fatalities, given at least one ground fatality, per accident.

[18] Collisions of aircraft of different categories are counted as half an accident for both categories.

3.3.4 Ground fatality rates per million airport operations

In the previous sections the groundling accident rates and the expected number of ground fatalities per groundling accident have been calculated for air carriers, air taxies and general aviation. The ground fatality rate per million operations is estimated by multiplying the groundling accident rate per million operations and the expected number of ground fatalities per groundling accident. The results are presented in Table 3.5.

These rates can be used to estimate the expected number of ground fatalities in 2000. The TAF system[19] provided predictions for the airport activity in 2000 (Table 3.2) and this resulted in the following estimate of the expected number of ground fatalities in the U.S. in 2000:

$0.047 \times 15.5 + 0.053 \times 14.6 + 0.018 \times 113.1 = 0.73 + 0.77 + 2.0 = 3.5$ ground fatalities

Air carriers and air taxis each account for about 20% of the expected ground fatalities, general aviation is responsible for the other 60%. Estimates for the expected number of ground fatalities in 2005, 2010 and 2015 are presented in Table 3.6.

The U.S. census bureau estimated the U.S. resident population in 2000 as approximately 275 million. The average risk of ground fatalities due to uncontrollable aviation accidents in 2000 can be estimated as (formula (2.13)):

$$P\left(B_{2000}(X)\right) = \frac{3.5\,\text{ground fatalities}}{275\,\text{million U.S residents}} = 1.310^{-8}$$

	Air Carrier	Air Taxi	General Aviation
Groundling accident rate per million airport operations	0.011	0.023	0.0099
Expected number of ground fatalities per groundling accident	4.3	2.3	1.8
Ground fatality rate per million airport operations	0.047	0.053	0.018

Table 3.5 Estimates of the current ground fatality rates per million operations.

	Air Carrier	Air Taxi	General Aviation	Total
2000	0.73	0.77	2.04	3.5
2005	0.82	0.84	2.11	3.8
2010	0.92	0.90	2.20	4.0
2015	1.03	0.97	2.28	4.3

Table 3.6 The expected annual number of ground fatalities per aviation category for 2000-2015.

---

[19] http://www.apo.data.faa.gov/faatafall.HTM

3.4 Accident distribution

3.4.1 Airport-related accidents

It is expected that most accidents occur in the vicinity of airports for two reasons:

1. The landing, final approach, takeoff and initial climb are the most risky operations and accident data show that approximately 70 % of the accidents occur during these phases[20].
2. Airport areas are the most congested aviation areas and consequently they are the most probable scenes for an aviation accident.

However, there is an important difference between commercial flights (air carriers and air taxis) and non-commercial flight (general aviation). General aviation aircraft can land on and take off from lakes, rivers, non-registered airfields and highways (emergency landing when the aircraft runs out of fuel), while commercial flights have a restricted flight plan that at least defines a registered destination[21] and commercial aircraft normally do not run out of fuel. Consequently, general aviation accidents tend to be less often airport-related[22] than air carrier and air taxi accidents.

A review of the accident abstracts revealed that 13 of the 35 accidents in which general aviation aircraft were involved (period 1978-1999) were airport-unrelated. Only 2 of 15 accidents with ground fatalities caused by air carriers or air taxis were airport-unrelated. Table 3.7 shows the percentage of groundling accidents that were airport-related for commercial and non-commercial aircraft.

	Air Carrier and Air Taxi	General Aviation
Number of airport-related accidents	13	22
Percentage	87%	63%
Number of airport-unrelated accidents	2	13
Percentage	13%	37%

Table 3.7 Number and percentage of airport-related and airport-unrelated accidents since 1978.

---

[20] *Statistical Summary of Commercial Jet Airplane Accidents, Worldwide Operations 1959-1999,* Boeing, 1999.

[21] The destination is an airport that is registered in the FAA 5010 record.

[22] An accident is related to an airport if and only if the airport is the origin or the final destination of the flight and the distance between the accident location and the airport is less than 10 miles.

3.4.2 Crash distribution for groundling accidents

To estimate the variability of the risk of ground fatalities, an estimate of the distribution function $G_{X,2000}(d)$ is needed (see equation (2.12)). The airport-related accidents account for the variability of the risk associated with the dimension of distance to an airport. Airport-unrelated accidents do not significantly contribute to the variability in this dimension. The risk of ground fatalities due to airport-unrelated aviation accidents is assumed to be distance independent. Therefore, $G_{X,2000}(d)$ only depends on locations were ground fatalities occurred due to airport-related aviation accidents.

The cumulative distribution function $G_{X,2000}(d)$ that is estimated in this section, can be interpreted as follows (consider individual X who dies due to an airport-related uncontrollable aviation accident in 2000):

$G_{X,2000}(d)$: The probability that X dies within distance d of the related airport given that X dies due to an airport-related, uncontrollable aviation accident in 2000.

Two assumptions are made to generate a reasonable estimate of $G_{X,2000}(d)$:

1. $G_{X,t}(d)$ has not changed since 1978, so accident data from 1978-1999 can be used to estimate $G_{X,2000}(d)$.
2. The expected number of ground fatalities per groundling accident does not depend on the distance between accident location and related airport. Consequently, $G_{X,2000}(d)$ can be seen as a crash distribution for groundling accidents.

In general, crash distribution are not expected to change significantly over time and in this research field, it is accepted to use historical crash data to estimate current crash distributions. The ground fatality data did not provide any reason to assume that the number of ground fatalities per groundling accident depends on the distance between accident location and related airport.

The previous section concluded that approximately 87% of the groundling accidents with commercial aircraft and 63% of general aviation accidents occurred within 10 miles of the origin or the final destination of the flight. In Section 3.1.4 the accident locations of these 34 airport-related accidents were determined and the distance to the related airport was measured. This resulted in distances $d_1, d_2, \ldots, d_{34}$ and these distances are presented in a histogram in Figure 3.4. The data set and the histogram seem to show two characteristics of the crash density function:

- The density function is approximately constant between 4 and 10 miles
- The density function tends to decrease exponentially to this constant value.

Consequently, it seems reasonable to assume that the density function $g_{X,2000}(d)$ is of the following form:

$$g_{X,2000}(d) = a.e^{-bd} + c, \quad 0 \le d \le 10, \qquad (3.2)$$

(where d is the variable, a, b and c are parameters). A crash distribution of the same form is used in the Schiphol survey performed by Technica Consulting [23]. The parameter c is chosen equal to the relative probability in the interval [5,10] (4 of 34 accidents occurred between 5 to 10 miles from the related airport):

$$\hat{c} = \tfrac{4}{34} * \tfrac{1}{5} = 0.023$$

The maximum likelihood estimator (see Appendix B) was used to determine the most probable a and b given the other 30 realizations. It was a 1-dimensional optimization because the area under the density function has to equal 1. The analysis resulted in a density function that is shown in Figure 3.7 (b=0.97, a=0.75), overlaying the histogram.

Although this density function properly represents the likelihood of accident locations, Table 3.8 suggests that the density function depends on the category. Air carriers and air taxis tend to crash closer to airports than general aviation aircraft. Therefore, different crash density functions are estimated for commercial aircraft (air taxis and air carriers) and general aviation. Both density functions are assumed to be of the form $a.e^{-bx} + c$ and the same method is used to estimate the parameters. The results are presented in Table 3.9 and the density functions for commercial aviation and general aviation are plotted in Figure 3.8. When the crash density functions for commercial aviation and general aviation are added up, weighted by the percentage of groundling accidents they account for, the resulting function overlaps the crash density function for civil aviation.

In Chapter 5, the density functions for commercial aviation, air carriers and air taxies (AC and AT), and for general aviation (GA) are used to quantify the variability of the risk. These density functions will be denoted as follows:

$$g^{AC}_{X,2000}(d) = g^{AT}_{X,2000}(d) = 1.33 \times e^{-1.57 \times d} + 0.015, \quad 0 \le d \le 10 \qquad (3.3)$$

$$g^{GA}_{X,2000}(d) = 0.53 \times e^{-0.73 \times d} + 0.027, \quad 0 \le d \le 10 \qquad (3.4)$$

distance to airport	Civil Aviation	Air carrier	Air taxi	General Aviation
<0.5	11	5	2	4
0.5-1.0	6	1	1	4
1.0-1.5	4	-	-	4
1.5-2	3	-	1	2
2-3	3	1	1	2
3-4	2	-	-	2
4-5	1	-	-	1
5-10	4	-	1	3

Table 3.8 Number of groundling accidents within certain distances to the airport.

---

[23] Smith, E., Spouge, J. "*Risk Analysis of Aircraft Impact at Schiphol Airport*", Technica Consulting, May 1990.

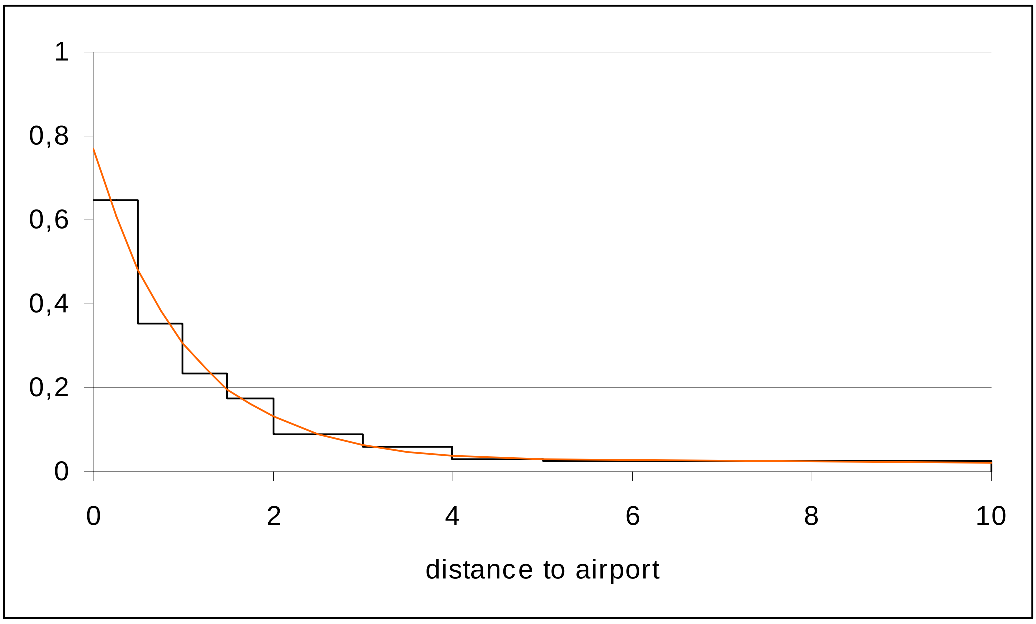

Figure 3.7 Crash density function for groundling accidents (all aviation categories).

	a	b	C
Civil Aviation	0.75	0.97	0.023
Commercial Aviation	1.33	1.57	0.015
General Aviation	0.53	0.73	0.027

Table 3.9 Estimates for a, b and c when the density function is of the form $a.e^{-bd} + c$.

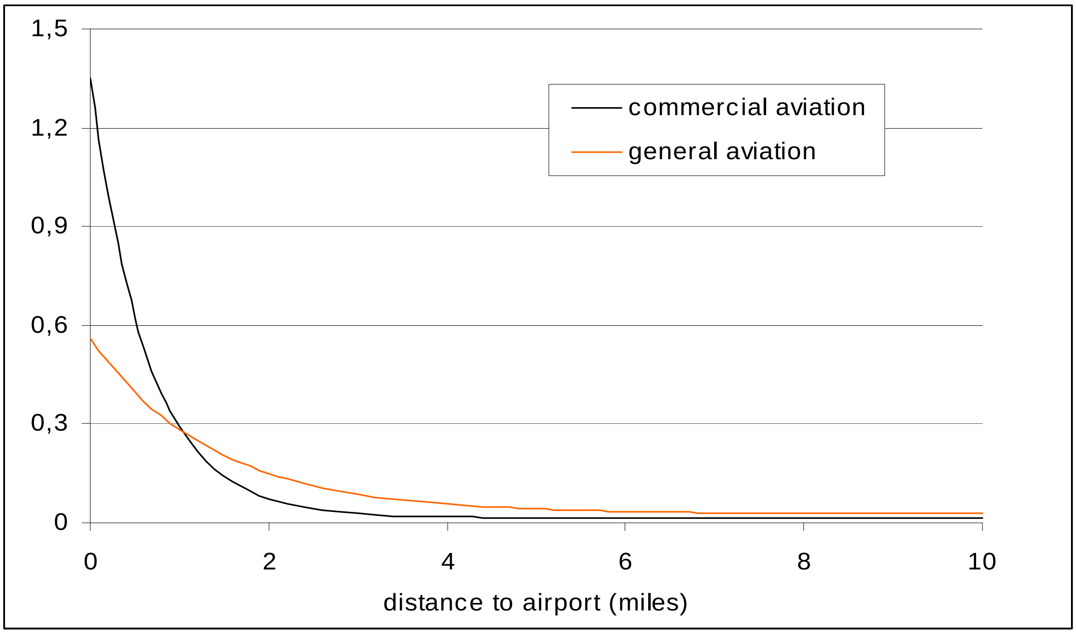

Figure 3.8 Crash density functions for commercial aviation (air carriers and air taxis) and general aviation.

4 Geographic Information System Modeling

In this chapter a Geographic Information System model is developed that relates the American resident population to a selected list of airports. This GIS model makes it possible to derive a distribution of the population as a function of the distance to the nearest airport (in the selected group), in other words how many people are living within x miles of one of the selected airports. Results are presented for several different airport lists.

4.1 How close do Americans live to airports?

The distribution function $F_{X,2000}(d)$ needs to be approximated in order to estimate the quantity that represents the variability of the risk in 2000, $P\left(B_{2000}(X) \mid A^{d}_{2000}(X)\right)$ in expression (2.11). The distribution function $F_{X,t}(d)$ can be interpreted as follows:

$F_{X,t}(d)$ =P(random resident X is within distance d of an airport at time t)

People are mobile and tend to travel, go to work or to school. Sometimes they even go on holiday or decide to stay abroad for a year. It is impossible to trace a resident at a random time during the year or even a random time during the day. The only available data about how Americans are spatially distributed are the U.S. census data, which contain resident locations for all U.S. residents. In this analysis $F_{X,t}(d)$ is estimated based on where U.S. residents live according to the database of the Census Bureau. Although U.S. residents do not spend 24 hours a day at their resident location, most Americans probably spend most of their time at home. This study noted the issue of mobility but uses a stationary distribution for $F_{X,t}(d)$ by assuming that the total mobility will not influence $F_{X,t}(d)$ over time. In other words, it is expected that any resident who leaves his residence area is replaced by someone else. This might not be totally true, because some areas like business areas and educational institutions are busiest during working hours and residential areas are more crowded during the evening and night. On the other hand, it is not automatically true that business areas and educational institutions tend to be situated closer or further from airports than residential areas. Because the distribution is parameterized by distance to an airport, the stationary assumption seems to be not unreasonable.
At the time of this study, the 2000 census data was not available yet and therefore 1990 census data had to be used.

Another issue concerns the collection of airports. Over 10,000 airports are registered in the United States and the diversity in activity is enormous. At some international airports, over 300,000 operations are performed each year, while at some local airports only once a week a plane takes off. To compare the risk of ground fatalities at busy and non-busy airports, spatial variability of the risk due to a selected list of airports is quantified for different airport lists. Variability due to airports that are not in the list at stake is ignored. Therefore the following distribution functions are determined for different airport lists L:

$F_{X,1990,L}(d)$ = P(random U.S. resident X is living within d miles of an airport in airport list L according to the 1990 census data) (4.1)

A Geographic Information System model is built to estimate these distribution functions. The model is based on the following procedure (notation of $F_{X,1990,L}(d)$ is shortened to $F_L(d)$ for convenience reasons) :

1. Select a relevant list of airports L.

2. Generate a sample of 1000 random U.S. residents and the locations of their residences.

3. Determine the distance to the nearest airport that is in L for each resident in the sample.

4. Use these distances to construct an empirical distribution function that approximates $F_L(d)$.

Section 4.2 describes how the residents were sampled from the 1990 U.S. residents. Section 4.3 discusses the model and the software program that were used to estimate the distance between a residence and an airport. Finally, the cumulative distribution functions $F_L(d)$ are presented for several relevant airport lists in Section 4.4. This section also discusses the uncertainty analysis and the assumptions.

4.2 Sample of the population

4.2.1 Census data

The U.S. Census Bureau is the main provider of information about the U.S. population and its official mission is formulated as follows:
to be the preeminent collector and provider of timely, relevant, and quality data about the people and economy of the United States.

Every 10 years the Census Bureau takes a census. The most recent census data available at the time of this project was the 1990 census, although compilation of the 2000 census data was underway. The results of the 2000 census are expected to be delivered to the President of the United States on December 31st, 2000 and the districting data are made available for the public on March 31st, 2001.

It is not possible to get a direct sample of random residents because the U.S. Census Bureau does not provide addresses of individuals to third parties due to strict privacy regulations. Fortunately, the Census Bureau provides online access[1] to census data for states, counties, census tracts, census block groups and census blocks. Population data from states and counties do not have the degree of precision of where people live that is needed for the detailed modeling. However, census tracts, census block groups and especially census blocks contain valuable information about U.S. residents on a scale that is applicable to this project.

Census blocks are the smallest geographic units used in tabulating the 1990 census. The census block data are stored in a database called Master Area Block Level Equivalency (MABLE). The MABLE database is a collection of 51 state-level data sets containing a total of just under 7 million census blocks.

---

[1] http://www.census.gov/plue/

For each block the MABLE database provided the following information:

- Population within the block,
- Area of the block, and
- Coordinates (longitude, latitude) of the interior point of the block.

The interior point corresponds to the spatial centroid of the census block except in those few cases that the centriod is outside the block, then it is moved to a location just inside. Census block groups are the second smallest geographic units and contain several adjacent census blocks. Likewise, census tracts consist of several adjacent census block groups. The census block groups in the area around John F. Kennedy Airport in New York and their interior points are shown in Figure 4.1 and the census tracts in this area are presented in Figure 4.2. The maps are provided by the Tiger map service of the U.S. Census Bureau[2]. No online map services that provide digital maps with census blocks have been found. Figure 4.1 shows that the census *block groups* can be reasonably used to relate the U.S. residence population to a list of airports. Using census *blocks* to derive a distribution of the population as a function of the distance to the nearest airport in an airport list, will generate even better results, because the census blocks are significantly smaller than the census block groups (see Table 4.1).

The differences between the three geographic units are shown by the statistics in Table 4.1 for the state of Delaware. For a population of approximately 666,000 there are 175 census tracts, 535 census block groups and 15,136 census blocks. Figure 4.3 presents the cumulative distribution function of the population per census block for the state of Delaware. Note that no one lives in approximately 23% of the census blocks (population is 0) and 95% of the census blocks contain a population that is smaller than 150 residents. Other statistics of the tracts, block groups and blocks for Delaware are presented Table 4.1. The number of residents in and the area of a census block are negatively correlated; census blocks that cover a small area mostly contain a large number of residents.

	Census tract	Census block group	Census block
Number of units	175	535	15,136
Average population per unit	3806	1245	44
Maximum population per unit	18897	8333	4376
Average area of a unit (sq. m.)	11.2	3.7	0.13
Maximum area of a unit (sq. m.)	107	40	7.7

Table 4.1 Statistics of the three geographic units for the state Delaware based on 1990 census data.

---

[2] http://tiger.census.gov/

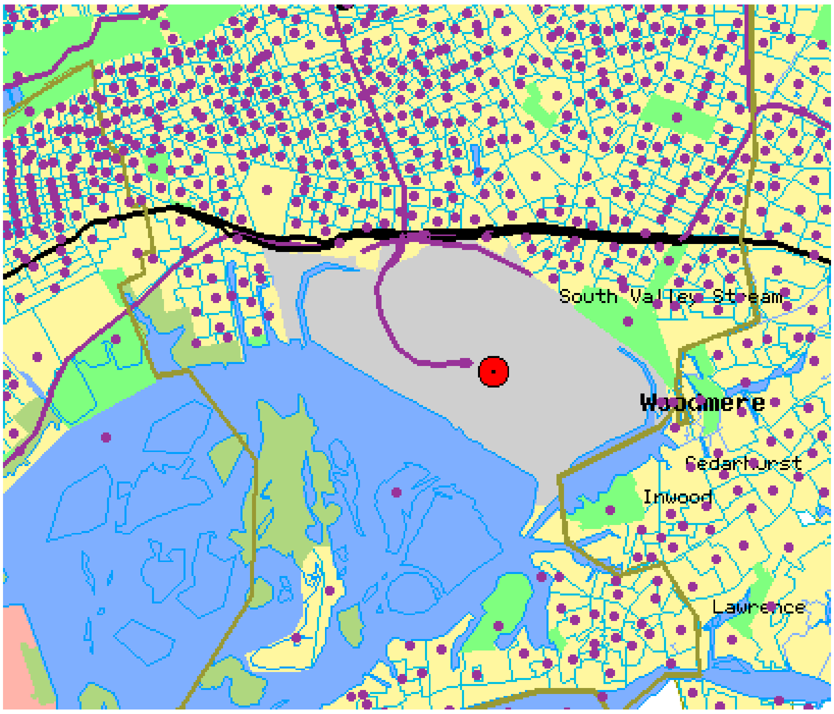

Figure 4.1 Map of the area around John F. Kennedy Airport, showing the census block groups and their interior points.

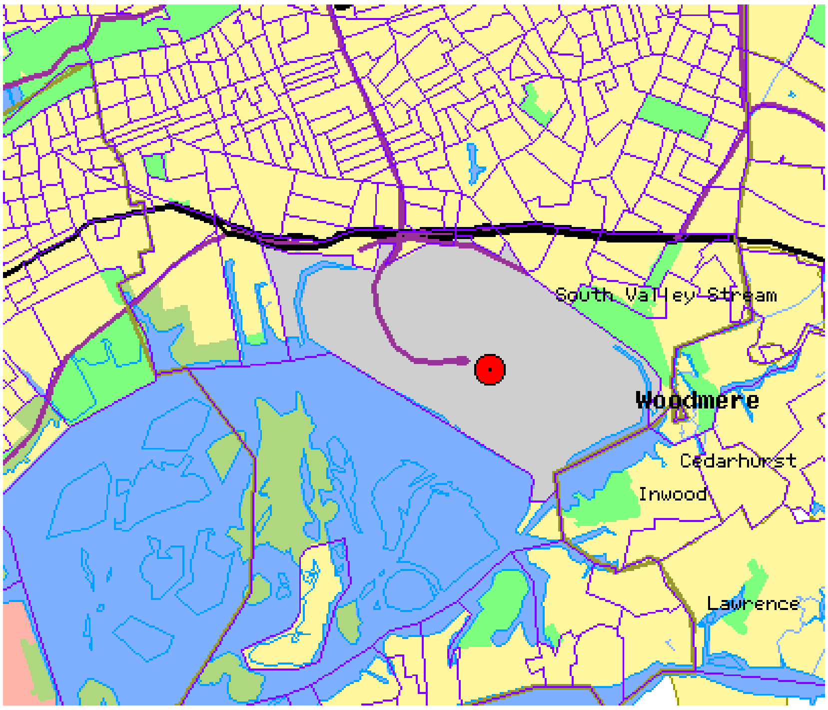

Figure 4.2 Map of the area around John F. Kennedy Airport, showing the census tracts.

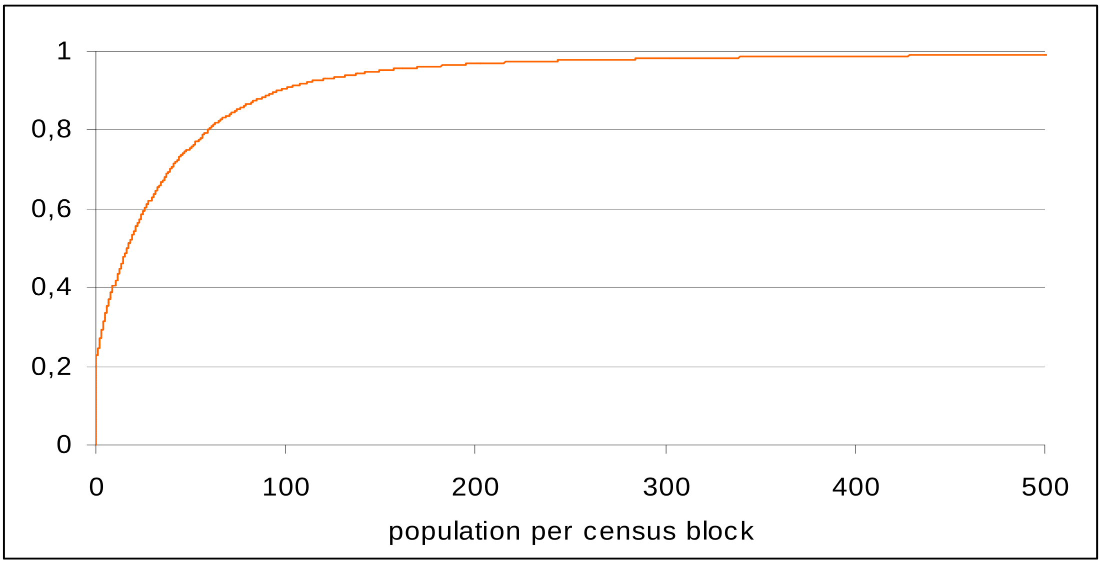

Figure 4.3 Cumulative distribution function of the population per census block in Delaware.

4.2.2 Sampling method

Due to strict privacy regulations it was not possible to directly sample 1000 residents (and addresses) from the U.S. resident population. However, the previous section showed that the collection of census blocks provides information about the U.S. population on a spatial scale that has the required accuracy for the objective of this project. The collection of approximately 7 million 1990 census blocks can be downloaded per state from the homepage[3] of the U.S. Census Bureau.

The following procedure was applied to sample 1000 U.S. residents from the 1990 U.S. residence population:

- First, the 1000 samples were partitioned among the 51 states[4] proportional to the state populations (see Table 4.3).
- Second, for each state a Monte-Carlo simulation was performed to pick the required number of census blocks (Table 4.3). The probability of being picked for a census block was proportional to the population in the census block.
- Finally, the location of a residence of a random resident was simulated by assuming a 2-dimensional uniform distribution on the area of the census block. This is further discussed in Section 4.4.4.

Stratified sampling (step 1) was applied to decrease the sampling error by geographically covering the whole United States on the state level. The first two steps resulted in a set of 999 census blocks (because of rounding, 999 residents are sampled instead of 1000 (Table 4.3)) and in the next step one location was randomly picked within each block (uniform distribution). In fact, these 999 locations in the U.S. were sampled from a spatial density function that is proportional to the population density. The population

---

[3] http://www.census.gov/plue/geocorr/data/mable-ascii/

[4] including District of Columbia.

density and consequently the spatial density function were estimated per census block and assumed constant within each block.

The first two steps of the sampling procedure resulted in a geographically stratified, population-weighted sample of 999 census blocks (possibly repeated).
Figure 4.4, Figure 4.5 and Table 4.2 provide information about the area of and the population in sampled census blocks. Recall that the population per census block and the area of the census block are negatively correlated. Census blocks with a large population are in general rather small. These statistics will be used to better interpret the results in Section 4.4. Note that the average population per census block of the sample is much higher than for Delaware, because the blocks were weighted by their population during the sampling.

	average	maximum
population per census block	281	13,004
area of a census block (sq. m.)	0.38	19.6

Table 4.2 Statistics of the sample of census blocks.

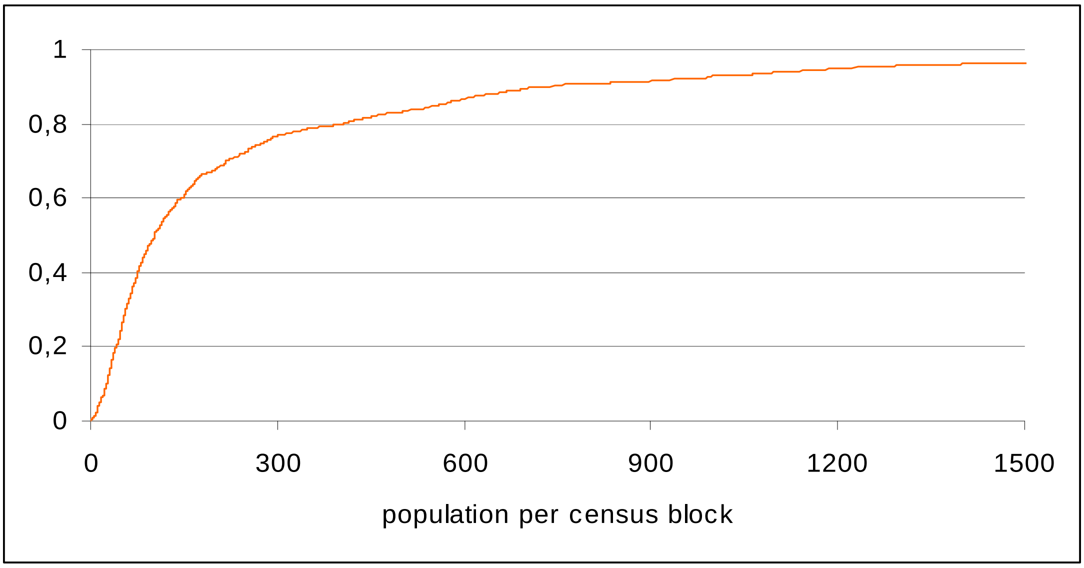

Figure 4.4 Cumulative distribution function of the population per census block in the sample.

	1990 census Population	Percentage of U.S. population	Number of samples
U.S.	248,790,925	100	999
Alabama	4,040,389	1.62	16
Alaska	550,043	0.22	2
Arizona	3,665,339	1.47	15
Arkansas	2,350,624	0.94	9
California	29,811,427	11.98	120
Colorado	3,294,473	1.32	13
Connecticut	3,287,116	1.32	13
Delaware	666,168	0.27	3
District of Colombia	606,900	0.24	2
Florida	12,938,071	5.20	52
Georgia	6,478,149	2.60	26
Hawaii	1,108,229	0.45	4
Idaho	1,006,734	0.40	4
Illinois	11,430,602	4.59	46
Indiana	5,544,156	2.23	22
Iowa	2,776,831	1.12	11
Kansas	2,477,588	1.00	10
Kentucky	3,686,892	1.48	15
Louisiana	4,221,826	1.70	17
Maine	1,227,928	0.49	5
Maryland	4,780,753	1.92	19
Massachusetts	6,016,425	2.42	24
Michigan	9,295,287	3.74	37
Minnesota	4,375,665	1.76	18
Mississippi	2,575,475	1.04	10
Missouri	5,116,901	2.06	21
Montana	799,065	0.32	3
Nebraska	1,578,417	0.63	6
Nevada	1,201,675	0.48	5
New Hampshire	1,109,252	0.45	4
New Jersey	7,747,750	3.11	31
New Mexico	1,515,069	0.61	6
New York	17,990,778	7.23	72
North Carolina	6,632,448	2.67	27
North Dakota	638,800	0.26	3
Ohio	10,847,115	4.36	44
Oklahoma	3,145,576	1.26	13
Oregon	2,842,337	1.14	11
Pennsylvania	11,882,842	4.78	48
Rhode Island	1,003,464	0.40	4
South Carolina	3,486,310	1.40	14
South Dakota	696,004	0.28	3
Tennessee	4,877,203	1.96	20
Texas	16,986,335	6.83	68
Utah	1,722,850	0.69	7
Vermont	562,758	0.23	2
Virginia	6,189,197	2.49	25
Washington	4,866,669	1.96	20
West Virginia	1,793,477	0.72	7
Wisconsin	4,891,954	1.97	20
Wyoming	453,589	0.18	2

Table 4.3 The number of resident samples per state.

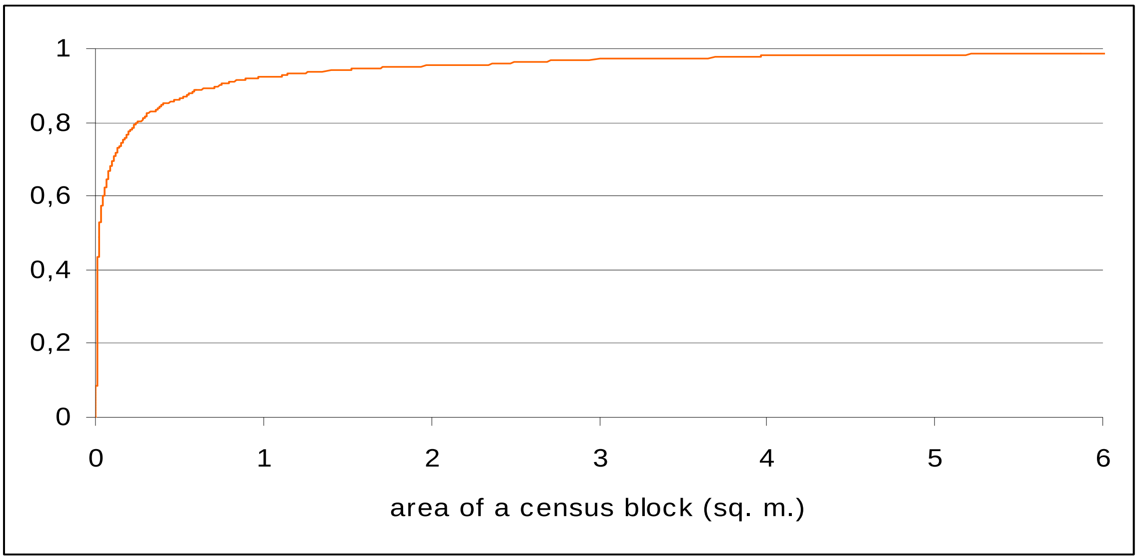

Figure 4.5 Cumulative distribution function of the area per census block in the sample.

4.3 Distance to Airports

4.3.1 Airports in the United States

According to the 5010 Record of the FAA, 13,749 airports, 4982 heliports, 462 seaplane bases and 88 stolports were registered in the United States in 1998. Table 4.4 shows that the number of airports has significantly changed over time, but it turned out that most closings and openings concern small private airports or airstrips. There is a huge diversity among the airports in activity and facilities. Approximately 35% of the airports have paved runways, about 650 airports have a FAR[5] part 139 certification that allows air carriers to use these airports and the 30 largest airports account for 70% of the enplanements[6].

This study analyzed the spatial variability of the risk associated with different lists of airports. It was expected that risks of ground fatalities are higher around busy airports than in the vicinity of less busy airports. The first list contains the 100 busiest airports in the United States. The FAA Terminal Area Forecasting System and the FAA Master record, discussed in Section 3.3.2, provided estimates of the number of airport operations in 2000 for each registered U.S. airport. The activity was separately estimated per aviation category. This raised the question on how to measure the activity at an airport; does an air carrier operation weigh as much as a general aviation operation?

---

[5] Federal Aviation Regulations

[6] http://www.faa.gov/arp/A&D-stat.htm

Because this analysis evaluates the risk of ground fatalities, the number of operations of each aviation category is weighted by its ground fatality rate (Table 3.5). This results in the potential risk of an airport, example:

Consider Airport Y and its activity:

- *202,340 air carrier operations*
- *15,233 air taxi operations*
- *36,122 general aviation operations*

The potential risk of airport Y (i.e. the expected number of ground fatalities per year around airport Y according to the ground fatality rates in Table 3.5):

$202{,}340 \times 0.047\ 10^{-6} + 15{,}233 \times 0.053\ 10^{-6} + 36{,}122 \times 0.018\ 10^{-6} = 0.011$ *ground fatalities.*

The U.S. airports are ranked by their potential risk and the 100 highest ranked airports form the Top100 list. According to the ground fatality rates in Table 3.5 the total activity of all Top100 airports account for 36% of the airport-related ground fatalities in 2000. In the same way the airports for the Top250 list and Top2250 list are selected. The Top250 airports account for 51% of the ground fatalities and the Top2250 airports for 95% (results are summarized in Table 4.5).

The ranking process assumes constant ground fatality rates among all airports, which does not generally hold true because local circumstances influence these rates, consequently it is designated as "potential risk". Issues concerning the applicability of this assumption in the context of this project are discussed in Chapter 5. In the next Sections, the U.S. resident population is related to each of the airport groups presented in Table 4.5. Alphabetic lists of Top100 and Top250 list are shown in Appendix C.

Year	All facilities[7]	Airports
1987	17,015	12,907
1988	17,327	12,950
1989	17,446	12,946
1990	17,490	12,920
1991	17,581	12,904
1992	17,846	13,016
1993	18,317	13,228
1994	18,343	13,202
1995	18,224	13,145
1996	18,292	13,175

Table 4.4 The number of registered civil airports, heliports, stolports and seaplane bases in the United States[8].

---

[7] All facilities includes airports, heliports, seaplane bases and stolports

[8] Source: FAA statistical Handbook for Aviation (calendar year 1996).

Airport list[9]	Number of airports	Percentage of the airport-related ground fatalities in 2000
Top 100	100	36%
Top 250	250	51%
Top 2250	2250	95%

Table 4.5 Airport lists and the percentage of airport-related ground fatalities they account for based on their airport activity.

4.3.2 Software

This section discusses a visual basic application that was written to relate a sample of census blocks to a list of airports. For each block the program searches for the nearest airport and calculates the distance. These distances can be used to estimate a cumulative distribution function of the population as a function of the distance to the nearest airport in the list. An ExcelSheet is used to present the output and to import the inputs.

The program requires the following inputs:

- List of airports.
  For each airport three attributes are needed (provided by the FAA 5010 Master Record):
  1. Degrees longitude of the location of the airport.
  2. Degrees latitude of the location of the airport.
  3. Area of the airport in square miles.

- Sample or list of census blocks
  For each census block two attributes are required:
  1. Degrees longitude of the interior point.
  2. Degrees latitude of the interior point.

The software program does the following:

For each census block in the sample the program searches for the two nearest airports in the list and returns the airport identifiers[10] and the calculated distances.

The most important aspect of the program is how it calculates the distance between a census block and an airport. In Chapter 3 the distance between airport and accident location is defined as the distance from the accident scene to the fence of the airport. The distance reference applied for the population has to be consistent with the reference used for the accident locations, because the results are combined to estimate the groundling risk in Chapter 5. However, the inputs are not sufficient to exactly determine the distance from the interior point to the fence of an airport, because the shape of the airport area can only be observed from maps and influences the distance.

---

[9] Each list only contains airports and seaplane bases, no heliports or stolports.

[10] The FAA uses three character expressions to identify airports in their databases.

The program uses a simple model to estimate the distance between an airport and the interior point of a census block (the distribution of the population within a census block is discussed in Section 4.4.4). Each airport in the 5010 Master Record is identified by the longitude and latitude of the airport location that more or less corresponds to the centre of the airport area (red dot in Figure 4.6). The distance from the fence of the airport to the interior point of the census block is estimated by assuming a circular airport. The centre of the circle corresponds to the airport location and the area of the circle is equal to the airport area. Figure 4.6 shows the circular airport that was used to estimate the distances to John F. Kennedy Airport.

Assuming a circular airport makes the actual calculation of the distance rather easy. The distance between the boundary of the circular airport and the interior point of the census block is equal to the distance between the airport centre and the interior point minus the radius of the circle. Because the area of the circular airport is assumed to be equal to the real airport area $A_{airport}$, the radius r can be calculated as follows:

$$A_{airport} = r^2\pi \Rightarrow r = \sqrt{\frac{A_{airport}}{\pi}} . \qquad (4.2)$$

The distance between the airport centre and the interior point of the census block can be calculated by using the formula for the so-called great circle distance[11]. The distance d of the shortest path between two points located at latitude δ and longitude λ of (δ_1, λ_1) and (δ_2, λ_2) on a sphere of radius a is equal to:

$$d = a_{earth} \arccos[\cos(\delta_1) \cos(\delta_2) \cos(\lambda_1-\lambda_2) + \sin(\delta_1) \sin(\delta_2)], \qquad (4.3)$$

where δ_1, δ_2, λ_1 and λ_2 are in radians and the radius of the earth a_{earth} is 3963 miles. Formula (4.3) does not take into account elevation and the flattening of the earth but this will not cause problems for this application because the distances are rather small.

The software program uses equations (4.2) and (4.3) to estimate the distance between the airport and the interior point of the census block. However, Figure 4.6 shows that the circular airport reasonably covers the real airport area, but it is definitely not perfect. Therefore, the accuracy of this estimation method is discussed in the next section.

The program writes the output, for each census block the two nearest airports and the estimated distances, to an ExcelSheet where it can be used for further analysis. A print out of the program is presented in Appendix D.

---

[11] http://mathworld.wolfram.com/GreatCircle.html

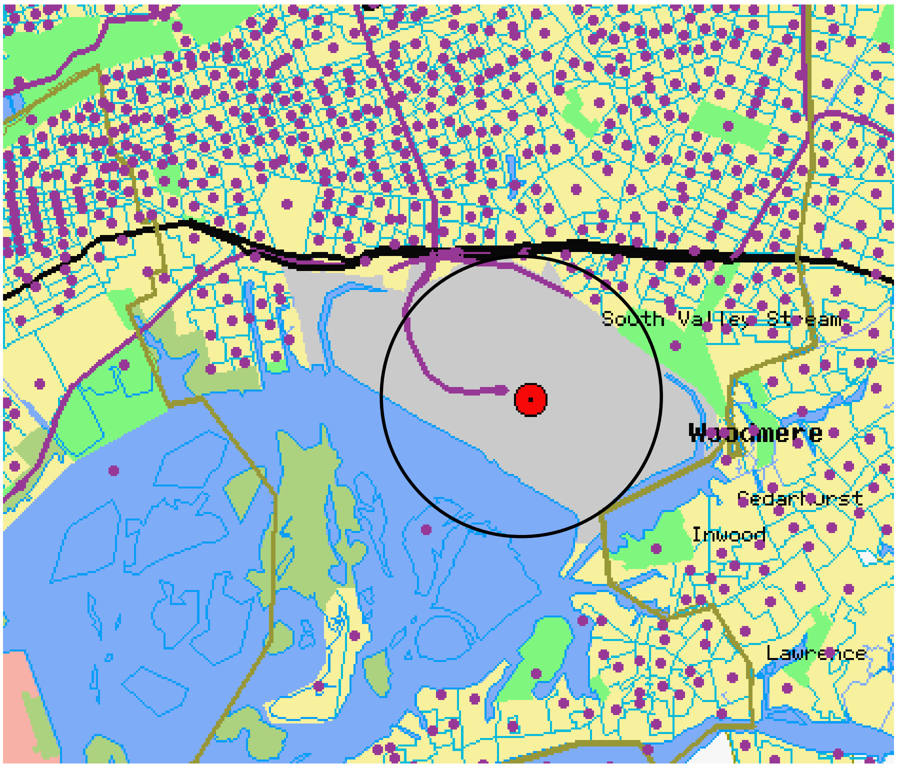

Figure 4.6 Circular approximation of John F. Kennedy Airport.

4.3.3 Error

In the previous section a model is discussed that estimates the distance between the fence of an airport and the interior point of a census block. Because the software program only estimates the real distance, the error has to be quantified to better interpret the output results. The estimation error ε is defined as follows:

$$\varepsilon = d_{real} - d_{estimate}, \tag{4.4}$$

where d_{real} is the real distance and $d_{estimate}$ is the estimated distance according to the model applied in the software program.

The real distance is unknown but can be determined by using online map services that show the airport area (Yahoo! map service or Tiger map services). Therefore, 120 estimated distances were picked out of the output lists and the 120 corresponding real distances were determined by using the Yahoo! Map service[12]. A scatter plot of the real and estimated distances is shown in Figure 4.7. This plot shows that the distribution of the error can be split up in the following three intervals:

---

[12] http://maps.yahoo.com/py/maps.py

- Estimated distance is negative.
  In this interval the real distances seem to be independent of the estimated distances.
- Estimated distance is greater than 1 mile.
  For these distances the expected value of the real distance tends to be equal to the estimated distance and the error seems to be independent of the estimated distance.
- Estimated distance is greater than 0 but smaller than 1 mile.
  This interval can be seen as the intermediate between the two situations.

A close examination of the data suggested that the error (d_{real}-$d_{estimate}$) is normally distributed if $d_{estimate}>1$ and that the real value D_{real} can be approximated by a normal distribution (independent of $d_{estimate}$) if $d_{estimate}<0$. The following model was applied to describe the distribution of the real value D_{real} given the estimated distance $d_{estimate}$:

$$D_{real} = \begin{cases} \beta & \text{if } d_{estimate} < 0.32 \quad \text{with } \beta \sim \text{Normal}(0.37, 0.17) \\ d_{estimate} + \varepsilon_1 & \text{if } 0.32 < d_{estimate} < 1 \quad \text{with } \varepsilon_1 \sim \text{Normal}(0.05,\ 0.17 + 0.29 \times (d_{estimate} - 0.32)) \\ d_{estimate} + \varepsilon_2 & \text{if } d_{estimate} > 1 \quad \text{with } \varepsilon_2 \sim \text{Normal}(0.05, 0.37) \end{cases} \qquad (4.5)$$

Note that $E(D_{real}|d_{estimate})$ is a continuous function of $d_{estimate}$ and that $\sigma_D(d_{estimate})$ is continuous as well. The 5% and 95% interval lines for $D_{real}|d_{estimate}$ together with the $E(D_{real}|d_{estimate})$ line are plotted in Figure 4.7. Because all distances have to be greater than zero, $D_{real}|d_{estimate}$ will be a truncated distribution.

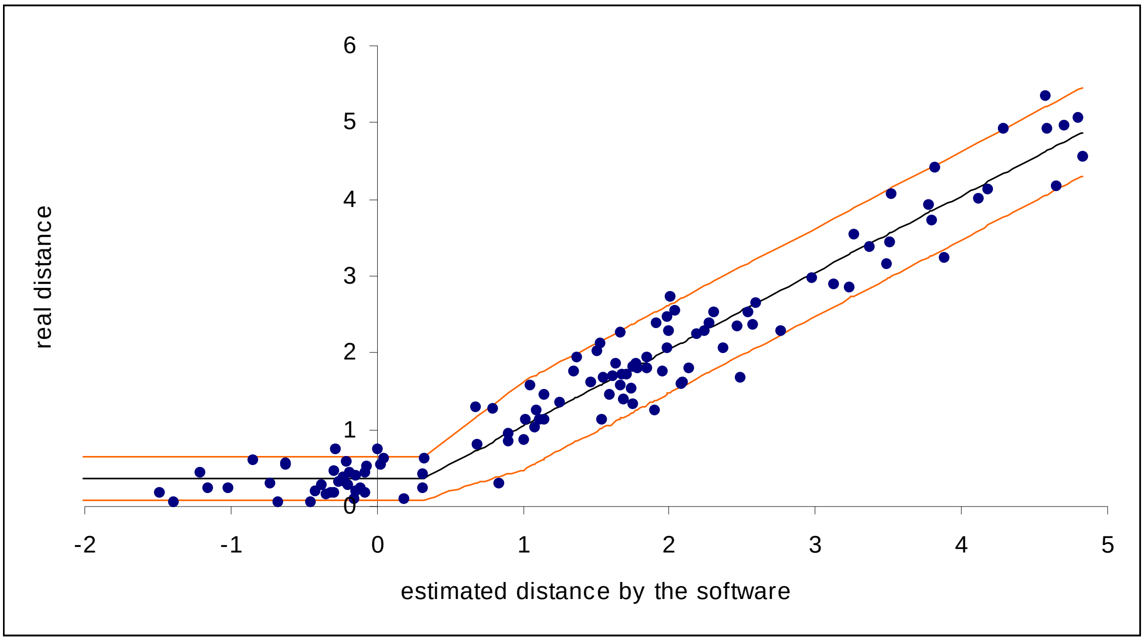

Figure 4.7 Scatter plot of the estimated and real distances.

In order to verify whether the stochastic variables ε_2 and β can be reasonably approximated by normal distributions, two probability plots were generated and two Kolmogorov-Smirnov tests were performed. Both the probability plots (Figure 4.8) and the K-S tests showed that it is reasonable to assume that ε_2 and β are normally distributed (μ_ε=0.05, σ_ε =0.35 and μ_β=0.37 σ_β=0.17). According to the Kolmogorov-

Smirnov test based on 75 realizations, the hypothesis that the error ε_2 is normally distributed would not have been rejected at a 0.1 significance level. Similar results were obtained for β, the hypothesis would not have been rejected at a 0.1 significance level (K-S test was based on 37 data points). Figure 4.7 shows that it is reasonable to interpolate the normal distribution within the interval [0.32,1] assuming a linear increasing standard deviation.

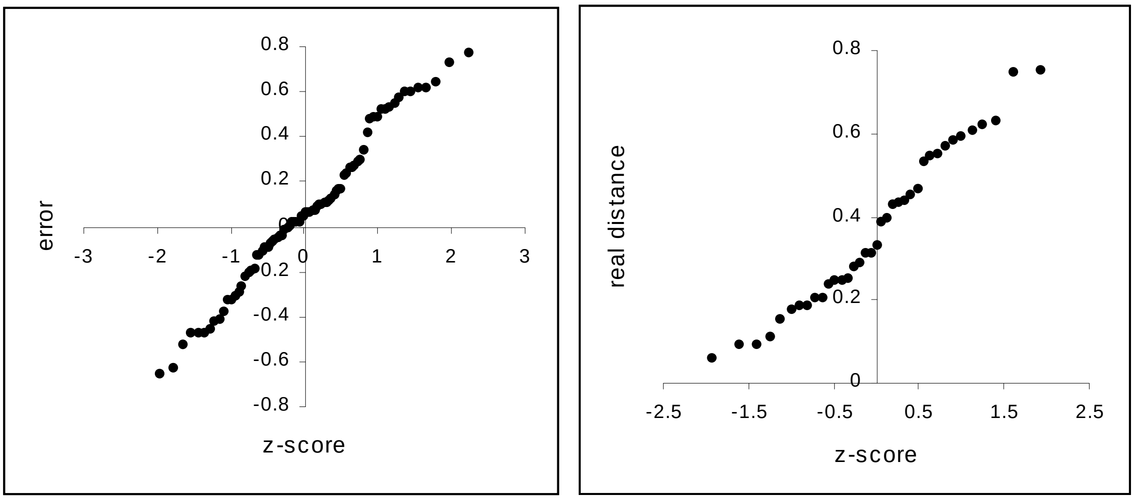

Figure 4.8 Probability plots for the error ε_2 (75 data points) and β (37 data points).

4.4 Results

4.4.1 Output of the software program

For several airport lists L, the final estimates of the distribution functions $F_L(d)$ (see Formula (4.1), page 33, the notation of 1990 is dropped for reasons of convenience) are presented in this section (Section 4.4). The output of the software program can directly be used to estimate the empirical distribution function. However, there are several problems and pitfalls that require adjustments. The total procedure that led to the final estimates will be discussed for the airport list *Top100*. For the other lists only the final approximations of the distribution functions are presented in Section 4.4.6.

The first step in the procedure was to run the program for the 100 airports and the sample of 999 census blocks. For each census block the program estimated the distance to the nearest of these 100 airports and returned the values in an ExcelSheet. These 999 distances were used to construct an empirical distribution function that is shown in Figure 4.9. The airport list contains only 100 airports and therefore the number of distances smaller than 1 mile is only 4 (Table 4.6). Consequently, it is not possible to construct a reasonable approximation of $F_{Top100}(d)$ for d<1. However, the groundling risk within 1 mile is most interesting, because ground fatalities are most likely to occur in this area. Therefore, the next section focuses on how to construct a better approximation of $F_{Top100}(d)$ for small d.

	≤ 1 mile	1-2 miles	2-3 miles	3-4 miles	4-5 miles	> 5 miles
number of blocks	4	21	40	36	37	861
relative frequency	0.004	0.021	0.040	0.036	0.037	0.862

Table 4.6 The number of census blocks in the sample and their distances to the nearest airport in the Top 100 list.

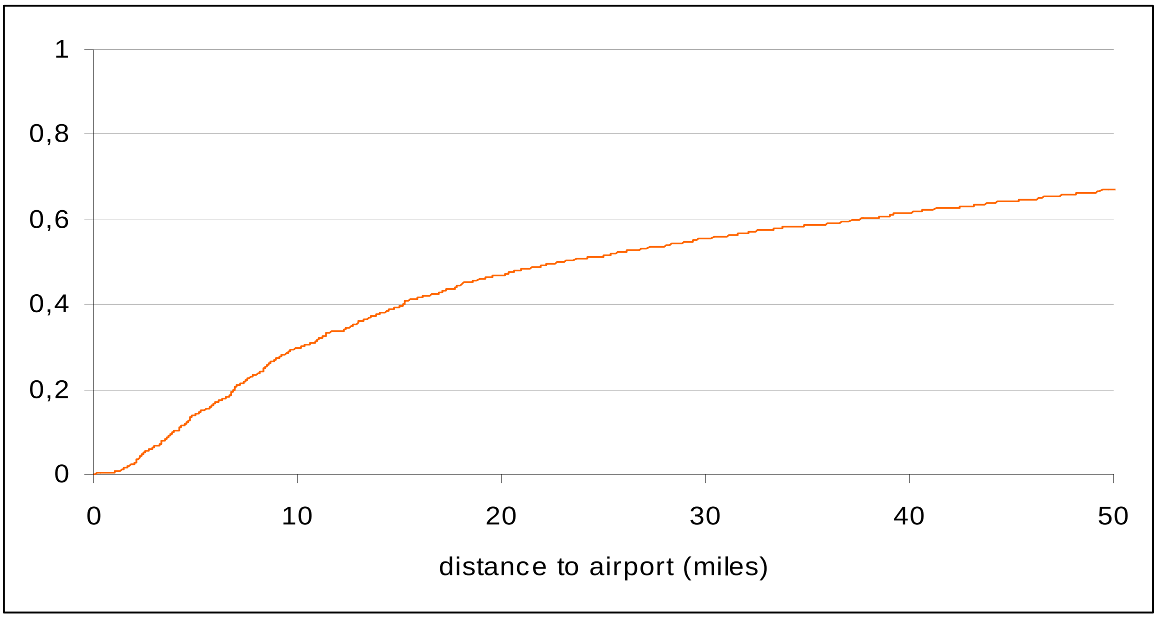

Figure 4.9 Empirical approximation of $F_{Top100}(d)$ directly based on the program's output.

4.4.2 Extra samples in first mile

Although the function in Figure 4.9 looks smooth, it cannot be used to reasonably estimate $F_{Top100}(d)$ for small d, because the number of estimated distances within 1 mile is too low. Increasing the number of samples can solve this, but an online software application of the Census Bureau provided another opportunity to tackle this problem. The geographic correspondence engine[13] (Geocorr) makes it possible to select all the census blocks that are within a certain distance of a given location. This application has been run for each airport in the list and provided all the census blocks that are within 1.5 miles[14] of an airport according to the model in Section 4.3.2.

The 100 runs of the online application resulted in about 23,000 census blocks that covered approximately 0.95% of the 1990 population (2,354,967 of 248,790,925). A sample of 350 census blocks was drawn from these 23,000 blocks (blocks were weighted by their population) and used as input for the software

---

[13] http://www.census.gov/plue/

[14] within 1.5 mile according to the model means within r+1.5 miles of an airport location (for definition of r see Formula (4.2))

program (Geocorr only selects the blocks, but does not calculate the distances). The 350 sampled distances made it possible to construct an empirical distribution function for d<1.5 with a small sampling error (Figure 4.10). However, the output values of the software program are estimates and consequently the estimation error is a source of uncertainty. Therefore, the error uncertainty is discussed in the next section.

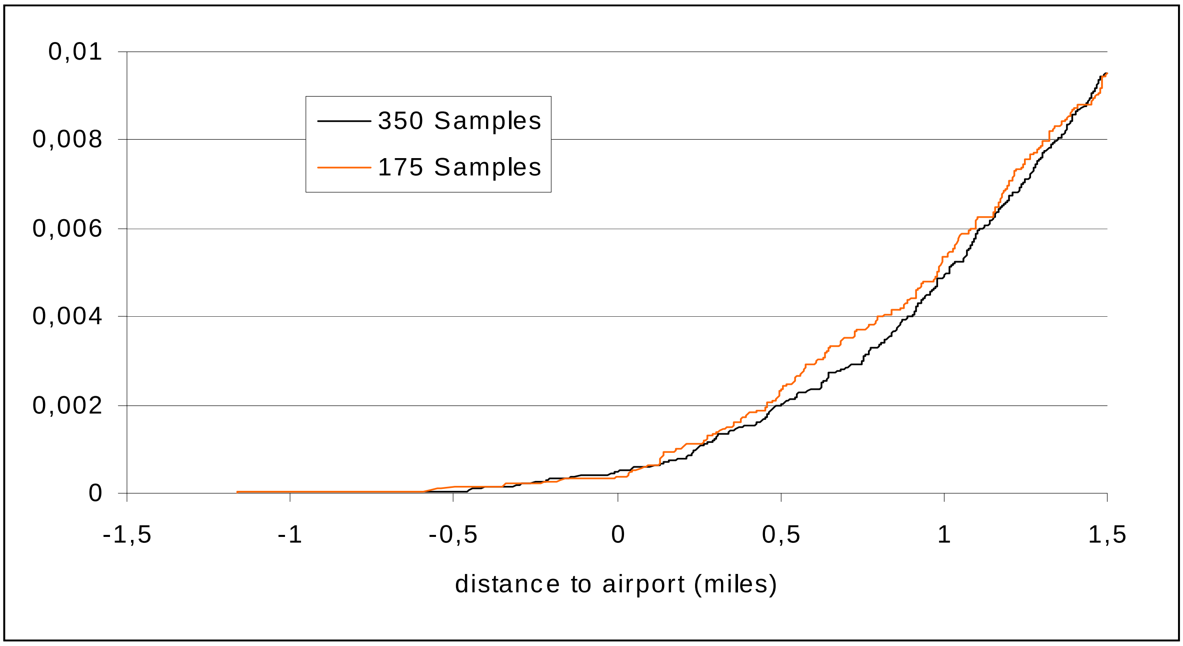

Figure 4.10 Empirical approximation of $F_{Top100}(d)$ based on 175 and 350 samples.

4.4.3 Uncertainty associated with the estimation error

In the previous two sections, empirical distribution functions of $F_{Top100}(d)$ were constructed based on the estimated distances by the software application. However, the output values are just estimates and the negative estimated distances do not make sense. Therefore, the estimation error has to be taken into account when $F_{Top100}(d)$ is constructed. In Section 4.3.3 the distribution of the real distance D_{real} given the estimated distance $d_{estimate}$ has been derived (equation (4.5)).

The two samples provided the following data:

- 350 distances $d^*_{estimate,1},\ldots,d^*_{estimate,350}$ smaller than 1.5 miles. These samples represent 0.95% of the population.
- 987 distances $d_{estimate,1},\ldots, d_{estimate,987}$ greater than 1.5 miles that represent the rest of the population.

For each estimated distance $d^*_{estimate,i}$ and $d_{estimate,j}$ a *real* distance is sampled from the distribution of $D_{real}|d_{estimate}$ defined by (4.5), resulting in $d^*_{real,i}$ (i=1,..,350) and $d_{real,j}$ (j=1,..,987). Based on these distances a distribution function $F^1_{Top100}(d)$ can be constructed:

$$F^{1}_{\text{Top100}}(d) = \frac{\#\{i : d^{*}_{\text{real},i} \le d\}}{350} \times 0.0095 + \frac{\#\{j : d_{\text{real},j} \le d\}}{986} \times 0.9905 \qquad (4.6)$$

By repeating this sampling procedure 99 times $F^{2}_{\text{Top100}}(d),\ldots,F^{100}_{\text{Top100}}(d)$ were generated. These distribution functions were used to construct the expected distribution function $E(F_{\text{Top100}}(d))$:

$$E\left(F_{\text{Top100}}(d)\right) = \tfrac{1}{100}\sum_{k=1}^{100} F^{k}_{\text{Top100}}(d)\,, \qquad (4.7)$$

The expected distribution function $E(F_{\text{Top100}}(d))$ and a 90% confidence interval are presented in Figure 4.11. Note that it concerns d-wise confidence bounds. The uncertainty of $F_{\text{Top100}}(d)$ is relatively small for $d<1.5$ because of the large number of samples in this interval.

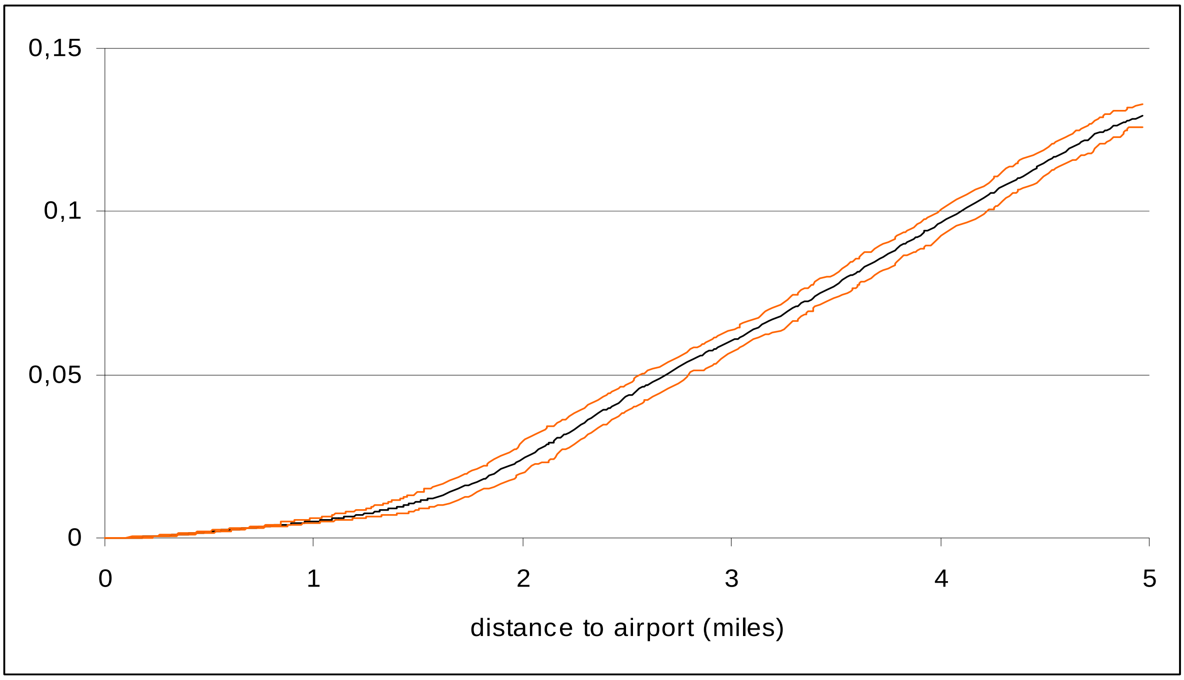

Figure 4.11 The expected value and the 90%-confidence interval for $F_{\text{Top100}}(d)$.

4.4.4 Population within a census block

The analysis in the previous two sections resulted in two sets of samples of the distance between airports and census blocks, one set of 988 distances greater than 1.5 miles and another set of 350 distances smaller than 1.5 mile. These distances represent the distance between the interior point of a census block and the fence of an airport. When it is assumed that all residents are living at the interior point of their census block, these distances can be seen as resident-airport distances. However, it is hard to imagine that residents only live at interior points of census blocks and therefore this section focuses on the question whether this assumption has significantly changed the results.

The MABLE database provided besides the interior points (longitude, latitude), also the areas of the sampled census blocks. The initial sampling strategy was to pick a random location within the census block assuming a 2-dimensional uniform distribution over the block area. The uniform distribution is chosen because it is the least informative and it is known that the blocks are bounded. Assuming a uniform distribution implies a constant population density within the census block. Alas, it is impossible to perform the suggested sampling strategy, because the MABLE database does not provide any information about the shape of the census block.

Consider the estimated distances $d^*_{estimate,1},..., d^*_{estimate,350}$ (smaller than 1.5 miles) and $d_{estimate,1},.., d_{estimate,987}$ between airports and interior points of census blocks and the corresponding areas of the census blocks $a^*_1,.., a^*_{350}$ and $a_1,..., a_{987}$. In order to determine whether the results will significantly change when the population is assumed to be uniformly spread out over the census block instead of concentrated at the interior point, two different situations with square census blocks were evaluated. The geometry for each situation is presented in Figure 4.12.

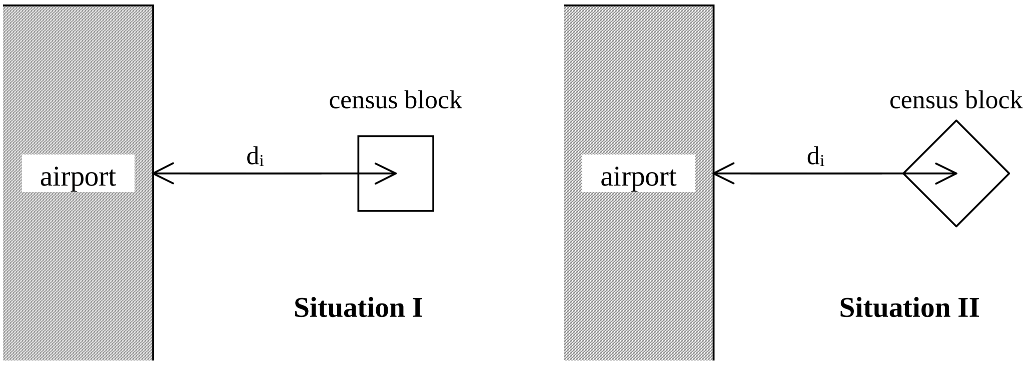

Figure 4.12 Geometry for each of the evaluated situations.

First, $d^*_{real,i}$ and $d_{real,j}$ were generated as in the previous section by sampling according to the distribution defined by (4.5) for each $d^*_{estimate,i}$ and $d_{estimate,j}$. Second, 10 locations within each block were sampled assuming one of the two situations considered and a 2-dimensional uniform distribution within the blocks.

Consider situation I, if the population is uniformly spread out over the census block, then the distance between the resident and the airport $D_{resident,i}$ or $D^*_{resident,i}$ is also uniform distributed:

$$D_{resident,j} \sim \text{Uniform}[d_{real,j} - \tfrac{1}{2}\sqrt{a_j}, d_{real,j} + \tfrac{1}{2}\sqrt{a_j}], \quad \text{and} \tag{4.8}$$

$$D^*_{resident,i} \sim \text{Uniform}[d^*_{real,i} - \tfrac{1}{2}\sqrt{a^*_i}, d^*_{real,i} + \tfrac{1}{2}\sqrt{a^*_i}] \tag{4.9}$$

The next step was to simulate 10 distances $d_{resident,j,1},.., d_{resident,j,10}$ for each $d_{real,j}$ according to formula (4.8) and $d^*_{resident,i,1},..., d^*_{resident,i,10}$ for each $d^*_{real,i}$ according to (4.9), assuming all resident-airport distances are distributed as defined in these formulas. This resulted in two sets of sampled distances:

- A set of 9,870 distances ($d_{resident,1,1},..,d_{resident,987,10}$) that represents the random sample of 987 residents and 99.05% of the population.
- A set of 3,500 distances ($d^*_{resident,1,1},..., d^*_{resident,350,10}$) that represents the random sample of 350 residents and 0.95% of the population.

The important difference between these sets and the original sets is that these sets account for the variability of the population within the census blocks. Both sets are used to construct an empirical distribution function of $F^1_{Top100}(d)$:

$$F^1_{Top100}(d) = \frac{\#\{(i,k) : d^*_{resident,i,k} \leq d\}}{3500} \times 0.0095 + \frac{\#\{(j,l) : d_{resident,j,l} \leq d\}}{9870} \times 0.9905 \qquad (4.10)$$

$F^2_{Top100}(d),.., F^{100}_{Top100}(d)$ are derived identically and again the expected distribution function is defined as follows:

$$E\left(F_{Top100}(d)\right) = \tfrac{1}{100} \sum_{k=1}^{100} F^k_{Top100}(d)\ . \qquad (4.11)$$

In situation II, the resident-airport distance is triangular distributed:

$$D_{resident,j} \sim \text{Triangle}(d_{real,j} - \tfrac{1}{\sqrt{2}}\sqrt{a_j}, d_{real,j}, d_{real,j} + \tfrac{1}{\sqrt{2}}\sqrt{a_j})\text{, and} \qquad (4.12)$$

$$D^*_{resident,i} \sim \text{Triangle}(d^*_{real,i} - \tfrac{1}{\sqrt{2}}\sqrt{a^*_i}, d^*_{real,i}, d^*_{real,i} + \tfrac{1}{\sqrt{2}}\sqrt{a^*_i}) \qquad (4.13)$$

The same sampling strategy provided 100 distribution functions $F^1_{Top100}(d),.., F^{100}_{Top100}(d)$, this time using a triangle distribution to describe the resident-airport distance within the census block. Figure 4.13 presents the expected distribution functions and the 90%-confidence intervals (d-wise) for the original sample and the enlarged samples (for uniform and triangle distribution for airport-resident distance). Even though the scale is small (only the first two miles are shown), the expected distribution functions overlap and the difference between the 5% and 95% boundaries of the confidence intervals are rather small.

Note that both situations in Figure 4.12 are hypothetical, because none of the census blocks is really square. However, Figure 4.13 shows that using a triangle or a uniform distribution to describe the airport-resident distance does not really influence the results. In fact, every unbiased distribution for the airport-resident will generate approximately the same results. It is reasonable to assume that the real, unknown distribution of the airport-resident distance (given d_i) is unbiased, because the interior points correspond to the spatial centroids of the census blocks. Consequently, Figure 4.13 shows that the results will not significantly change when the population is assumed to be uniformly spread out over the census block instead of concentrated in the interior point.

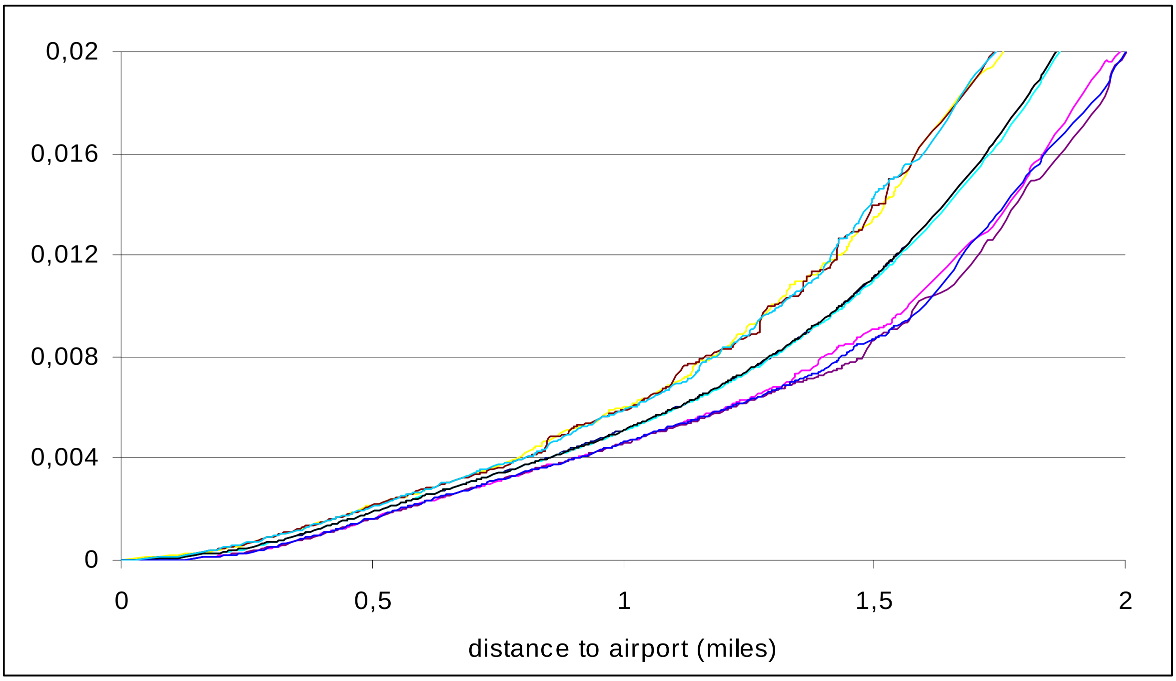

Figure 4.13 The expected values and the 90%-uncertainty intervals of $F_{Top100}(d)$ for the three different approaches (population concentrated in interior point, situation I and situation II).

4.4.5 Uncertainty

The empirical distribution function shown in Figure 4.13 is the final non-parametric approximation of $F_{Top100}(d)$. In the next Section, a parametric approximation of $F_{Top100}(d)$ is derived mostly based on the expected distribution function in Figure 4.13. However, several assumptions have been made to construct this function. In order to correctly interpret the results, the following sources of uncertainty need to be considered:

1. Sampling error
   For $F_{Top100}(d)$ for $d<1.5$, the sampling error is very small because of the large number of samples in this interval. Outside the 1.5-mile area, $F_{Top100}(d)$ is based on a sample of 995 resident locations and consequently, there is a larger sampling error. Still, the sampling error is relatively small compared to other uncertainties.

2. Estimation error
   A software application (Section 4.3.3) was written to estimate the distances between airports and interior points of census blocks. The distances were estimated using a simple model that assumed a circular airport. Section 4.3.3 derived the distribution for the real distance given the estimated distance by the software application. A Monte-Carlo simulation was performed to derive the mean and the d-wise 90%-confidence interval of $F_{Top100}(d)$ for the uncertainty caused by the error (Figure

4.11 or Figure 4.13). The error is relatively small for d<1, because of the large number of samples in [0,1.5].

3. Population within census block
   The population density is assumed to be concentrated in the interior point of each census block. It would be better to sample resident locations assuming that the population is uniformly spread over the census block, but the MABLE database does not provide any information about the shape of the blocks and therefore this is impossible. However, two hypothetical situations assuming square census blocks (Figure 4.12) were evaluated and the results did not significantly differ from the outcomes that were obtained when the residents were expected to live at the interior points. Consequently, it seems reasonable to assume that the population is concentrated in the interior points, because it does not significantly change the results.

4. Errors in census data
   The U.S. Census Bureau provided the 7 million census blocks that are the basis of the GIS analysis in this chapter. Although the Census Bureau has to deal with the non-responses, false responses and other inquiry related problems, the census data can be considered as a reliable source of population data. Errors in census data are not a significant source of uncertainty in this project. However, it concerns census data from 10 years ago.

5. 1990 census data
   The analysis that is performed in this chapter is totally based on U.S. census data from 1990. In fact, the empirical distribution function in Figure 4.13 is an approximation of $F_{Top100}(d)$ on April 1st, 1990. It is very difficult to analyze whether $F_{Top100}(d)$ has changed since 1990, because a list of airports is not a commonly used reference in GIS-models and therefore this issue is not a popular topic in GIS-literature. The U.S. resident population has increased from 248 million in 1990 to 275 million in 2000. In order to determine the development of $F_{Top100}(d)$ over time, the sampling procedure needs to be repeated with the 2000 census data that will become available in April 2001.

Despite several uncertainties, the cumulative distribution function in Figure 4.13 can be seen as reasonable approximation for $F_{Top100}(d)$ in 1990. However, the main source of uncertainty is linked to the question whether $F_{Top100}(d)$ has changed over time. A large effort is needed to quantify this uncertainty and it seems more reasonable to deal with this issue when the 2000 census data become available. Therefore, this project does not focus on this uncertainty, but if necessary the analysis can be repeated with the 2000 census data later. In the next section a parametric distribution function will be derived to approximate $F_{Top100}(d)$.

4.4.6 Parametric distribution functions

Besides the cumulative distribution function $F_{Top100}(d)$, an approximation of the probability density function $f_{Top100}(d)$ is needed in Chapter 5 to relate the groundling risk to the distance to a Top100 airport. The expected distribution functions shown in Figure 4.11 and Figure 4.13 are empirical distribution functions and are constructed using equation (4.6) and (4.10). Consequently, $f_{Top100}(d)$ cannot be determined by simply taking the derivative of the empirical approximation of $F_{Top100}(d)$, because this is a discontinuous step function. It does not make sense to fit a traditional distribution, because nothing indicates that the considered distribution can be described by any of the standard distributions.

Therefore, the empirical distribution function is approximated by several polynomials in different intervals. This parametric approximation of $F_{Top100}(d)$ and the expected empirical distribution are plotted in Figure 4.14 on two different scales. Both functions can barely be distinguished and the parametric distribution function seems to be a good fit. The parametric distribution is used in the further analysis.

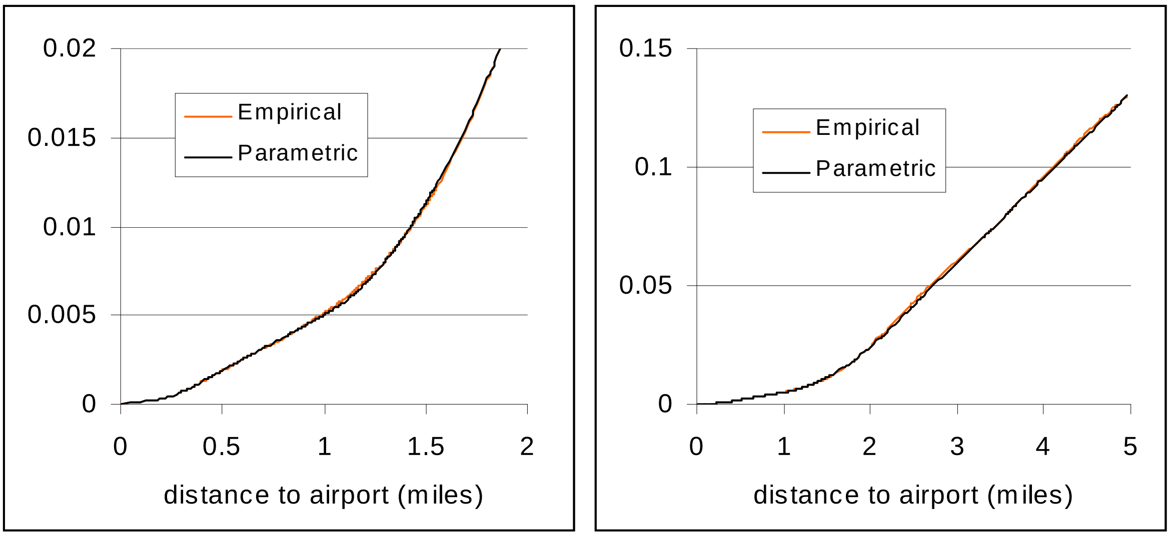

Figure 4.14 Parametric and empirical distribution function for $F_{Top100}(d)$ on two different scales.

4.4.7 Different airport lists

This section presents the final results for the different airport lists in Table 4.5. For each airport list the same analysis as for the Top100 list is performed. These analysis resulted in parametric distribution functions $F_L(d)$ for the Top250 and Top2250 airport lists and all three distribution functions are presented in Figure 4.15. In the next chapter, the distribution functions $F_L(d)$ are used to quantify the spatial variability of the risk associated with distance to an airport in list L.

Confidence bounds for $F_L(d)$ and $f_L(d)$ are needed to quantify the uncertainty of the risk. It is not possible to sensibly transform the 90%-confidence intervals for the distribution functions into 90%-confidence intervals for the density functions. Theoretically, the local uncertainty in the density functions is not bounded (the confidence interval of $F_L(d)$ does not exclude any possible slope of $F_L(d)$), but it is obvious that the parametric density functions reasonably describe the trends.

To quantify the uncertainty associated with an estimate of the groundling risk in the distance interval 2 to 3 miles of an airport, the distribution of $F_L(3)$-$F_L(2)$ is needed. These distributions were recognized as normal distributions and the parameters are presented in Table 4.7. The probability plots in Figure 4.16 show that the normal distribution is a reasonable fit for these variables.

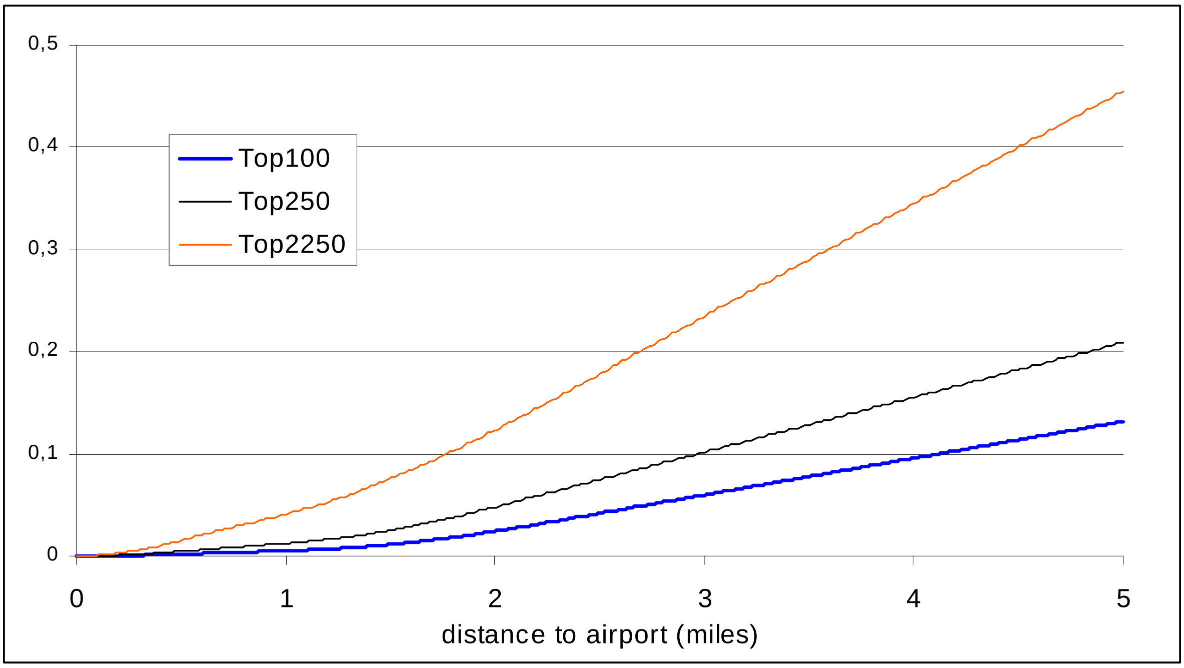

Figure 4.15 The parametric distribution functions $F_L(d)$ for L=Top100, Top250, Top2250.

$F_L(d)$		Top100	Top250	Top2250
	Distribution	Normal	Normal	Normal
$F_L(1)$	μ	0.0051	0.012	0.041
	σ	0.00049	0.00074	0.0033
	Distribution	Normal	Normal	Normal
$F_L(2)- F_L(1)$	μ	0.018	0.036	0.081
	σ	0.0026	0.0034	0.0055
	Distribution	Normal	Normal	Normal
$F_L(3)- F_L(2)$	μ	0.036	0.054	0.11
	σ	0.0032	0.0045	0.0068
	Distribution	Normal	Normal	Normal
$F_L(4)- F_L(3)$	μ	0.036	0.054	0.11
	σ	0.0033	0.0045	0.0068
	Distribution	Normal	Normal	Normal
$F_L(5)- F_L(4)$	μ	0.036	0.054	0.11
	σ	0.0033	0.0045	0.0068
	Distribution	Normal	Normal	Normal
$F_L(10)- F_L(5)$	μ	0.16	0.21	0.36
	σ	0.0032	0.0040	0.0051

Table 4.7 Mean and 90%-confidence intervals for $F_L(d)$ in 1990.

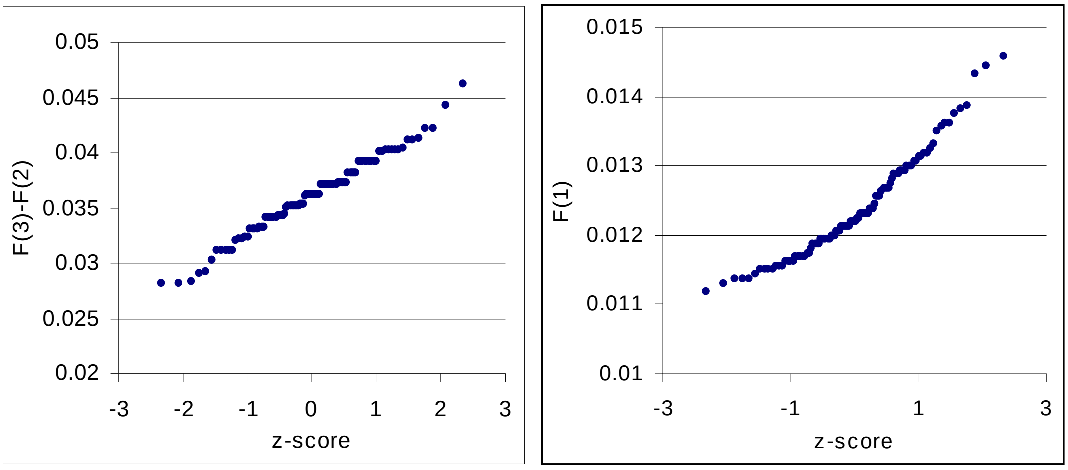

Figure 4.16 Probability plots for two uncertain variables $F_{Top100}(3)$-$F_{Top100}(2)$ and $F_{Top250}(1)$.

5 Risk of ground fatalities due to uncontrollable aviation accidents in the United States

Chapter 3 describes the crash distribution and fatality rates derived and Chapter 4 provides a Geographic Information System model developed to relate the U.S. resident population to a list of airports. This chapter combines the results of Chapters 3 and 4 and presents the final estimates of the current ground fatality risk and the spatial variability of this risk. Section 5.1 focuses on the average risk of ground fatalities and in Section 5.2 the variability is quantified. Section 5.3 compares the results of this analysis with the outcome of a study that evaluated the risk around a specific airport (Schiphol airport, Amsterdam, Netherlands).

5.1 Average risk of ground fatalities

5.1.1 Exposed population; U.S. residents

To estimate risks like $\mathrm{P}\left(\mathrm{B}_{2000}(\mathrm{X})\right)$ the following formula is applied:

$$\text{Annual Risk} = \frac{\text{Expected annual number of fatalities due to the risk}}{\text{population exposed to the risk}} \qquad (5.1)$$

This raises the question of who is exposed to the groundling risk. In some degree, everybody who lives in the U.S. is exposed to the risk of death due to a crashing airplane. This analysis assumes that the population exposed to the groundling risk is equal to the U.S. resident population as known to the U.S. Census Bureau. The bureau works under strict privacy regulations and is not allowed to provide any personal data to third parties like the immigration service or police. Thus while people are more willing to provide their personal data, these data are not available.

The U.S. Census Bureau provides projections of the U.S. resident population until 2100. The method used to produce the projections is described by Hollmann et al[1]. Future U.S. populations are derived from the projection of population change for each of its major demographic components: births, deaths and migration. The projections of these major components are driven by the composition of the population by age, sex, race, and Hispanic origin, and the way these variables influence the propensity to bear children, die, and migrate to or from the United States. The projections do not include a systematic measurement of uncertainty, but middle, low and high projections are generated for each demographic component. The low and high projections were combined to produce the extreme estimates for the future U.S. population[1]:

---

[1] Hollmann, F.W., Mulder, T.J. and Kallan, J.E., "*Methodology and Assumptions for the Population Projections of the United States: 1999 to 2100.*", January 2000.

Applying variant assumptions for each component individually resulted in the range of population series that would be identified with the maximum likely variance of that component. To produce our lowest and highest series, we combined the extreme values of all three major components that favored, respectively, the lowest and highest population growth. Therefore, the extreme projections do not represent likely scenarios in themselves, but purport to represent the extremes between which most likely outcomes should fall.

The extreme projections are rather conservative and cannot directly be associated with a confidence interval. Given the description of the extreme projections, interpreting the low and high projections as boundaries of a 90% confidence interval seems to be a conservative approach. The projections by the Census Bureau are presented in Table 5.1.

	Middle	Low	High
2000	275,306	274,853	275,816
2005	287,716	284,000	292,339
2010	299,862	291,413	310,910
2015	312,268	297,977	331,636

Table 5.1 Middle, low and high projections of the U.S. resident population provided by the U.S. Census Bureau (numbers in thousands).

5.1.2 "Observed" risk of ground fatalities

When Goldstein et al estimated the groundling risk in 1992, they approximated the expected number of ground fatalities per year by taking the average over an 11-year period. This was a reasonable approach because the number of groundling accidents was rather large in this period (this analysis identified 33 groundling accidents in 1975-1985 resulting in 75 ground fatalities, Goldstein et al assumed that 150 ground fatalities occurred in that same period). In the period 1987-1999 only 19 groundling accidents occurred and therefore a more analytical approach was chosen to estimate the expected number of ground fatalities per year (Chapter 3). However, it still is interesting to analyze the *observed* risk of ground fatalities in the last 30 years.

Table 5.2 shows the observed groundling risks for three periods. The numbers are calculated by dividing the average number of ground fatalities per year by the average population[2] in the period (formula (5.1)). The uncertainty of the risk in the 1964-1979 period results from a number of ground fatalities for which the degree of control the individuals had over the aviation activity that caused their deaths remained uncertain. The table clearly shows that the risk for groundlings due to aviation accidents has decreased in the last 30 years, both due to a decrease in the number of ground fatalities (Figure 3.2) and an increase in population.

	observed annual risk	corresponding 70-year lifetime risk
1964-1979	$[3.8\ 10^{-8}, 4.5\ 10^{-8}]$	$[2.7\ 10^{-6}, 3.2\ 10^{-6}]$
1980-1989	$2.4\ 10^{-8}$	$1.7\ 10^{-6}$
1990-1999	$8.5\ 10^{-9}$	$5.9\ 10^{-7}$

Table 5.2 Observed risk of ground fatalities in U.S. for different periods.

---

[2] http://www.census.gov/population/estimates/nation/popclockest.txt

5.1.3 Current average risk

The population or group risk, the expected number of ground fatalities due to aviation accidents, was estimated in Chapter 3 and the results are presented in Table 3.6. The group risk for 2000 was estimated as 3.5 ground fatalities. Applying equation (5.1) results in an estimate of the annual individual risk (using the middle estimate of U.S. resident population in 2000 is 275,306,000):

$$P\left(B_{2000}(X)\right) = \frac{3.5\,\text{ground fatalities}}{275{,}306{,}000} = 1.3\ 10^{-8},$$

(for interpretation $P(B_{2000}(X)$ see Section 2.3) and the corresponding 70-year lifetime risk is equal to 8.9 10^{-7}. Note that this estimate is below the commonly used risk management threshold of 1 in a million lifetime risk.

The estimate 1.3 10^{-8} is a factor 1.5 higher than the observed risk in 1990-1999 (8.5 10^{-9}), because it takes into account the rather rare accidents that kill a large number of groundlings (see Section 3.3.1). In the period 1990-1999 the largest number of ground fatalities associated with a single aviation accident in the U.S. was 2 fatalities.

5.1.4 Projections for the risk of ground fatalities in 2005, 2010 and 2015

Table 5.2 shows that the observed risk of ground fatalities has decreased in the last decades due to increases in aviation safety. The projections for the risk of ground fatalities presented in this section are based on a scenario analysis. This section discusses the outcome and evaluates the scenario analysis. Consider the following baseline scenario:

- The estimates for the airport activity provided by the FAA are correct (Table 3.2).
- All other quantities of interest for the ground fatality risk, including the ground fatality rates, will remain unchanged in the next 15 years.

The baseline scenario shows the consequences of projected airport activity for the risk of ground fatalities given that all other quantities remain unchanged. The projections of the airport activity are based on a scenario analysis that assumes a strong relationship between the economic climate and the aviation activity. Annual percent changes of activity are derived for each aviation category making fundamental economic assumptions concerning future economic growth, fuel prices and interest rates. The FAA does not provided any uncertainty analysis or extreme values.

The most important implication of the second assumption is that the ground fatality rate per operation is expected to remain unchanged for the next 15 year. In the last 50 years accident and fatality rates in aviation have significantly decreased due to technological developments[3]. However, the decrease has been flattening in the last decade[4] by a slightly increasing number of flights. The ground fatality rate per airport operation seems to follow the same trend, but it is more difficult to observe because groundling accidents are rather rare events. This analysis assumes that the fatality rates remain constant until 2015, even though the population is expected to grow from 275 million to 312 million in 2015. An increase in

---

[3] National Transportation Safety Board, *We are all safer now*, 1998.

[4] Federal Aviation Administration, *The Global Analysis and Information Network*, Office of System Safety, June 1997.

the population might increase the ground fatality rate, because a crashing aircraft is more likely to hit a groundling. On the other hand, new technologies and better information sharing between airlines, NTSB, and FAA[4] concerning near-accidents probably decreases the accident and fatality rates in general. In the last 13 years the groundling accident rate seemed to remain constant despite an increase in population from 242 million in 1987 to 275 million in 1999 and introductions of several new technologies. Probably, both effects canceled each other resulting in a approximately constant groundling accident and ground fatality rate per operation in the last decade. Therefore, ground fatality rates per operation remaining constant in the near future seems to be the most reasonable scenario.

The results are presented in Table 5.3. The baseline scenario predicts that the average individual risk of ground fatalities in the U.S. remains almost unchanged, although the population risk increases from 3.5 fatalities in 2000 to 4.3 fatalities in 2015. The individual risk remains constant because the U.S. resident population is expected to grow from 275 million in 2000 to 312 million in 2015.

Year	Population risk[5]	Middle projection of U.S. population (in thousands)	Annual individual risk	Corresponding 75-year lifetime risk
2000	3.5	275,306	$1.3\ 10^{-8}$	$9.5\ 10^{-7}$
2005	3.8	287,716	$1.3\ 10^{-8}$	$9.9\ 10^{-7}$
2010	4.0	299,862	$1.3\ 10^{-8}$	$1.0\ 10^{-6}$
2015	4.3	312,268	$1.3\ 10^{-8}$	$1.0\ 10^{-6}$

Table 5.3 Estimated risk of ground fatalities due to uncontrollable aviation accidents 2000-2015.

5.2 Spatial variability of the groundling risk: results.

5.2.1 Mathematical formulation of the variability of the risk

In Chapter 2 a mathematical representation of the variability of the risk of ground fatalities due to uncontrollable aviation accidents was derived. However, introducing airport-related and airport-unrelated accidents, aggregating ground fatality rates by aviation category and defining different crash distributions for commercial aviation and general aviation requires a more sophisticated model to quantify the variability of the risk. This Section describes the adjustments that were made, but the basic approach discussed in Chapter 2 remains unchanged.

Consider the spatial variability of the risk associated with airports in list L in 2000. Assume a random resident X and define the following function:

$D_L(X,t)$: Distance between individual X and the nearest airport in list L at time t.

Consider the following events:

---

[5] Expected annual number of ground fatalities

$B_{2000,L\text{- related}}(X)$: {X dies due to an airport-related[6] uncontrollable aviation accident in 2000, related airport is in list L}

$B_{2000,\text{ unrelated}}(X)$: {X dies due to an airport-unrelated uncontrollable aviation accident in 2000}

$A_{2000}^{d_L}(X)$: $\{D_L(X,t) = d : t \in [2000,2001)\}$

Events $B_{2000,L\text{-related}}(X)$ and $B_{2000,\text{unrelated}}(X)$ are disjunct and define $B_{2000,L}(X)$ as follows:

$B_{2000,L}(X) = B_{2000,L\text{-related}}(X) \cup B_{2000,\text{unrelated}}(X)$

$P\left(B_{2000,L}(X) \mid A_{2000}^{d_L}(X)\right)$ is the quantity that represents the spatial variability of the risk associated with the airports in list L. This quantity can be written as (disjunct events):

$$P\left(B_{2000,L}(X) \mid A_{2000}^{d_L}(X)\right) = P\left(B_{2000,L\text{-related}}(X) \mid A_{2000}^{d_L}(X)\right) + P\left(B_{2000,\text{unrelated}}(X) \mid A_{2000}^{d_L}(X)\right) \tag{5.2}$$

This analysis assumes that the risk of ground fatalities due to airport-unrelated accidents is location independent, so:

$$P\left(B_{2000,\text{unrelated}}(X) \mid A_{2000}^{d}(X)\right) = P\left(B_{2000,\text{unrelated}}(X)\right) \tag{5.3}$$

The risk of ground fatalities due to airport-unrelated accidents is estimated in Section 5.2.2. The following events are defined to estimate the first term on the right hand side of (5.2):

$B_{2000,L\text{-related}}^{AC}(X)$: {X dies due to an uncontrollable, airport-related, air carrier accident in 2000, related airport is in list L}

$B_{2000,L\text{-related}}^{AT}(X)$: {X dies due to an uncontrollable, airport-related, air taxi accident in 2000, related airport is in list L}

$B_{2000,L\text{-related}}^{GA}(X)$: {X dies due to an uncontrollable, airport-related, general aviation accident in 2000, related airport is in list L}

These events are disjunct because in the data analysis ground fatalities due to collisions of aircraft of different categories are evenly divided over the categories and are associated with only one category. A model that separately included collisions between aircraft has been considered, but was rejected due to the lack of data to support such a model. Consequently,

$$\begin{aligned} P\left(B_{2000,L\text{-related}}(X) \mid A_{2000}^{d_L}(X)\right) = {} & P\left(B_{2000,L\text{-related}}^{AC}(X) \mid A_{2000}^{d_L}(X)\right) + P\left(B_{2000,L\text{-related}}^{AT}(X) \mid A_{2000}^{d_L}(X)\right) + \\ & + P\left(B_{2000,L\text{-related}}^{GA}(X) \mid A_{2000}^{d_L}(X)\right) \end{aligned} \tag{5.4}$$

---

[6] *An accident is related to an airport if and only if the airport is the origin or final destination of a flight and the distance between airport and accident scene is less than 10 miles.*

In Chapter 2, an expression for $P\left(B_{2000}(X) \mid A^{d}_{2000}(X)\right)$ is derived by applying Bayes' theorem. The same approach can be applied to each of the terms on the right hand side of (5.4), resulting in (compare to (2.11)):

$$P\left(B^{AC}_{2000,L\text{-related}}(X) \mid A^{d_L}_{2000}(X)\right) = \frac{g^{AC}_{X,2000}(d)}{f_{X,2000,L}(d)} P\left(B^{AC}_{2000,L\text{-related}}(X)\right), \tag{5.5}$$

$$P\left(B^{AT}_{2000,L\text{-related}}(X) \mid A^{d_L}_{2000}(X)\right) = \frac{g^{AT}_{X,2000}(d)}{f_{X,2000,L}(d)} P\left(B^{AT}_{2000,L\text{-related}}(X)\right), \tag{5.6}$$

$$P\left(B^{GA}_{2000,L\text{-related}}(X) \mid A^{d_L}_{2000}(X)\right) = \frac{g^{GA}_{X,2000}(d)}{f_{X,2000,L}(d)} P\left(B^{GA}_{2000,L\text{-related}}(X)\right). \tag{5.7}$$

where

$F_{X,2000,L}(d)$: The probability that X is within distance d of an airport in list L at a random time in 2000.

$G^{AC}_{X,2000}(d)$: The probability that X dies within distance d of the related airport given that X dies due to an uncontrollable, airport-related, *air carrier* accident in 2000.

$G^{AT}_{X,2000}(d)$ and $G^{GA}_{X,2000}(d)$ can be similarly explained. In fact, $G^{AC}_{X,2000}(d)$, $G^{AT}_{X,2000}(d)$ and $G^{GA}_{X,2000}(d)$ can be seen as groundling crash distribution functions for the three aviation categories. These distribution functions have been estimated in Section 3.4.2, equations (3.3) and (3.4).

In this analysis $F_{X,2000,L}(d)$ and $f_{X,2000,L}(d)$ are approximated by $F_{X,1990,L}(d)$ and $f_{X,1990,L}(d)$. $F_{X,1990,L}(d)$ is strictly based on 1990 U.S. census data and derived in Chapter 4. Figure 4.15 presents $F_{X,1990,L}(d)$ for different airport lists L.

Combining equations (5.2) to (5.7) results in the following expression:

$$\begin{aligned} P\left(B_{2000,L}(X) \mid A^{d_L}_{2000}(X)\right) &= P\left(B_{\text{unrelated}}(X)\right) + \\ &+ \frac{g^{AC}_{X,2000}(d) \times P\left(B^{AC}_{2000,L\text{-related}}(X)\right) + g^{AT}_{X,2000}(d) \times P\left(B^{AT}_{2000,L\text{-related}}(X)\right) + g^{GA}_{X,2000}(d) \times P\left(B^{GA}_{2000,L\text{-related}}(X)\right)}{f_{X,2000,L}(d)} \end{aligned} \tag{5.8}$$

The spatial variability of the risk associated with the dimension of distance to an airport in L is quantified in this Chapter by using equation (5.8). The quantity $P\left(B_{2000,L}(X) \mid A^{d_L}_{2000}(X)\right)$ can be interpreted as follows:

$P\left(B_{2000,L}(X) \mid A^{d_L}_{2000}(X)\right)$ *is the hypothetical risk for a groundling of being killed due to an uncontrollable aviation accident in 2000 given that the groundling stayed at a distance d away from the nearest airport in list L for the whole year.*

5.2.2 Airport-unrelated groundling risk

Chapter 3 showed that most groundling accidents occurred in the vicinity of airports, but that a significant number of ground fatalities were not at all related to airports. The expected number of ground fatalities in 2000 caused by air taxi and air carrier accidents is 1.5 and 2.0 for general aviation accidents (Table 3.6). Table 3.7 showed that 87% of the commercial aviation accidents is airport-related and 63% of the general aviation accidents is related to an airport. This implies that the expected number of airport-unrelated ground fatalities in 2000 is equal to:

$1.5 \times 0.13 + 2.0 \times 0.37 = 0.94$ airport-unrelated ground fatalities.

Airport-unrelated ground fatalities can occur everywhere in the United States, but flight paths and other crowded aviation areas are more likely to be the scene of an accident. Although there is some variability in the airport-unrelated groundling risk, the variability in this risk is relatively small compared to variability in the airport-related groundling risk. Consequently, it is reasonable to assume that the airport-unrelated groundling risk is location independent. The risk of airport-unrelated ground fatalities in 2000 can be estimated by dividing the expected number of fatalities in 2000 by the U.S. population in 2000 (formula (5.1)):

$$P\left(B_{2000,\text{unrelated}}(X)\right) = \frac{0.94}{275{,}306{,}000} = 3.4\ 10^{-9}\ . \qquad (5.9)$$

U.S. residents who live within 10 miles of an airport are exposed to both airport-unrelated and airport-related risk (next section). Section 5.2.5 discusses the assumptions that are made to estimate the airport-unrelated and the airport-related groundling risk.

5.2.3 Airport-related groundling risk

In this section the spatial variability of the groundling risk in the dimension distance to an airport is estimated by applying equation (5.8). The analysis is presented for the airport list Top100 and the results are shown for the other two lists, Top250 and Top2250.

Consider the Top100 airports as selected in Section 4.3.1 and equation (5.8). The following variables and functions are required as input:

$P\left(B_{2000,\text{unrelated}}(X)\right)$: Estimated as $3.4\ 10^{-9}$, Equation (5.9)

$g^{AC}_{X,2000}(d)$, $g^{AT}_{X,2000}(d)$ and $g^{GA}_{X,2000}(d)$: Equation (3.3) and (3.4) in Section 3.4.2

$f_{X,2000,\text{Top100}}(d)$: Approximated by $f_{X,1990,\text{Top100}}(d)$, see Figure 4.15

$P\left(B^{AC}_{2000,\text{Top100-related}}(X)\right)$, $P\left(B^{AT}_{2000,\text{Top100-related}}(X)\right)$ and $P\left(B^{GA}_{2000,\text{Top100-related}}(X)\right)$: To be estimated.

$P\left(B^{AC}_{2000,\text{Top100-related}}(X)\right)$ is estimated by applying formula (5.1):

$$P\left(B^{AC}_{2000,\text{Top100-related}}(X)\right)=\frac{\begin{array}{c}\text{expected number of ground fatalities due to uncontrollable,}\\ \text{Top100 - related, air carrier accidents in 2000}\end{array}}{\text{U.S. population in 2000}} \tag{5.10}$$

The numerator of (5.10) is estimated by taking the product of the ground fatality rate per air carrier operation (Table 3.5), the total number of air carrier operations at Top100 airports and the percentage of accidents that is airport-related (Table 3.7). Results are presented in Table 5.4. Projections of the total number of air carrier operations at Top100 airports in 2000 are provided by the FAA Terminal Area Forecast system[7]. Calculations are performed for all three aviation categories and for 2000, 2005, 2010 and 2015.

The numbers in the last column of Table 5.4 are filled in for the numerator of (5.10), resulting in (estimates of the U.S. resident population are presented in Table 5.1):

$$P\left(B^{AC}_{2000,\text{Top100-related}}(X)\right)=\frac{0.54}{275{,}306{,}000}=2.0\ 10^{-9}$$

$$P\left(B^{AT}_{2000,\text{Top100-related}}(X)\right)=\frac{0.32}{275{,}306{,}000}=1.2\ 10^{-9}$$

$$P\left(B^{GA}_{2000,\text{Top100-related}}(X)\right)=\frac{0.13}{275{,}306{,}000}=4.7\ 10^{-10}$$

This means that estimates are available for all the required inputs of equation (5.8). The same analysis has been performed for the Top250 and Top2250 airport lists. Results for all three airport groups are shown in Figure 5.1. Recall that the quantity $P\left(B_{2000,L}(X)\,|\,A^{d_L}_{2000}(X)\right)$ is a hypothetical risk, it represents the probability for a groundling of death due to an uncontrollable aviation accident in 2000 given that this groundling stays at a distance d away from the nearest airport in L for the whole year.

The following can be concluded from Figure 5.1:

1. Risks of ground fatalities are higher in the vicinity of an airport. The risk varies by a factor 100 in the dimension distance to an airport. The variability mainly applies to the first 2 miles around an airport. The average risk further than 2 miles from an airport is below 10^{-8} annually regardless the distance.

2. Risks of ground fatalities associated with busier airports are higher than risks associated with less busy airports. Again, the variability in the ground fatality risk mainly applies to the first 2 miles around an airport. The estimate of the average annual risk within 0.2 miles of a Top100 airport just exceeds 10^{-6}.

Note that $P\left(B_{2000,L}(X)\,|\,A^{d_L}_{2000}(X)\right)$ only takes into account the risk associated with the nearest airport. This means that the analysis might not be completely valid if airports are highly clustered. This problem does

---

[7] http://www.apo.data.faa.gov/faatafall.HTM

not show up for the Top100 airports, but might be a problem for the larger airport lists. The software program in Section 4.3.2 provided for each sampled resident the two nearest airports. The output about the second nearest airport showed that even for the Top2250 list the second nearest airport tends to be further than 10 miles away and that it is very rare that the second nearest airport is within 2 miles. The spatial variability of the risk applies mainly to the first two miles around an airport (Figure 5.1) and consequently the results remain valid.

	Year	Estimate of the total number of airport operations at Top 100 airports (in millions)	Ground fatality rate (per million operations)	Percentage airport-related accidents	Expected number of Top100-related ground fatalities
Air Carrier	2000	13.26	0.047	0.87	0.54
	2005	15.10	0.047	0.87	0.62
	2010	17.18	0.047	0.87	0.70
	2015	19.40	0.047	0.87	0.79
Air Taxi	2000	6.93	0.053	0.87	0.32
	2005	7.78	0.053	0.87	0.36
	2010	8.61	0.053	0.87	0.40
	2015	9.44	0.053	0.87	0.44
General Aviation	2000	11.35	0.018	0.63	0.13
	2005	12.26	0.018	0.63	0.14
	2010	13.24	0.018	0.63	0.15
	2015	14.31	0.018	0.63	0.16

Table 5.4 Estimates of the expected number of Top100-related ground fatalities for air carriers, air taxis and general aviation.

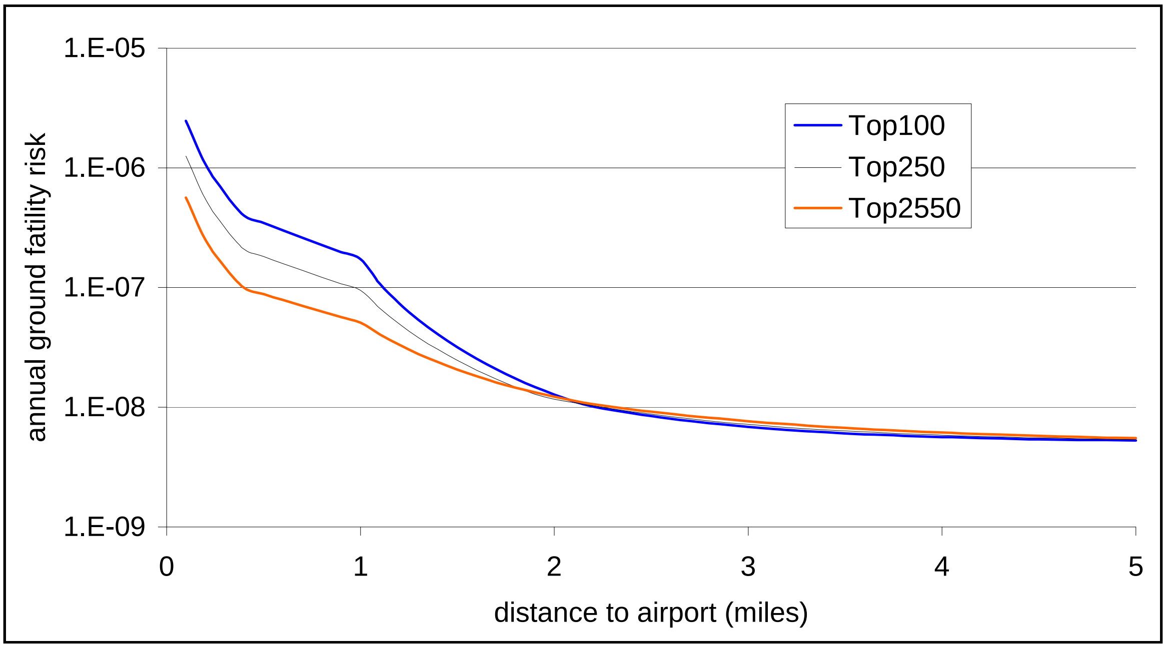

Figure 5.1 Variability of the risk of ground fatalities in the dimension distance to an airport represented by the quantity $P\left(B_{2000,L}(X) \mid A_{2000}^{d_L}(X)\right)$, for L=Top100, Top250, Top2250.

5.2.4 Uncertainty analysis

In this section the uncertainty in the groundling risk associated with the following sources is quantified:

- Uncertainty in the estimates for the U.S. resident population (Section5.1.1).
- Uncertainty of distribution function $F_{X,1990,Top100}(d)$ (Section 4.4.6).

The U.S. Census Bureau provided low, middle, and high estimates for the population for 2000, 2005, 2010, and 2015. The low and high estimates *represent the extremes between which most likely outcomes should fall* and in this analysis these estimates are interpreted as the bounds of a 90%-confidence interval. The U.S. population in future years is expected to be normally distributed and the mean μ is chosen equal to the middle estimate. The standard deviation σ was estimated by:

$$\sigma = \frac{\text{high estimate population - low estimate population}}{z_{95\%} - z_{5\%}} .$$

The parameters for the normal distributions of the U.S. population for 2000, 2005, 2010, and 2015 are shown in Table 5.5. Interpreting the extreme values as bounds of a 90%-confidence interval is a rather conservative approach and it is more likely that the uncertainty is overestimated than underestimated.

	Distribution	μ	σ
2000	Normal	275,306,000	585,463
2005	Normal	287,716,000	5,069,754
2010	Normal	299,862,000	11,853,339
2015	Normal	312,268,000	20,463,227

Table 5.5 Distributions for the U.S. resident population in 2000, 2005, 2010 and 2015.

The Geographical Information System modeling in Chapter 4 provided d-wise confidence bounds for $F_{X,1990,L}(d)$, but these bounds cannot be reasonably transformed to bounds for $f_{X,1990,L}(d)$. Therefore, the uncertainty analysis focused on quantities like $P\left(B_{2000,L}(X) \mid A_{2000}^{[0,1)_L}(X)\right)$ with the following interpretation:

$P\left(B_{2000,L}(X) \mid A_{2000}^{[0,1)_L}(X)\right)$ *is the hypothetical risk for a groundling of being killed due to an uncontrollable aviation accident in 2000 given that the groundling stayed within 1 mile of an airport in list L for the whole year.*

$P\left(B_{2000,L}(X) \mid A_{2000}^{[2,3)_L}(X)\right)$, $P\left(B_{2000,L}(X) \mid A_{2000}^{[3,4)_L}(X)\right)$, etcetera are defined similarly. To estimate these risks, $F_{X,1990,L}(1)$, $F_{X,1990,L}(2)$-$F_{X,1990,L}(1)$ and $F_{X,1990,L}(3)$- $F_{X,1990,L}(2)$ are required and for these quantities uncertainty distributions have been derived (see Table 4.7).

A Monte Carlo simulation was performed to quantify the uncertainty in the groundling risk for each interval. During the simulation the uncertain parameters were considered independent. The results for the groundling risk in the vicinity of Top100 airports in 2000 are presented in Figure 5.2. Despite the uncertainty, Figure 5.2 clearly shows that the risk in [0,1] is significantly higher than in the other intervals. The variability seems to be a more important factor than the uncertainty. However, uncertainties associated with some assumptions have not been quantified and are discussed in the next section.

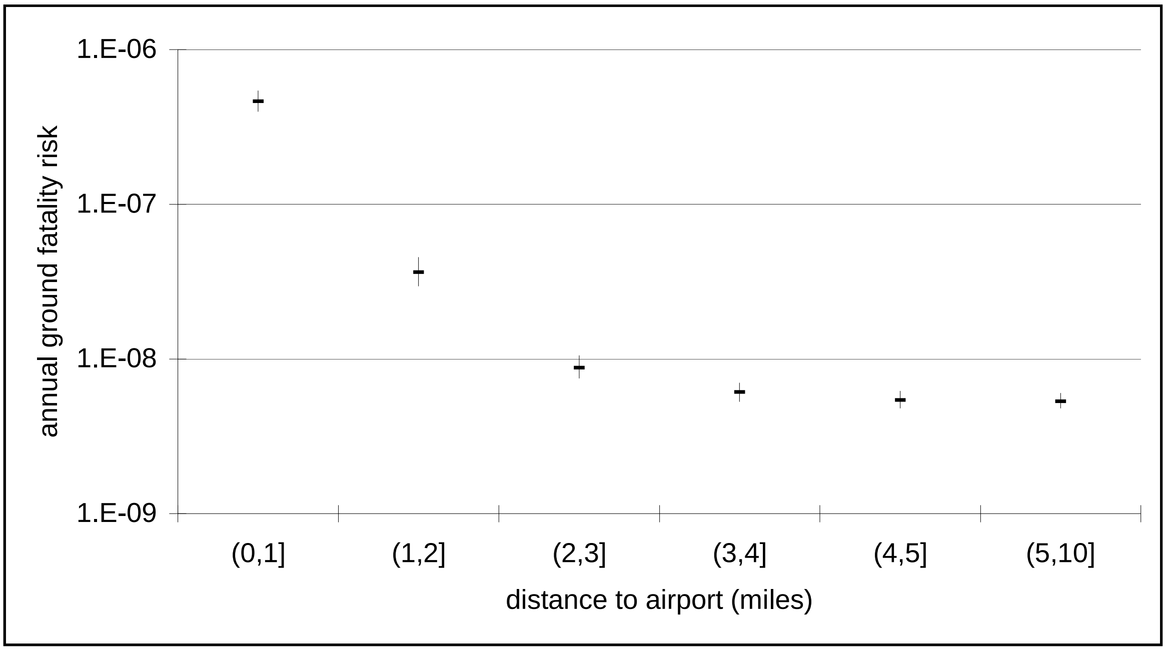

Figure 5.2 Plot of the uncertainty in the ground fatality risk in 2000 for each distance interval (mean and 90% confidence interval).

5.2.5 Assumptions

The key assumptions that are made to estimate the groundling risk and that are not taken into account in the uncertainty analysis are the following:

- Constant fatality rates among airports.
- $F_{X,2000,L}(d)$ does not depend on the time and is approximated using 1990 census data.

This analysis assumes that the expected number of ground fatalities per year only depends on the airport operations (i.e. number of flights) and applies a constant fatality rate per aviation category. This approach raises the following question: Do the ground fatality rates depend on the airport, in other words does the groundling safety differ among airports?

The answer to this question is simple; yes, ground fatality rates per operation differ among airports. Local circumstances like climate, population density around the airport, length of the runways etc. influence the ground fatality rate. For each aviation category, the applied rate is an estimate of the average ground fatality rate per operation for all airport operations in the United States. Because the fatality rates differ among airports, it does not make sense to apply this approach to estimate the expected number of ground fatalities per year related to a single airport. Therefore, this method is only applied to lists of at least 100 airports that account for a significant part of the total airport activity. However, the Top100 airport list only contains busy airports and it might be possible that the fatality rates at the larger airports are different than at the smaller airports. To verify whether constant ground fatality rates per operation can be applied to the airport lists used in this analysis, the observed ground fatalities per airport list were compared with the projections based on constant fatality rates per operation. The TAF-system again provided estimates of the airport activity for each airport in the list for the period 1987-1999. The expected number of ground fatalities per airport was calculated (same approach as in Section 5.2) assuming constant ground fatality rates per operation for each aviation category and summed for each list. According to this assumption, the Top 100 airports should have accounted for 38% of the airport-related ground fatalities between 1987 and 1999. Evaluation of the accident data revealed that 6 of the 18 (33%) airport-related fatalities occurred in the vicinity of a Top100 airport. Considering the small number of ground fatalities in 1987-1999, the 5% difference does not support a rejection of the assumption. Similar analysis of the Top250 and Top2250 airport lists did not provide serious indications to doubt the assumption concerning constant fatality rates. Consequently, it seems reasonable to assume constant ground fatality rates among groups of airports that account for a significant part of the total airport activity in the United States. A more sophisticated model[8,9] is needed to estimate the groundling risk in the vicinity of a specific airport.

	Model	Observed
Top100	38%	33%
Top250	53%	44%
Top2250	95%	94%

Table 5.6 Percentage of ground fatalities per airport list according to the model (assuming constant fatality rates) and the observed percentage in the 1987-1999.

Two assumptions were made to derive an approximation of $F_{X,,2000,L}(d)$. First, it was assumed that $F_{X,2000,L}(d)$ is stationary within the year 2000 and this stationary distribution is estimated based on where people live according to the U.S. census data. Second, U.S. census data from 1990 was used to estimate the distribution.

The stationary assumption (discussed in Section 4.1) does not directly ignore all mobility. It is assumed that the total mobility does not influence the distribution. In other words, each resident who leaves his residence area is expected to be replaced by some one else. The facts that the area at stake is the whole United States and that the distribution is parameterized by distance to an airport strengthen this assumption.

The 2000 census data were not available at the time of this study and therefore the 1990 data were used. This analysis assumes that $F_{X,2000,L}(d)$ can be approximated by $F_{X,1990,L}(d)$. It is hard to analyze whether this distribution has significantly changed or to quantify the uncertainty associated with this assumption. An analysis of the 2000 census data can answer this question, but is not option now.

Although not all uncertainties have been quantified, it seems justified to conclude that variability is a more important factor than uncertainty. Considering Figure 5.2, it is hard to imagine that including these other uncertainties will dramatically change the trend as shown in the figure.

---

[8] Hillestad, R., Solomon, K., Chow, B. et al, *Airport Growth and Safety, A study of the external risk of schiphol airport and possible safety-enhancement measures,* RAND, 1993.

[9] Piers, M.A. et al, *Study of the external risk around Schipol airport*, National Aerospace Laboratory (The Netherlands), 1993.

5.2.6 Variability in other dimensions

Figure 5.1 and Figure 5.2 present the spatial variability of the groundling risk related to the distance to an airport. The figures show that people who spend most of their time in the vicinity of an airport are at higher risk than others. However, the spatial variability of the groundling risk is also related to other dimensions. For example, busy airports expose their neighbors on average to a higher risk than non-busy airports. The spatial variability related to this dimension has been determined by analyzing different airport groups. Figure 5.1 clearly shows that risks in the vicinity of busier airports (Top100) are significantly higher.

However, there are some important dimensions for which the spatial variability has not been estimated. A dimension that has been ignored concerns the flight paths of aircraft. The risks of ground fatalities associated with airport-unrelated accidents are assumed to be uniformly spread out over the U.S. population. But it is hard to believe that someone living in the middle of Arizona where an aircraft passes once every two years is at equal risk as someone who lives right beneath a busy flight path (even if it is at least 10 miles from an airport). However, the average airport-unrelated groundling risk is rather low and estimated as $3.4\ 10^{-9}$ in 2000. Although the airport-unrelated risk is variable among the U.S. population, it is obvious that this variability can be neglected compared to the variability of the airport-related risk shown in Figure 5.1.

Accident data show that most aircraft crash in the path of a runway. This observation implies that residents who live at the end or the extended path of a runway are on average at higher risk than those who do not. It is hard to say what the magnitude of the variability is in this dimension, but the survey that estimated the ground fatality risk around the Schiphol airport identified individuals who were exposed to an annual risk of 10^{-4}. These individuals were working or living right at the end of a runway and it concerned a small group of people. Figure 5.1 shows that the annual risk within 0.2 miles of an airport is approximately 10^{-6} implying that the runway dimension can account for another order of magnitude of variability. Section 5.3 discusses the Schiphol Survey and compares the results with the outcomes of this analysis.

5.3 Comparison with the Schiphol survey performed by RAND in 1993

5.3.1 Description of the model used to evaluate the external risk of Schiphol Airport

On October 4th, 1992 an El Al carrier crashed into an apartment building in the residential area ‘de Bijlmermeer’, near Amsterdam. As a result, RAND was asked by the Dutch government to perform an analysis of the risks to residents due to crashing planes in the areas surrounding Schiphol Airport. RAND undertook an evaluation of the external risks at Schiphol and concluded that Schiphol was a safe airport. The survey was presented in three reports[10] and a set of proceedings[11] of a conference organized by

---

[10] Hillestad, R., Solomon, K., Chow, B., et al, *“Airport Growth and Safety, a Study of the External Risk of Schiphol Airport and Possible Safety-Enhancement Measures*”, RAND, 1993.
Hillestad, R., Solomon, K., Chow, B., et al, *“Airport Growth and Safety, Executive Summary of the Schiphol Project”,* RAND, 1993.
Brady, S.D, Hillestad, R., *Modeling the External Risks of Airports for Policy Analysis*, RAND, 1995.

RAND on the subject of Managing Safety In and Around Airports. In this Section the risk model that RAND developed and applied to Schiphol Airport is described and the results are compared with the outcomes of this analysis.

The analysis that led to RAND’s model can be split up in the following steps:

1. Analysis of international accidents rates to derive accident rates that are applicable to Schiphol.

2. Analysis of 53 crash sites to construct a crash distribution function.

3. Collection of population data around Schiphol.

4. Development of a quantitative model to estimate the risk of a crashing airplane killing an individual on the ground.

International accident rates were used because too few accidents occurred at Schiphol Airport to generate sensible accident rates. However, all airports are different and the accident rates were adjusted to Schiphol by omitting accidents that could not have occurred at Schiphol. RAND determined by examining 114 hull losses world wide that 14% of these accident could be excluded. For example, RAND excluded crashes that occurred during flight shows and airplanes crashing into mountains, because Schiphol does not organize flight shows and the area around Schiphol is relatively flat. The accident rates depend on the size of the aircraft and the model distinguished three categories (large, medium, and small aircraft).

RAND constructed a two dimensional crash density function and the end of the runway was chosen as the reference point. The crash distribution function was modeled as follows:
- 58% of the crashes occur along the centerline of the end of the runway. The distance from the end of the runway is exponentially distributed with a mean of 3.5 miles.
- 42% of the crashes do not occur along the centerline. The distance from the end of the runway, x-coordinate, for these crashes is assumed to be normally distributed (mean and standard deviation both of about 5 miles), and the distance from the centerline, y-coordinate, is modeled as a normal distribution with mean zero and standard deviation of about 6 miles.

The parameters were estimated based on an evaluation of 53 crashes and their locations and RAND tried several other distributions but decided that this model was the best fit considering the data.

The population data around Schiphol airport was provided by a Dutch institution in Delft. The cell size for each data point was 100 by 100 meters and the population within each cell was estimated for 1991, 2003 and 2015, and for business hours as well as non-business hours.

RAND evaluated the possible consequences for each airport operation (landing or takeoff) that takes place in a certain year by a scenario analysis. An aircraft that takes off or lands from a certain runway can crash or not and the probability of a crash is equal to the applicable accident rate. If the aircraft crashes, for each cell the probability that it is affected is estimated using the crash distribution, the affected area by the crash and the location of the cell. The affected area of a crash depends on the aircraft size, type of operation (aircraft carrier more fuel during take off) and the angle of the crash. If a cell is affected by a crash, the percentage of people who die in this cell depends on the aircraft size and the type of operation.

---

[11] Conference Proceedings, Managing Safety In and Around Airports, RAND, 1995.

These percentages vary from 0.9 (takeoff and large aircraft) to 0.15 (landing and small aircraft). Using the population data per cell leads to an estimate of the expected number of fatalities per cell for that particular flight. Summing over all the flights in a certain year results in the expected number of ground fatalities per cell per year.

Summing the fatalities over all cells results in an estimate of the expected annual number of ground fatalities around Schiphol (group risk). The individual risk in particular cell can be estimated by dividing the expected number of fatalities in the cell by its population. This implies that the individual risk is assumed to be constant for all persons within the same cell. The analysis resulted in a group risk of 0.6 expected ground fatalities in 2003.

5.3.2 Scenario-based versus data-based approach

The model that RAND developed to evaluate the risk to groundlings around Schiphol airport and the approach of this analysis are different, but more complementary than competitive. The main differences are the following:

- RAND's model can be used to evaluate the risk of a single airport and to support local safety decisions, while this analysis focuses on the risk to groundlings in the whole United States and only distinguishes between groups of airports. Therefore, the results can be used in public health analysis and risk communication issues.
- The model RAND used is mostly scenario based, while this analysis applied a more data-based approach.

Although Schiphol Airport can be compared with American airports, the differences between the two approaches make it difficult to compare the results of both studies. According to RAND's model, the group risk, or number of expected fatalities around Schiphol is 0.6 ground fatalities per year. This analysis estimated that the expected number of ground fatalities due to crashing aircraft in the U.S. is 3.5 in 2000 and 3.8 in 2005. Based upon its airport activity in 2003 (see Table 5.7), Schiphol Airport would be ranked somewhere around place 25 to 30 among U.S. airports. However, the expected number of ground fatalities related to one of the Top100 airports in the U.S. is 0.99 in 2000 according to this analysis. It is hard to imagine that the Top100 airports together only account for less than twice as many fatalities a year than Schiphol itself. From that perspective, the 0.6 looks quite high or the 0.99 quite low.

In the last ten years only 6 ground fatalities occurred at Top100 airports and 43 ground fatalities, in one accident, occurred at Schiphol. The evaluation undertaken by RAND concluded that Schiphol is a safe airport and that the external risks to groundlings are comparable with the external risks associated with other airports in Europe and the United States. However, this analysis estimated the number of expected ground fatalities related to the Top100 airports as 0.99 (average of 0.01 per airport), while Schiphol accounts for 0.6 fatalities per year according to RAND.

The 0.99 is based on an analysis of historical data concerning crashes that accounted for ground fatalities with little or no control over the aviation activity that caused their death. The estimates of the groundling accident rates per operation in the U.S. are based on accidents and activity data for the period 1987-1999. The crash consequences were modeled as the number of ground fatalities per groundling accident, of course depending on the category. For example, the expected number of ground fatalities per accident was estimated using data for 1964-1999 resulting in 4.3 ground fatalities per air carrier crash, given at least one ground fatality (see Table 2.3 for crash consequences for air taxis and general aviation).

In the last 36 years, 205 U.S. residents were killed as a result of an aviation accident over which they had little or no control and the degree of control for 23 fatalities remained unknown. This corresponds to an average of 6.3 fatalities a year (including the 23 fatalities in the analysis) for the whole U.S. In the last 22 years 33 ground fatalities occurred due to aviation accidents at Top100 airports, averaging 1.5 fatalities a year.

Consider the following:

- the number of large and medium aircraft operations at Schiphol in 2003 is approximately 2% of the total number of air carrier[12] operations at all Top100 airports in 2000,
- the number of small commercial aircraft operations at Schiphol in 2003 is approximately 0.7% of the total number of air carrier operations at all Top100 airports in 2000,
- the expected number of ground fatalities related to Top100 airports for 2000 is estimated as 0.99 and for 2005 as 1.09,
- the average number of ground fatalities related to Top100 airports over the last 22 years is 1.5 and the time series analysis in Section 3.2. showed that this number has been decreasing until the last decade, and
- Schiphol is comparable to most U.S. airports concerning safety according to RAND,

Given these insights, the questions remains: why is the group risk at Schiphol according to RAND's model is so much higher than this analysis of the risk to groundlings in the U.S. would suggest? The probable reason lies in the difference between how this analysis and RAND's analytical model differently treat the crash consequences. This analysis uses empirical data, while RAND applies a crash consequence model based on affected areas and the percentage of people that die within these areas. These crash consequence models assume that the pilots do not try to avoid densely-populated areas or buildings, but they crash according to the crash distribution. This assumption leads to higher probabilities of ground fatalities because pilots often try and succeed to avoid killing people on the ground. The probabilities of accidents with a lot of fatalities on the ground become larger when you assume that pilots do not try to avoid buildings. However, even for their own and their passengers' chances of surviving, they will try to avoid a collision with a building. Although sometimes a pilot totally loses control and crashes into building causing a lot fatalities on the ground, for example the El Al crash in Amsterdam in 1992. But this appears to be a more rare event than RAND's model predicts, at least in the U.S.

Although the previous paragraph criticizes an assumption of the Model RAND developed to evaluate the risk to groundlings around Schiphol, it strengthens RAND's conclusion that Schiphol is a safe airport. On the other hand, RAND's model and this analysis both do not answer the question why an accident involving 43 ground fatalities "happened" to occur at Schiphol and not at another airport.

RAND's model is a sophisticated approach to evaluate the external risks associated with a single airport. The model generates risk contours that show individual risks on a 2-Dimensional scale and supports policy decisions concerning the land use and safety at and around Schiphol Airport. However, one should be careful in identifying the external risks associated with an airport as unacceptable only based on RAND's model. A comparison between their results for Schiphol Airport and empirical data for the

---

[12] An aircraft with a seating capacity of more than 60 seats or a maximum payload capacity of more than 18,000 pounds (=8165 kg) carrying passengers or cargo for hire or compensation.

whole U.S. showed that RAND's model is conservative and probably overestimated the risk of ground fatalities.

Aircraft size	Characterization of aircraft		Number of airport operations at Schiphol
	by seat capacity	by maximum takeoff weight	
Large	>180 seats	>105,000 kg	97,377
Medium	70-180 seats	27,000-105,000 kg	173,614
Small, commercial	<70 seats	<27,000 kg	49,009
Total			320,000

Table 5.7 Projections of the number of airport operations at Schiphol airport in 2003.

6 Conclusions and discussion

In the previous chapter the results from the data analysis and the geographic information system model were combined and the variability of the risk of ground fatalities was quantified in the dimension distance to an airport. This chapter sums the important results and conclusions and puts it in a risk communication perspective.

6.1 Summary of results

6.1.1 Average risk of ground fatalities in the United States

This analysis estimated the population risk in 2000 in the United States as 3.5 ground fatalities. A baseline scenario, all quantities of interest, including fatality rates, remain constant and airport activity increases according to FAA estimates, resulted in projections of the group risks of 3.8 fatalities in 2005, 4.0 in 2010 and 4.3 in 2015. The corresponding individual risks are calculated by dividing the expected number of ground fatalities by the U.S. resident population, and are presented in Table 6.1. The table shows that according to the baseline scenario the average individual risks will remain almost constant in the next 15 years. Note that that all estimates of the 75-lifetime risk do not exceed the risk management threshold of 1 in a million lifetime risk.

Year	Population risk (expected number of ground fatalities)	Average individual risk	
		Annual	75-year lifetime
2000	3.5	$1.3\ 10^{-8}$	$9.5\ 10^{-7}$
2005	3.8	$1.3\ 10^{-8}$	$9.9\ 10^{-7}$
2010	4.0	$1.3\ 10^{-8}$	$1.0\ 10^{-6}$
2015	4.3	$1.4\ 10^{-8}$	$1.0\ 10^{-6}$

Table 6.1 The Population and individual risk to U.S. residents due to crashing aircraft in 2000-2015

This data-driven analysis estimates the expected number of ground fatalities related to any of the Top100 airports in 2000 as 1 fatality in total (on average 0.01 per airport), while Schiphol Airport itself accounts for 0.6 ground fatalities annually according to a scenario-based model built by RAND in 1993. The difference can probably be explained by the fact that pilots try to avoid populated areas just before they crash and the scenario-based models do not take this into account. Consequently, one should be careful in classifying the risk of ground fatalities around an airport as unacceptable by just applying RAND’s model. The data-driven approach in this thesis would estimate the expected number of ground fatalities around Schiphol Airport as 0.014 per year, a factor 40 lower than RAND’s estimate.

Although the group risk in 2010 is estimated as 4 fatalities, this does not mean that the number of ground fatalities in 2010 will be exactly 4. The occurrence of ground fatalities due to aviation accidents is a stochastic process. Although general aviation crashes tend to have a smaller impact than air carrier crashes, general aviation accidents account for the majority of the expected ground fatalities (57%, 2.0 of 3.5 in 2000).

Besides the number of ground fatalities per year varies due to stochastic fluctuations, the risk itself is an uncertain variable. Furthermore, the risk differs among individuals and consequently is variable. In this analysis, uncertainty and variability have been separately analyzed.

6.1.2 Spatial variability

This analysis determined the spatial variability of the risk in two dimensions, distance to the airport and busy versus non-busy airports. Figure 5.1 presents the variability in both dimensions.
The following can be concluded:

1. Risks of ground fatalities are significantly higher in the vicinity of an airport. The spatial variability of the exposure associated with the dimension distance to an airport is approximately a factor of 100. The variability of the exposure to the ground fatality risk mainly applies to the first 2 miles around an airport.

2. Risks of ground fatalities associated with busier airports are higher than risks associated with less busy airports. Differences in exposure between busy and less busy airports mainly applies to the first 2 miles around the airport. The estimate of the average annual exposure within 0.2 miles of a Top100 airport just exceeds 10^{-6}.

This analysis only quantified the spatial variability in the two mentioned dimensions, but other dimensions account for spatial variability as well. For example people who live at the extension of a runway are at higher risk of death due to a crashing airplane. Evaluations of the risks at airports in Europe (Schiphol Airport, the Netherlands and Farnborough Aerodrome[1], United Kingdom) show that annual individual risks of 10^{-4} may exist at the extension of the runway. When these evaluations are correct, the runway dimension accounts for an additional factor of 100 in the variability of the risk. In addition to risks of ground fatalities differing among airports due to differences in airport activity, risks also differ because of the local variables like safety procedures, local climate, land use around the airport, etc.

The variability of the annual exposure to the ground fatality risk among the population associated with the dimension of distance to an airport is smaller than shown by Figure 5.1. The quantity plotted in this figure represents the annual exposure to the risk at a certain distance of an airport given that you stayed there for a whole year. However, people do not stay at the same place for a whole year, people go out, go on vacation or more regularly they go to work or to school. Their distance to the nearest airport changes almost continuously and consequently their exposure to the ground fatality risk does too. However, it is expected that people spend most of their time in their residence and consequently people who live in the vicinity of an airport are exposed to significantly higher risks. A more detailed analysis with respect to the mobility of Americans is needed to really answer this question. Only a brief evaluation concerning this issue is described in the following paragraph.

According to data of the Dutch Bureau of Statistics[2], Dutch residents older than 12 spend on average 16 hours a day at home[3]. Assume that the average American also spends 16 hours a day in her or his

---

[1] National Air Traffic Services, *Third Party risk at Farnborough Aerodrome*, July 1998 , United Kingdom.

[2] Centraal Bureau voor de Statistiek

[3] Schmeets, H., Wieling, M., *De leefsituatie van de Nederlandse bevolking 1997*, Centraal Bureau voor de Statistiek, 1999.

residence including the whole night. A review of the local accident times showed that it is very rare that aviation accidents occur between midnight and 6:30 am. Groundling accidents seem to occur uniformly over the period 6:30 am- midnight and consequently, it can be concluded that the exposure to this risk is concentrated in this day interval. This means that 9.5 of the 17.5 hours of exposure are spend at home, about 54% and the variability is reduced by about a factor 2. Consequently, the spatial variability of the risk among the population associated with where Americans spend their time relative to airports will be approximately a factor 50. On the other hand, there are Americans, like new-born babies and their mothers who live in the vicinity of an airport and who spend almost all their time in their residence and who may be exposed to an annual risk of 10^{-6}.

6.1.3 Uncertainty

An uncertainty analysis was performed (Section 4.2.3) to better interpret the results. The following variables were considered uncertain:

- Estimates of the U.S. resident population in 2000, 2005, 2010 and 2015.
- The population distribution function $F_{X,L,1990}(d)$ (see formula (4.1)).

The 90%-uncertainty intervals of the average risk for the different distance intervals are shown in Figure 5.2. Note that the second uncertainty excludes the uncertainty associated with the fact that $F_{X,L,1990}(d)$ is used to approximate $F_{X,L,2000}(d)$. This analysis assumes constant ground fatality rates per airport operation among all U.S. airports and uncertainty associated with this assumption is not included in the uncertainty analysis. Furthermore, $F_{X,L,2000}(d)$ is expected to be a stationary distribution.

Figure 5.2 demonstrates that although the uncertainty is not negligible, the variability is a more important factor for the risk of ground fatalities than uncertainty. Although not all sources of uncertainty have been quantified, this conclusion seems legitimate.

6.2 The risk of ground fatalities due to a crashing airplane; a useful risk communication tool?

In 1992, Goldstein et al estimated the 70-year lifetime risk to groundlings due to a crashing airplane as 4.2 in a million and suggested to use it as a risk communication tool for risks that fit the four criteria discussed in Section 1.4.

Goldstein et al used data from 1975-1985 to estimate the risk and noted that their estimate of the lifetime risk exceeds 1 in a million, a commonly used threshold for these types of risks in many U.S. regulatory approaches. However, this analysis updated the estimate of the 75-year lifetime risk of ground fatalities due to aviation accidents as 0.95 in a million, just below the threshold.

Goldstein et al noted that "*our society seems to have achieved a consensus that governmental action to protect public health is appropriate in environmental matters in the range of* 10^{-5} *to* 10^{-6} *lifetime risk*". The current lifetime risk of ground fatalities due to aviation accidents is just outside the range raising the question of whether this risk is still an appropriate communication tool.

Apparently the public does not seem to care about lifetime risks that are below 1 in a million and from that perspective the risk of ground fatalities due to aviation accidents seems useless for communication purposes in the context Goldstein et al suggested. On the other hand, the risk is just below the 1 in a million and can be used to represent the lower bound of the range.

Another issue that affects applicability of the risk to groundlings as a risk communication tool is the spatial variability of the risk. This analysis showed that people who spend most of their time in the vicinity of an airport are exposed to significantly higher risks of death due to crashing airplanes than those who spend their time further from airports. As argued in Section 6.1, the spatial variability of the risk causes significant variability of the ground fatality risk among the population, despite mobility. Recall that the spatial variability is only quantified in the dimension distance to an airport and that variability associated with the path of the runway is ignored. Goldstein et al ignored the variability and presented the risk of ground fatalities due to crashing airplanes as completely random.

Due to the spatial variability of the risk, both individuals who are exposed to a lower lifetime risk than 10^{-6} and to a higher lifetime risk than 10^{-5} exist in the United States. An evaluation of the risk of ground fatalities around Schiphol airport showed that annual individual risks of 10^{-4} are possible. The majority of the U.S. residents, people who spend most of their time further than 2 miles from an airport, probably have an individual lifetime risk of dying due to a crashing airplane below 10^{-6}. The variability reduces the power of the risk of ground fatalities due to aviation accidents as a risk communication tool, because the criteria Goldstein et al (Section 1.4) formulated are only met by selected group of people, not the whole U.S. population.

The first criterion[4] does not apply to those who spend most of their time in the vicinity of a busy airport (and especially not to people who live in the path of a busy runway), because the risk for these people is not so low that most of them are not concerned by the risk when they go about their daily lives. The awareness of the risk is raised by the noise of airplanes passing over their houses and the media attention that all large airplane crashes get. The fact that the European Transportation Safety Council noted that public tolerance versus third party risks at airports is decreasing supports the conclusion that people who live in the vicinity of airports are concerned about the risk.

Goldstein et al said that adjusting your lifestyle to reduce the risk of dying due to a crashing airplane seems ludicrous considering the magnitude of the risk (criterion 2, uncontrollable risk[5]). However, the variability of the risk puts this argument in a different perspective. U.S. residents are free to decide where to live, i.e. where to spend most of their time, and nearly everybody knows that the risks of dying due to an aviation accident are higher in the vicinity of an airport. These risks cannot be directly defined as uncontrollable or involuntary, because these people chose to live there. Furthermore, people who spend most of their time near an airport tend to benefit (criterion 3)[6] more from the airport because they are more likely to work at the airport or profit from it indirectly.

The importance of variability and uncertainty in risk management and regulation has been mentioned several times, but it does not make life easy. At first, the risk of ground fatalities due to a crashing airplane was presented as a “random” risk (i.e. non-variable) and it was used for risk communication purposes several times (and actually is still used) in that context. A more close evaluation of the risk

---

[4] Criterion 1: The risk is low- so low that we generally do not consider ourselves with risk at this level during our daily lives.

[5] Criterion 2: Though we are not in control of the event that leads to the risk, we can reduce the exposure by adjusting our lifestyles. However, the magnitude of the risk does not warrant the change in lifestyle and it seems ridiculous to do so.

[6] Criterion 3: We do not derive direct benefit of the activity that is responsible for the risk.

showed that the risk is highly variable and that the criteria formulated for the comparison do not necessary apply to most U.S. residents. Although the risk looses most of its strength as a communication tool for non-variable risks, it could be used to communicate with people about variability and uncertainty in risk. In fact, the ground fatality risk could be used as a communication tool for other variable risks, like risks associated with chemical factories, nuclear power plants or firework depots.

References

I Literature

1. Boeing, *Statistical Summary of Commercial Jet Airplane Accidents, Worldwide Operations 1959-1999*, 1999.
2. Brady, S.D, Hillestad, R, *Modeling the External Risks of Airports for Policy Analysis,* RAND, 1995.
3. Cooke, R.M., *Experts in Uncertainty*, 1991.
4. European Safety Transportation Council, *Safety In and Around Airports*, Brussels, 1999.
5. Federal Aviation Administration, *Terminal Area Forecasts – Fiscal years 1998-2015,* U.S. Department of Transportation, FAA-APO-98-10, October 1998.
6. Federal Aviation Administration, *FAA Statistical Handbook for Aviation (calendar year 1996),* Statistics and Forecast Branch of the Office of Aviation Policy and Plans, 1998.
7. Federal Aviation Administration, *The Global Analysis and Information Network*, Office of System Safety, June 1997.
8. Goldstein, B.D., Demak, M., Northridge, M., Wartenberg, D., *Risk to Groundlings of Death Due to Airplane Accidents: A Risk Communication Tool*, Risk Analysis, Vol 12, No. 3, 1992.
9. Harvard Center for Risk Analysis, *Cellular phones and driving, weighting the risks and benefits*, Risk in Perspective, July 2000.
10. Hillestad, R., Solomon, K., Chow, B. et al, *Airport Growth and Safety, A study of the external risk of schiphol airport and possible safety-enhancement measures*, RAND, 1993.
11. Hillestad, R., Solomon, K., Chow, B., et al, *Airport Growth and Safety,Executive Summary of the Schiphol Project*, RAND, 1993.
12. Hollmann, F.W., Mulder, T.J. and Kallan, J.E., *Methodology and Assumptions for the Population Projections of the United States: 1999 to 2100*", January 2000.
13. Morgan, M.G., Henrion, M., *Uncertainty, a Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*, 1990.
14. National Transportation Safety Board, *Annual review of accident data – U.S. General Aviation – 1996*, ARG-99/01, 1999.
15. National Transportation Safety Board, *We are all safer now*, 1998.
16. Piers, M.A. et al, *Study of the external risk around Schiphol airport*, National Aerospace Laboratory (The Netherlands), 1993.
17. RAND, *Managing Safety In and Around Airports*, Conference Proceedings, 1995.
18. Smith, E., Spouge, J. *Risk Analysis of Aircraft Impact at Schiphol Airpor*", Technica Consulting, May 1990.
19. Wilson, R., *Regulating Environmental Hazards*, Regulation, Vol 23, 2000.

II Internet sites

I. http://aviation-safety.net/
II. http://www.apo.data.faa.gov/faatafall.HTM
III. http://www.census.gov/plue/
IV. http://www.census.gov/plue/geocorr/data/mable-ascii/
V. http://www.census.gov/population/estimates/nation/popclockest.txt
VI. http://www.faa.gov/arp/A&D-stat.htm
VII. http://www.gcr1.com/aims/5010_fed.htm
VIII. http://www.landings.com/_landings/pages/search_apt-pos.html
IX. http://maps.yahoo.com/py/maps.py
X. http://mathworld.wolfram.com/GreatCircle.html
XI. http://nasdac.faa.gov/asp/fw_ntsb.asp
XII. http://www.ntsb.gov/aviation/classify/classification.htm
XIII. http://www.ntsb.gov/aviation/Table10.htm
XIV. http://tiger.census.gov/