AIRCRAFT RECOGNITION WITH RADAR RANGE PROFILES

RENÉ VAN DER HEIDEN

Aircraft Recognition with Radar Range Profiles

ACADEMISCH PROEFSCHRIFT

Ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam,
op gezag van de Rector Magnificus
Prof. dr J.J.M. Franse
ten overstaan van een door het college voor promoties ingestelde
commissie in het openbaar te verdedigen in de Aula der Universiteit
op dinsdag 27 oktober 1998 te 15.00 uur.

door
RENÉ VAN DER HEIDEN

geboren te Rozenburg

PROMOTIECOMMISSIE:

Promotor:	Prof. dr ir F.C.A. Groen
Faculteit:	Wiskunde, Informatica, Natuurkunde en Sterrenkunde Kruislaan 403 1098 SJ Amsterdam The Netherlands
Overige leden:	Dr R.G. Aldous Prof. dr ir F. Absil Prof. dr M. Boasson Prof. dr L.O. Hertzberger Prof. dr C.A.J. Klaassen Dr ir R.P.W. Duin

This research was performed at TNO Physics and Electronics Laboratory, The Hague, The Netherlands and funded by the Ministry of Defence of The Netherlands.

Aircraft Recognition with Radar Range Profiles / René van der Heiden. - Amsterdam: Universiteit van Amsterdam, Faculteit WINS. Proefschrift Universiteit van Amsterdam - Met lit. opg., Trefw.: radar / patroonherkenning / EM-modellering

Cover illustration "Circel" painted by Kees Buurman (ink on paper, 1980).
Printed by Print Partners Ipskamp, Enschede.

"Real knowledge is to know the extent of one's ignorance"
— Confucius

Voor Hannelore

Contents

Chapter 1

Introduction

1.1 Motivation and objectives

[1]In both the first and the second World War, aircraft recognition did not pose serious problems: the distance at which an aircraft could visually be recognised, was considerably larger than the reach of the weapon systems aboard the aircraft. The object was easily recognised by signs showing the nation of origin and by its squadron emblems. Also, the shape of the aircraft was usually distinctive for its origin. Finally, recognition on each of the characteristics was facilitated by a low velocity.

Nowadays, all fighter aircraft have a comparable shape, they may fly at supersonic velocities and the reach of their weapon systems has increased dramatically. The consequence is that the aircraft is much too close once it has been recognised on sight.

Just before the beginning of World War II the first systems were being developed to examine the identity of a flying target. Improved versions of the most important technique, Identification Friend Foe (IFF), are still being used.

IFF is based on a simple principle: an unidentified aircraft receives from an *interrogator* (for example a ground-based radar station) a signal, referred to as the *challenge*. If the target is friendly and equipped with a compatible transponder, it automatically interprets the challenge and generates an answer. This answer contains an identification code. It is automatically transmitted back and decoded when it is received at the ground station. After successful completion of this question and answer loop, it is concluded that the aircraft is friendly.

If the target is hostile, it is assumed that the transponder, if any, cannot interpret the challenge and the aircraft is therefore unable to return a correct answer.

IFF systems have a number of limitations. Many IFF systems still in use are old and either have no encryption or have an encryption scheme which may no longer be completely effective. As a result, these systems can be deceived. Even with encryption it may still be possible for a hostile aircraft to listen to and re-transmit a reply from a friendly aircraft, in which case the hostile aircraft will appear to be friendly

[1]Most of this section and the next chapter were published in Dutch in [32].

unless the code validity interval is extremely short.

Another difficulty is that civil airliners are fitted with transponders that conform to open international standards. Thus a hostile military aircraft can be equipped with such a transponder to make it appear to be a civil aircraft. Moreover, the interrogation procedure can be unsuccessful even if the aircraft really is friendly. For example, the transponder aboard the aircraft can be overloaded with IFF challenges, the crypto key may have been entered incorrectly, or the transponder may simply be defective. Failure may also be caused by hostile jamming. Finally, there may be severe practical problems ensuring full interoperability between all pairs of interrogators and transponders, particularly in multi-national military operations.

Now we have arrived at a crucial point. If the IFF interrogation loop is interrupted for any reason, the inference drawn from the absence of a correct reply, namely "this is a hostile aircraft" can lead to fratricide, the elimination of a friendly aircraft.

Extra sources of information may be employed to reduce the probability of an erroneous decision, such as intelligence and *Electronic Support Measures* (ESM). In the latter technique one "listens" to the radio and radar signals that originate from the aircraft. An additional option is to agree on corridors which may not be left by friendly aircraft. If an unidentified object moves out of such a corridor, it is probable that this is an unfriendly object.

Still, these methods and procedures are not sufficient to face the problems with IFF. In his book on radar recognition [57] Nebabin lists several incidents of erroneous aircraft identification with casualties — notorious examples from the last twenty years:

1980 An Italian DC-9 passenger plane was shot down with a missile near Sicily. The missile (supposedly French or US made) was reaimed by mistake because of incorrect IFF and discrimination of the type of plane. 81 passengers died.

1988 In the Persian Gulf, an Iranian Airbus passenger plane was downed by an SM-2 Standard Missile from the Aegis system of the cruiser USS Vincennes. The aircraft was erroneously recognised as an F-14 fighter. 298 people died.

1992 A helicopter of the European Community observers' mission in Yugoslavia was brought down 80 km north of Zagreb by a missile launched from a Yugoslavian aircraft. Five men died. Incorrect IFF and recognition of the type of aircraft were supposed to be the cause.

1994 Two US Blackhawk helicopters were shot down in the no-fly zone of Iraq by two US F-15 fighter aircraft. Fourteen men died. The helicopters were erroneously identified as Iraqi Hinds. Careful investigation of this incident showed that the Turkish IFF key was used, whereas it should have been switched to the Iraqi key at the moment the Blackhaws passed the Turkish-Iraqi border. Consequently, no answer was returned once the helicopters were in Iraq.

To reduce the probability of fratricide, for some time now research is performed into techniques that provide *positive foe recognition* or *Non-Cooperative Target Recognition* (NCTR). With these techniques, the aircraft is identified without its active participation. The basic idea is that the geometry of the aircraft or the moving parts on the aircraft impose features in the reflected radar signal that are typical for the object. These features may then be used to classify the target.

The radar is a very attractive instrument for NCTR. It is able to detect and track aircraft at large distances (up to several hundreds of kilometres) in adverse weather conditions and at both day and night. Other sensors (i.e. optical, infra-red, laser) do not have the same reach and/or have comparable all-weather capability. Additionally, the last decades have also shown essential technological steps in radar hardware that make the measurement of signatures for target recognition possible.

Several types of radar signatures can be employed to acquire information about the aircraft characteristics. For example, one can exploit the modulation in the radar returns by the rotating parts on the aircraft, such as propellers, rotors, or the compressor blades on a jet aircraft. Another option is to make an actual radar image of the target such that information on the geometry of the aircraft is revealed. This can be done in one dimension, giving the distribution of aircraft

radar reflections in range. Such a signature is called a radar range profile. Using the motion of the aircraft, under certain circumstances also information in the direction perpendicular on the line-of-sight can be extracted to form a two-dimensional image.

To our knowledge, of all radar signature types, most research effort in the field of target classification with radar is currently devoted to radar range profiles. For example, in a recent symposium on NCTR with radar (Mannheim, Germany, April 22-24, 1998), the majority of the papers on techniques for aircraft recognition, ten out of twelve, used radar range profiles as the target signature. The reason for this is their operational value: they can be measured in a relatively short time, recognition is possible at all attitudes of the target and the demands on the radar hardware are not excessive. Moreover, with moderate effort many existing radars can be upgraded to measure range profiles. On these grounds we have chosen radar range profiles as the signature for aircraft recognition for the research presented in this thesis.

Before range profiles can be used for recognition, they first need to be collected with a radar. Several basic profile measurement options are available given the radar equipment. The data we have to our disposal come from radars that measures a profile by synthesising a sequence of snapshots that individually contain only a small amount of information on the target geometry. Integration of these pulses through digital signal processing techniques produces the range profile. The first objective of this thesis is to provide a comprehensive explanation on the principles of this technique and the signal processing involved. Considerable attention will be paid to the influence on the range profile of rotation and radial motion of the aircraft relative to the radar.

After profile measurement, we can proceed with its classification. It is the second goal of this thesis to show that range profiles do contain enough information on the aircraft geometry to allow recognition. We want to show which classifiers are available, and what the trade-offs are between the different properties of the classifiers. Also, we aim to demonstrate the advantageous effects of an elementary preprocessing step applied to the profiles.

Early on in this research we realised that a fundamental problem needs to be solved before range profile classification can be used in operational environments: it concerns the availability of sufficient data

for training a classifier. The problem is that we would have to carry out lengthy and very costly measurement campains to collect profiles from all aircraft of interest, seen at virtually all orientations. A way out, and the approach we shall assess in this thesis is to use computer simulations of radar range profiles. Our third objective is hence to show that it is in principle possible to recognise measured profiles with their simulated counterparts.

1.2 Our contributions

Range profiles can be measured with a radar using so-called *stepped frequency waveforms* that consist of a sequence of short emission intervals, i.e. pulses, where the radar frequency within each interval is different. We show that several options are available for the times at which these pulses can be emitted and choose the appropriate one for our purposes.

In a large number of experiments we compare four different range profile classification techniques and establish the trade-offs between classification accuracy and speed. For one particular combination of application and scenario, the classification accuracy may be of highest importance, whereas for the other a very quick answer is essential. Based on the trade-offs, a choice is made for eight combinations of scenario and application.

We will show that the preprocessing of radar range profiles is an extremely important step. A simple, computationally cheap, nonlinear scaling can improve the classification results dramatically. The application to other pattern recognition problems confirms that its usefulness extends far beyond radar data.

Then, we will show that computational electro-magnetics in conjunction with an accurate geometrical description of a target, can be used to reliably mimic radar range profiles. The comparison shows that, particularly when the aircraft is looked upon from broadside, the similarity between measured and computed range profiles is convincing.

The final contribution is the demonstration that such synthetic range profiles can actually be used for recognition of measured profiles.

1.3 Outline of this thesis

The next chapter reviews the prominent techniques for non-cooperative target recognition with radar, thereby introducing the radar range profile. Then, Chapter 3 gives the physical background of the range profile, reviews its variability as a function of target motion and radar characteristics, and shows which waveform types for a stepped frequency radar are available. The acquisition of range profiles is covered by Chapter 4. We show how the profiles have been measured and present the program for computational electromagnetics along with the aircraft models that are available. Four distinct classification techniques are described, tested and compared in the subsequent chapter. It is demonstrated which trade-offs can be made — application is on four different radar systems in two scenarios. Chapter 7 is devoted to a comparison between measured and simulated range profiles and provides a careful assessment of the causes of the differences. In Chapter 8, measured range profiles are classified with simulations. In the final chapter, we give the conclusions and provide directions for future research.

Chapter 2

Non-cooperative target recognition techniques

2.1 Introduction

This introductory chapter delineates the developments in the most important non-cooperative target recognition techniques, thereby providing a first, qualitative introduction to aircraft classification with radar range profiles.

The prominent methods may be subdivided into two families.

- The techniques from the first group are based on the fact that radar radiation is reflected by the rotating parts of the engine, such as the propellers on a prop-aircraft, the rotors on a helicopter or the compressor blades in a jet engine. The latter technique is quite successful — it will therefore be reviewed more extensively in the next section.

- Techniques from the second family classify an aircraft on the basis of its radar image. One-dimensional images are High Resolution Radar (HRR) range profiles, or simply *range profiles*. With the Two-Dimensional Inverse Synthetic Aperture Radar (2D-ISAR) technique radar images of aircraft in two dimensions can be produced. Sections 2.3 and 2.4 are devoted to a discussion of NCTR with range profiles and 2D-ISAR, respectively.

Section 2.5 presents a survey of several other approaches to target classification.

Before continuing let us first make two remarks on the terminology used in this thesis. Firstly, both the term *object* and *target* are used to indicate an aircraft in flight. Secondly, *classification* and *recognition* denote the same action: determination of the type of class of aircraft, e.g. "this is an Airbus 310". *Identification* denotes the assignment of the aircraft to three possible classes: friend, foe or neutral.

2.2 Jet Engine Modulation (JEM)

Using a radar, one can look into the engine of a jet aircraft if the aircraft is flying towards or away from the radar. The largest fraction of this radiation is reflected by the blades of the first rotor. A smaller portion

passes along the first rotor and is reflected by the blades on the second rotor. Theoretically, reflections from further blades are also included in the radar return, but in practice reliable attribution of spectrum features to third and subsequent rotor stages is seldom possible. Figure 2.1 shows a characteristic spectrum.

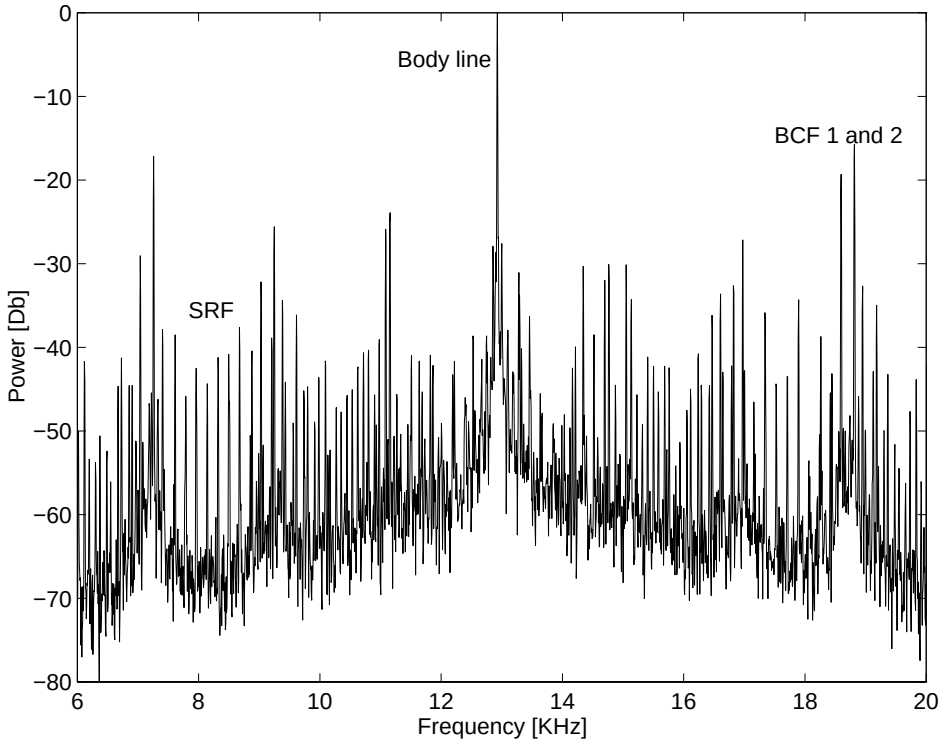

Figure 2.1: *JEM spectrum of a two-engine aircraft.*

The central peak, the *body line*, shows the reflection of the aircraft as a whole. The term BCF denotes the *Blade Chopping Frequency*: the frequency corresponding to the rotation of the first stage rotor over a single blade interval i.e. $\frac{360}{N_B}$ degrees where N_B is the number of blades. As two BCF lines can be seen, it follows that the aircraft has at least two engines. The somewhat lower peaks in the spectrum, clearly present under the phrase "SRF" are harmonics of the so-called *Shaft Rotation Frequency*. This frequency corresponds to a 360° rotation of a blade. Division of the BCF by the SRF gives the number of blades. This

information can be looked up from a table and classification follows. In the event that two engines have the same number of blades on the first rotor stage it is necessary to use features such as the engine Shaft Rotation Frequency and the second rotor stage returns to resolve the ambiguities.

The same technique can be applied if the radar looks in the rear side of the engine onto the turbine blades.

Looking at Figure 2.1 it seems quite complex to extract the desired parameters from such a signature. However, quite straightforward signal processing techniques already provide fairly good results [75].

Strong (+) and weak (–) points of JEM are:

\+ The target data base is small and simple

It contains a table with, for each engine type, information about the number of blades on the successive compressor blades (front of engine) and turbine blades (back of engine), and the range of Shaft Rotation Frequencies.

\+ A relatively short *time on target* is required

\+ Almost each aircraft has a unique engine type.

– The success of a JEM-classification depends on the aspect angle.

The radar signal has to reach the interior of the engine. Only close to head-on or tail-on, a useful spectrum can be obtained.

– A large signal-to-noise ratio is required for a successful classification.

It means that JEM is only suitable for classification at relatively short distances.

– The JEM spectrum cannot always be interpreted reliably in cases such as:

- aircraft with 3 or more engines where the engine Shaft Rotation Frequencies are not accurately synchronised and
- engine types where the first and second rotor stages are on different engine shafts which rotate at different rates.

JEM is a mature technology, applied in many operational situations such as the US F18 aircraft[1] and the US HAWK missile systems[2]. A thorough article on the modelling of JEM spectra has been written by Bell and Grubbs [9].

2.3 Radar Range Profiles

High-Resolution Radar (HRR) range profiles or simply "range profiles" are essentially one-dimensional images of aircraft. The scatterers, i.e. the parts on the aircraft that give a strong radar reflection, are projected onto the line of sight. See Figure 2.2

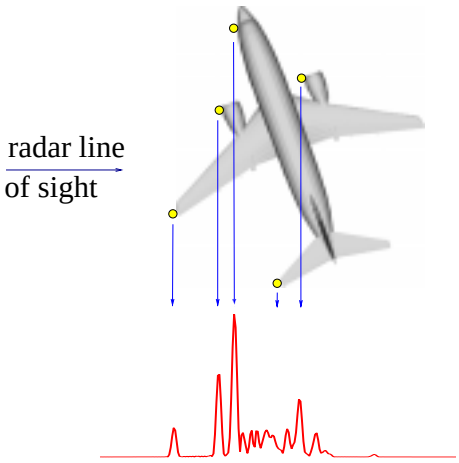

Figure 2.2: *A range profile of an aircraft viewed from the left hand side. Responses from the aircraft scatterers (dots) are projected onto the line of sight, resulting in a radar range profile.*

The profiles contain information on the *geometry* of the target and may therefore be characteristic for the type of aircraft. On the basis of previously measured profiles, a well-designed classification system is able to recognise an aircraft.

[1]Source: Department of Defense News Release, September 11, 1985, vol. 85, no. 543, p.3.

[2]Source: PR Newswire, March 12, 1990, p. 1

The advantages (+) and drawbacks (–) of aircraft recognition with range profiles are:

+ Classification of range profiles is possible at any aspect angle.

+ A relatively short time on target is required.

  Typically it takes several tenths of a second to measure a range profile using so called *stepped frequency waveforms*.

– The shape of a range profile depends strongly on aspect angle.

  It means that a large data set is necessary of all expected targets with profiles measured on a very dense aspect angle grid.

Of all NCTR techniques, the classification of range profiles currently draws most attention. Literature on this subject is steadily growing, an overview can be found in Chapter 4.

2.4 Two-Dimensional Inverse Synthetic Aperture Radar (2D-ISAR)

Range profiles offer resolution in the line-of-sight direction. In Two Dimensional Inverse Synthetic Aperture Radar (2D-ISAR) the technique is extended through the use of information in the cross-range direction. Employing the rotational component of the aircraft's motion perpendicular to the line of sight makes it possible to separate radar scatterers also in this direction [77]. Figure 2.3 shows an example of an Airbus A310.

The strong (+) and weak points (–) of target recognition using 2D-ISAR images are:

+ A 2D-ISAR image gives more detailed information on the target geometry than a range profile.

+ 2D-ISAR images are particulary suited for human interpretation.

– The target motion should have a rotational component perpendicular to the line of sight.

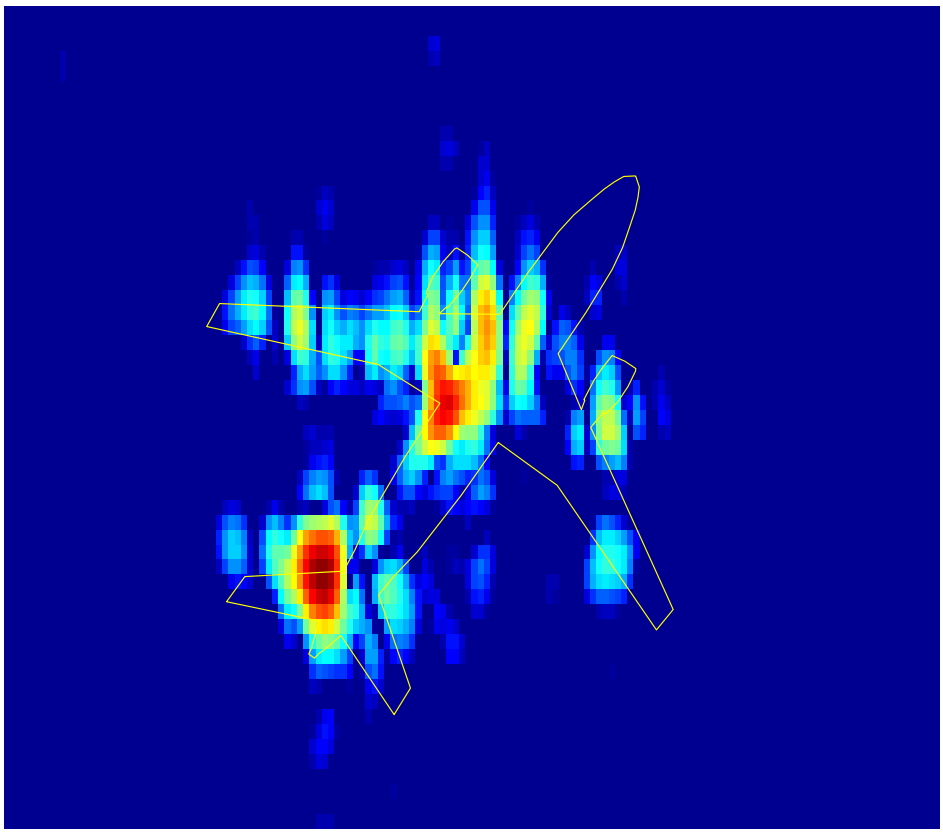

Figure 2.3: *Example of a 2D-ISAR image of an Airbus 310.*

– The motion compensation and the processing from raw radar data to focussed 2D-ISAR images is complex, not robust and computationally very expensive.

– A considerable time on target is needed.

– Knowledge about the precise projection plane (plane perpendicular to the rotation vector) is often imprecise which makes recognition complicated.

The most important problem in using 2D-ISAR images is the motion compensation [4, 72]. Normal tracking data are often not accurate enough for producing a well-focused image. Therefore the compensation has to be carried out using *autofocus techniques* where the motion parameters are estimated using the radar data itself [31, 73].

Only few publications address the problem of 2D-ISAR classification. Novak [58] uses a statistical pattern classifier on turntable measurements of four different objects. Bachmann *et al* [5] demonstrate how 2D-ISAR-like objects can be classified using three different neural network architectures. Ioannidis [42] shows classification results based on moment invariants and edge-detected images. Intriguing new developments have been reported by Rihaczek and Hershkowitz in the field of 2D-ISAR classification through the use of the so-called *complex-image analysis* [65].

Intuitively, an image contains much more information than a single range profile and should therefore lead to lower classification errors. However, none of the authors discusses whether the effort that has to be invested in the measurement and focussing of 2D-ISAR images pays off in an improved classification rate.

2.5 Survey of other techniques

Similar to JEM, a radar may also be used to identify propeller aircraft (*Propeller ROtor Modulation*, PROM) or helicopters (*HElicopter Rotor Modulation*, HERM) [62, 63]. In the latter technique the main and tail rotor cause characteristic Doppler spectra from which several parameters may be extracted, such as the number of rotors, the number of

blades on the rotors and the rotation frequencies. These numbers may be found in a table, and classification follows.

Natural resonances occur if a radar signal with a long wavelength is reflected by an aircraft. An introduction can be found by Morgan [54]. He makes an appealing comparison with a guitar string. The initial phase and the amplitude depend on the nature of exitation. However, the frequency and the damping of the string is characteristic for the string itself and could serve for the classification of the *type* of string. In the same way the resonant frequencies of an aircraft could be measured with a radar signal that has a wavelength in the order of the size of the aircraft.

A major advantage is that the method is practically aspect angle independent. It means that only a small data set of signatures is necessary. It could also be used for *stealth* aircraft. The method is, however, rather impractical as for realistic targets the radar has to emit a signal with a very large wavelength combined with a large bandwidth. This requires large radar antennas and a considerable number of frequency channels. Nevertheless, a relatively large amount of publications has appeared on this subject [1, 15, 16].

Several other physical effects, such as the *skin effect* and the *nonlinear effect* may theoretically be used for target recognition. A complete listing is provided by Nebabin [57].

2.6 Comparison of techniques for NCTR

Table 2.1 summarises the advantages and disadvantages of the three most important NCTR techniques.

This chapter has provided an overview of the available NCTR techniques. Several techniques are mature (JEM, HERM) but are aspect angle dependent (JEM) or work only for a small class of targets (HERM, PROM). Other methods are difficult to use in practical situations (Natural Resonances) or do not yet have the required robustness (2D-ISAR). For these reasons the classification on radar range profiles (HRR) currently draws most attention and is likely to have the largest potential for improved air target recognition.

	Advantages	Disadvantages
JEM	Short time on target	Aspect angle restrictions
	Simple target data base	Not suitable for large distances
	Fast classification	
	Mature technology	
HRR	Relatively short time on target	Large data set of signatures
	Applicable on all aspect angles	
2D-ISAR	Detailed object information	Depends on target motion
		Complex motion compensation
		Long time on target
		Large data set of signatures

Table 2.1: *Comparison of classification techniques.*

The most important topics that need to be addressed to move towards an operational range profile classification system are the measurement of range profiles, their nature, the classification and the fundamental lack of measured range profiles to train a classifier in operational situations. In this thesis all three problems are treated.

Chapter 3

Radar range profiles

3.1 Introduction

In this chapter we provide the physical background of radar range profiles.

Before range profiles of aircraft can be used for classification or any other task, they have to be measured with a high-resolution radar. The third and fourth section of this chapter give the basis for these measurements and show which options are present for the waveform choice. We evaluate the trade-offs to find an appropriate waveform for the radar equipment we have available for range profile measurement, the FELSTAR. In a measurement campaign, called ORFEO, we have acquired range profiles of civil aircraft with this waveform. In Section 3.6 the necessary background on the variability of range profiles due to various parameters is presented. This is an important subject, as the nature of the range profile fluctuations have a decisive influence on the design of the classifier. We will start now with several definitions to be used throughout this thesis.

3.2 Definitions

The *target aspect angle* can be expressed as a coordinate pair (α, θ) where α is the *aspect azimuth* and θ is the *aspect elevation*. See Figure 3.1. We define the aspect elevation θ as the angle between the radar line of sight and the plane through the wingtips and nose of the aircraft. The elevation is positive if the aircraft is viewed from below. Note that if the aircraft flies with zero pitch and roll angle, the aspect elevation equals the *radar* elevation (angle between horizontal plane through the radar and the aircraft direction).

We define the aspect azimuth α as the angle between

- the direction of the nose of the aircraft and
- the direction of the radar line of sight projected on the plane through nose and wingtips.

The aspect azimuth is positive if the aircraft is viewed from the starboard side, it is negative if it is viewed from port side, see Figure 3.2.

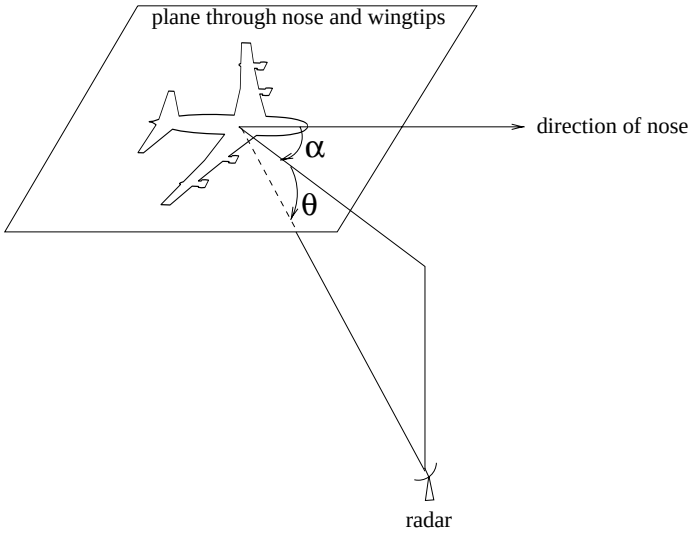

Figure 3.1: *Definition of the aspect elevation and aspect azimuth. In this particular orientation both α and θ are positive.*

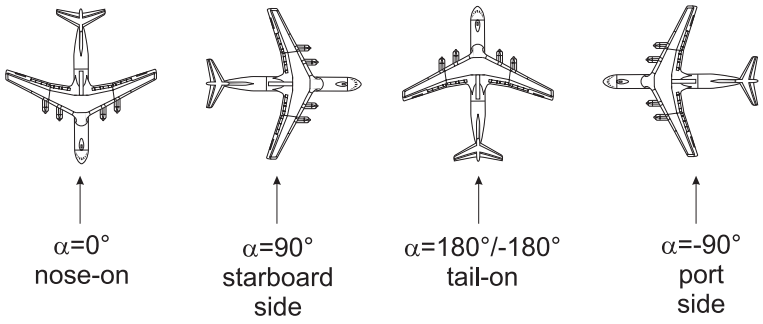

Figure 3.2: *Aspect azimuth.*

A major source for range profile fluctuations are changes in aspect angle of the aircraft during flight. Rotation about an arbitrary axis can be decomposed into roll-, yaw- and pitch motions, illustrated by Figure 3.3.

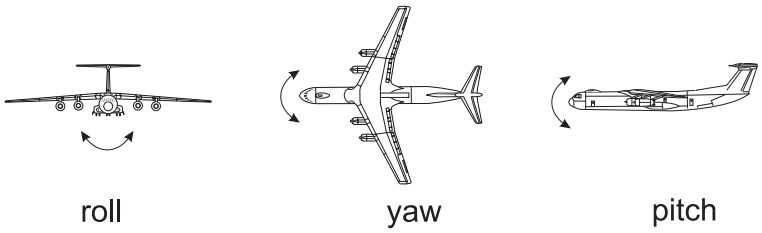

Figure 3.3: *The three basic rotational motions.*

3.3 Physical background

A range profile can be viewed as a one dimensional "image" of an aircraft, where the parts of the aircraft that reflect the radar radiation (*scatterers*) are projected onto the line of sight. An example is Figure 2.2 on page 17.

Wehner [78] provides in his book on high-resolution radar the basis for the production of range profiles: a range profile can be measured by recording, as a function of time, the target response on a very short pulse. For example, again view Figure 2.2, page 17. Imagine a short pulse being transmitted from the left hand side. The pulse hits the tip of the left-most wing first, then the nose and subsequently the horizontal stabiliser on the tail. In the same order, the echoes arrive back at the radar. Recording the backscattered energy as a function of time then produces the profile. To measure a profile in this way, the pulse should be extremely short: for the one shown in the figure under consideration roughly 2 ns would be required. This puts severe constraints on the radar hardware.

Instead of monitoring the incoming energy as a function of time, we can also measure the change in amplitude and phase as a function of radar frequency. The advantage is that less stringent hardware requirements need to be met: a profile is produced by emitting a number of

pulses with linearly increasing frequencies, a *Stepped Frequency Waveform* (SFW). In such a way, the desired bandwidth is synthesised from pulses that have a narrow bandwidth individually but together span a large frequency range. As the real measurements we use in this thesis have exclusively been made using SFWs, we will limit ourselves to a description of range profile generation with these waveforms.

A sequence of $L+1$ pulses is emitted with linearly increasing frequencies $f_l = f_0 + l\Delta f$ at time instances t_l, where l runs from 0 to L. See also Figure 3.4. The signal z_l^e that is transmitted in pulse l can be

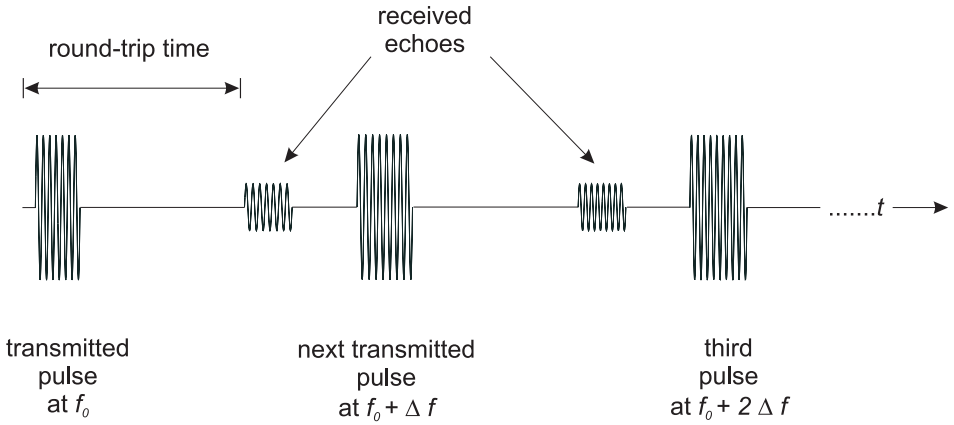

Figure 3.4: *First three pulses of a Stepped Frequency Waveform. Also shown are the echoes from the first two pulses.*

written as a complex exponential:

$$z_l^e = Ae^{2\pi j f_l (t+t_l) + 2\pi j \theta_{0,l}} \tag{3.1}$$

where $\theta_{0,l}$ is the initial phase for the l^{th} pulse and A is the amplitude. Suppose that the emitted energy bounces on a single, stationary point scatterer, then the received echo is given by

$$z_l^r = \rho A e^{2\pi j f_l (t+t_l - 2R/c) + 2\pi j \theta_{0,l}} \tag{3.2}$$

during the length of the pulse. Here ρ is the reflectivity of the point scatterer that is present at a distance R. Essentially, the radar measures the modulation of the original signal as it reflected on the scatterer:

$$G_l \equiv \frac{z_l^r}{z_l^e} = \rho e^{-4\pi j f_l R/c} = \rho' e^{-4\pi j l \Delta f R/c} \tag{3.3}$$

where the constant $e^{-4\pi j f_0 R/c}$ is absorbed in ρ'. For the full waveform l runs from 0 until and including L. As can be seen, the phase of the complex-valued sequence $G_l, l = 0, \ldots, L$, is linear in l and therefore contains exactly one harmonic component[1]. Now, observe that each additional scatterer i would give an extra term, i.e. harmonic component, in the sequence G_l, each with its own amplitude ρ_i and frequency depending on the range R_i to the scatterer. The appropriate technique to find the harmonic contents of this sequence is the Inverse Discrete Fourier Transform (IDFT).

For simplicity let us go back to one scatterer. Then we apply an IDFT to the sequence G_l; the resulting sequence we call g_m, defined for $m = 0, \ldots, L$. Taking the squared magnitudes $|g_m|^2$ we find (see Appendix A):

$$|g_m|^2 = \rho^2 \left| \frac{\sin \pi y_m}{\sin \frac{\pi y_m}{L+1}} \right|^2 \tag{3.4}$$

with

$$y_m \equiv m \Leftrightarrow \frac{2BR}{c}. \tag{3.5}$$

where $B \equiv L\Delta f$ is the bandwidth. Now the vector with elements $|g_m|^2, m = 0, \ldots, L$ is called a *high-resolution radar (HRR) range profile* or simply a *range profile*.

Let us derive two fundamental parameters from Equations 3.4 and 3.5: the amount of detail that can be seen in the range profile and the largest object that fits in the profile. Firstly, the distance in metres between two gridpoints ($\Delta m = 1$, i.e. the change in range ΔR for which $2B\Delta R/c = 1$), equals $c/2B$. This quantity is called the *nominal resolution* of the range profile

$$\Delta R_N = \frac{c}{2B}. \tag{3.6}$$

Secondly, scatterers at ranges $\ldots, R \Leftrightarrow 2\frac{c}{2\Delta f}, R \Leftrightarrow \frac{c}{2\Delta f}, R, R + \frac{c}{2\Delta f}, R + 2\frac{c}{2\Delta f}, \ldots$ will all appear in the same position in the range profile. The

[1] Later in this chapter we will look at moving targets: then the measured sequences are not necessarily linear in l and G_l contains a continuum of harmonics.

number $\frac{c}{2\Delta f}$ is therefore named the *unambiguous range interval*:

$$R_U = \frac{c}{2\Delta f}. \tag{3.7}$$

If a target is larger in range than R_U aliasing occurs: in the range profile reflections from the far side of the target are inseparably mixed with reflections from the near side of the target, illustrated by Figure 3.5.

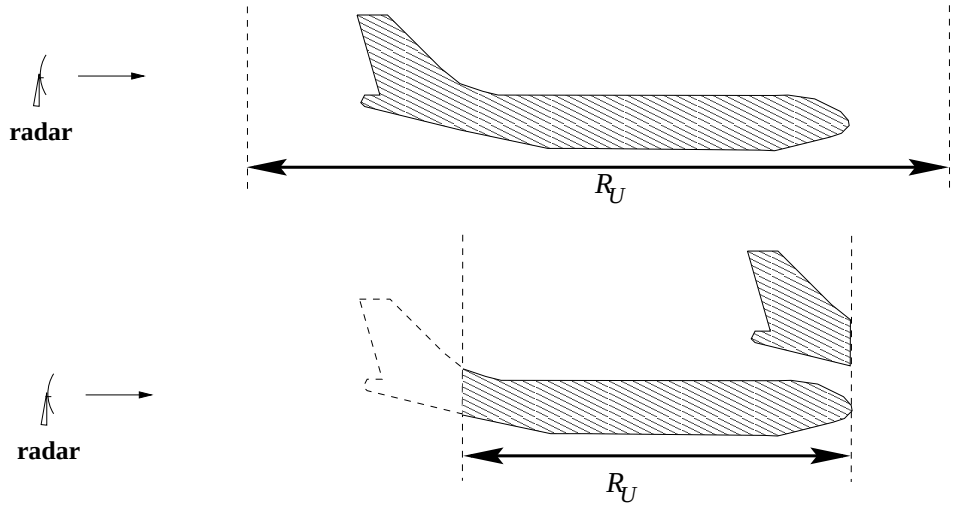

Figure 3.5: *In the top figure the unambiguous range R_U is larger than the target size and the aircraft fits into the window. The second figure illustrates the undesired effect of aliasing: the range window is smaller than the target dimension, and contributions from different parts of the aircraft are inseparably mixed.*

The sequence $|g_m|^2$ has a maximum for the integer value of m closest to $m_0 = \frac{2BR}{c}$ as in that case y_m is closest to zero. If it happens that $m = m_0$, then we find exactly $|g_m|^2 = \rho^2$. This is of course not the case in general — we shall therefore interpolate the profile to (partly) reveal contributions in between grid-points. This can be conveniently done by adding zeroes to the right of the sequence G_l before the IDFT is performed ("zero-padding").

The IDFT works on a selection of frequencies ranging from f_0 to $f_0 + B$. The edges of this choice are very sharp: effectively a square window is used. These abrupt changes give rise to high sidelobes next to the main peaks in the range profile which could easily obscure real scatterers of the aircraft. Therefore, we shall apply a weighting function to smooth the edges, at the cost of a worsened (that is, larger) resolution. Many windows are available, all with their own properties and trade-offs [30].

See Figure 3.6, solid line. In this thesis, and in the figure, we have used a Hamming window, giving a resolution of approximately $1.3\Delta R_N$ and a sidelobe reduction of 20 dB.

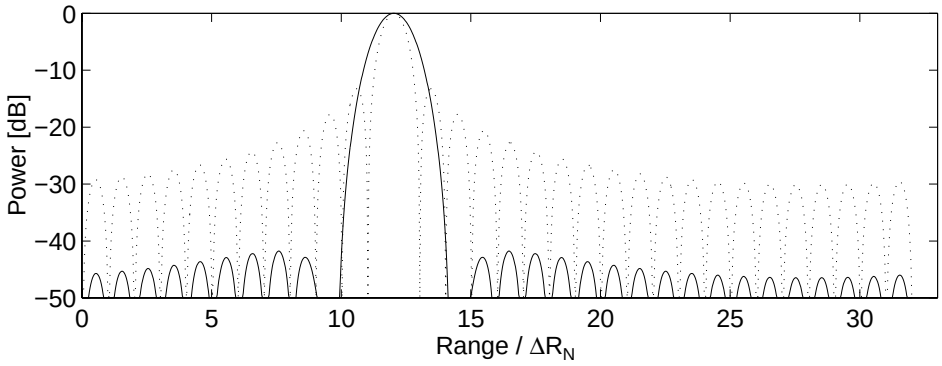

Figure 3.6: *Side-lobe levels for rectangular, default, window (dotted) and Hamming window (solid).*

3.4 Waveform parameters

At this point several trade-offs can be made for a stepped frequency radar system. We will not go into the different types of waveforms — this is left for the next section.

For classification purposes we want the profile to contain as much information as possible to make it maximally discriminating against profiles from other aircraft. This requires a high resolution and therefore it is desirable to use the maximum bandwidth available given the radar hardware.

The unambiguous window, R_U, must be chosen such that the size of the largest expected target does not exceed this window to avoid aliasing. If R_T is the overall size of the expected targets, then we must require that $R_U > R_T$.

It is advisable, however, to choose the unambiguous window even larger than the maximum expected target size as most available range profile classification techniques are based on *circular correlations* between range profiles, see [36, 41] and Chapter 3.

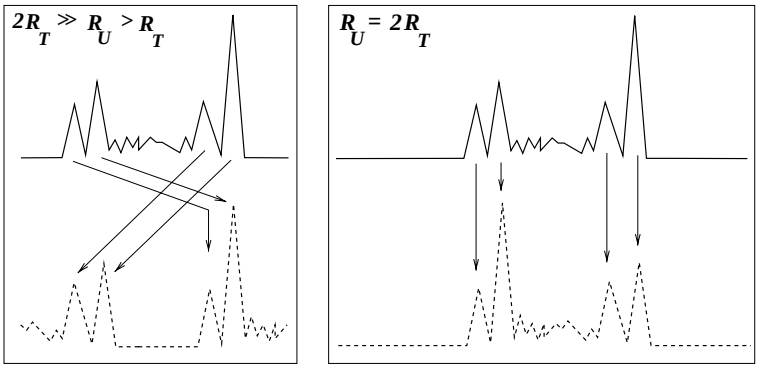

Figure 3.7: *In the left figure the unambiguous range R_U is larger than the target size but considerably smaller than $2R_T$. Then, a circular correlation mismatch is possible: the peaks originating from the far end of the aircraft correlate with the peaks from the near end of the aircraft. If R_U is $2R_T$, shown in the figure on the right hand side, such a miscorrelation is not possible.*

See Figure 3.7: incorrect correlation between peaks originating from the far side of the aircraft in the first profile and peaks originating from the near side of the aircraft in the second profile might occur. This may give rise to additional classification errors. If we want to be absolutely sure that this correlation mismatch does not occur, we have to leave one half of the range profile empty. An extra advantage of leaving a portion of the profile empty is an improved signal-to-noise ratio.

So we would like to choose

$$L \approx \frac{4R_T B}{c} \tag{3.8}$$

Note that this number of pulses is an upper limit. If a smaller portion of the profile is empty, say 25%, correlation mismatch is still very unlikely. Also, we use the maximum "optical" target dimension R_T which is usually larger than the maximum separation between any two scatterers that can be seen at the current aspect angle.

Another requirement to be considered is that we cannot choose L too large, because the larger L is the longer it takes to emit the entire bandwidth, and the more distortion is to be expected due to the movements of the target during the emission of the waveform.

If we consider civil aircraft we can expect a maximum size of $R_T =$ 70 m. The maximum bandwidth of the radar we have used for measurements, the FELSTAR, is 600 MHz and we thus find that $L + 1 = 421$. For a range profile measurement campaign, named ORFEO and held in 1995 with the FELSTAR radar, we decided to settle for a smaller number of pulses and a lower range resolution to improve the robustness for radial motions, see Table 3.1.

Carrier frequency	f_0	=	3073.8 MHz
Bandwidth	B	=	452.2 MHz
Number of pulses	$L+1$	=	324
Frequency step	Δf	=	1.4 MHz
Nominal Range Resolution	ΔR	=	0.33 m
Unambiguous Range	R_U	=	107.5 m
Minimum delay between two pulses	$\Delta t_{\min}$	=	420 μs

Table 3.1: *ORFEO waveform parameters.*

The minimum delay between two pulses is fixed by the round-trip time of the emitted pulse and a (hardware determined) *silent* or *dead* period before a new pulse can be emitted. As the latter period is $20\mu s$ for the FELSTAR and as we wanted to measure targets up to 60 km in range, we use $\Delta t_{\min} = \frac{2\times 60\times 10^3}{c} + 20\mu s = 420\mu s$. In the next section the choice for the waveform type will be further investigated.

3.5 Stepped Frequency Waveform types

An important problem in the production of range profiles is that the profiles are distorted by radial motions of the target. It is possible to compensate for these motions if the flight path is accurately known, but this information is often too coarse or not available. Furthermore, it is desirable to save the computation time needed for motion compensation and use it for the classification of the profiles. Finally, for accurate motion compensation the target has to be tracked for a certain time to estimate its velocity and acceleration which increases the time-on-target considerably.

In this chapter we will present three different ways to emit a stepped frequency waveform. We will start with the *Linear Waveform* (LW) in which the pulses are emitted at equidistant time instants (Subsection 3.5.2). Subsection 3.5.3 describes the *Velocity Tolerant Waveform* (VTW). During the emission of the frequencies the time step between the pulses decreases, such that any constant velocity of the aircraft is compensated automatically. In the *Acceleration Tolerant Waveform* (ATW) the time steps decrease at a faster rate compared to the VTW, and it gives an undamaged range profile independent of the acceleration of the aircraft (Section 3.5.4).

Of course, the LW and the ATW produce distorted range profiles if the aircraft has a non-zero radial velocity and the LW and the VTW produce distorted range profiles if the aircraft has a non-zero radial acceleration. We will, where possible, derive the magnitude of these effects as a function of the basic parameters involved in range profile generation, i.e. the carrier frequency f_0, the fractional bandwidth γ ($\equiv B/f_0$), the number of pulses $L+1$, the minimum delay between two pulses $\Delta t_{\min}$ and the target's velocity and acceleration. The equations will be applied to the ORFEO parameters given in the Table 3.1 on page 32. In the next section we will provide a general approach for deriving these quantities.

3.5.1 Distortion of a range profile due to motion

We have so far assumed that the distance to the target was a constant R. However, the object generally has a certain velocity and acceleration

and we shall thus write

$$r_l = R + vt_l + \frac{1}{2}at_l^2 \tag{3.9}$$

where v and a are the radial velocity and radial acceleration of the target at time $t_l = 0$, respectively. We omit higher terms in this approximation.

From Equation 3.3 we can see that the phase of a point scatterer is given by

$$\phi_l = 2f_l r_l / c. \tag{3.10}$$

where we have replaced R with r_l. For any choice of the emission time t_l of pulse l, we can write the phase of the received pulse for a point scatterer by substituting Equation 3.9 into 3.10:

$$\phi_l = \phi_l^0 + v\mathcal{V}_l + a\mathcal{A}_l \quad \text{for} \quad l = 0, \ldots, L \tag{3.11}$$

The three functions ϕ^0, $\mathcal{V}$ and $\mathcal{A}$ are independent of velocity and acceleration and are fully characterised by the waveform choice. Explicitly:

$$\phi^0 = \frac{2}{c}R(f_0 + l\Delta f) \tag{3.12}$$

$$\mathcal{V}_l = \frac{2}{c}t_l(f_0 + l\Delta f) \tag{3.13}$$

$$\mathcal{A}_l = \frac{1}{c}t_l^2(f_0 + l\Delta f) \tag{3.14}$$

Clearly, if the three sequences $\phi_l^0, \mathcal{V}_l$ and $\mathcal{A}_l$ are linear in l (the time variable) or if they are independent of l, the derivative of the phase, i.e. the frequency, is constant. In that case, harmonic analysis by the use of the Fourier transform gives a well-focussed range profile. Observe that for any choice of t_l, the first term ϕ_l^0 is linear in l. We write for the change of the phase as a function of l using Equation 3.11:

$$\phi_l \Leftrightarrow \phi_{l-1} \equiv (\Delta\phi)_l = (\Delta\phi^0)_l + v(\Delta\mathcal{V})_l + a(\Delta\mathcal{A})_l \quad \text{for} \quad l = 1, \ldots, L \tag{3.15}$$

To arrive at a measure of distortion, we want to find the change of $(\Delta\phi)_l$ over the range $l = 1, \ldots, L$, which is equivalent to the change of

frequency. Our approach is to find the deviation of the frequency of the three terms during the sweep through the bandwidth:

$$\mathcal{D}(\Delta\phi) = |\mathcal{D}(\Delta\phi^0) + v\mathcal{D}(\Delta\mathcal{V}) + a\mathcal{D}(\Delta\mathcal{A})| \tag{3.16}$$

where

$$\mathcal{D}(\Delta\phi^0) \equiv \sqrt{\frac{1}{L}\sum_{l=1}^{l=L}((\Delta\phi^0)_l \Leftrightarrow \overline{\Delta\phi^0})^2} \tag{3.17}$$

In the same way $\mathcal{D}(\Delta\mathcal{V})$ and $\mathcal{D}(\Delta\mathcal{A})$ are defined. As ϕ_l^0 is linear $\mathcal{D}(\Delta\phi^0)$, the first term on the right hand side of Equation 3.16, is always zero. $v\Delta\mathcal{V}$ and $a\Delta\mathcal{A}$ are the distortions due to the velocity and the acceleration, respectively. For the three waveforms, $\mathcal{D}(\Delta\phi)$ evidently has the following forms:

$$\text{LW}: \quad \mathcal{D}(\Delta\phi) = |v\mathcal{D}(\Delta\mathcal{V}) + a\mathcal{D}(\Delta\mathcal{A})| \tag{3.18}$$

$$\text{VTW}: \quad \mathcal{D}(\Delta\phi) = |a\mathcal{D}(\Delta\mathcal{A})| \tag{3.19}$$

$$\text{ATW}: \quad \mathcal{D}(\Delta\phi) = |v\mathcal{D}(\Delta\mathcal{V})| \tag{3.20}$$

To scale the frequency variation to actual meters we have to multiply these dimensionless quantities with the size of the range window $R_U = \frac{Lc}{2\gamma f_0}$ so that we find the observed width $\mathcal{W}$ of a point scatterer in meters:

$$\mathcal{W} = \frac{Lc}{2\gamma f_0}|\mathcal{D}(\Delta\phi)| \tag{3.21}$$

3.5.2 Linear Waveform

Definition

The time points of emission of the frequencies f_l are chosen as $l\Delta t_{\min}$ where l runs from 0 until L. See Figure 3.8.

Effect of velocity and acceleration

The phase of a point reflector initially located at a distance R, with a radial velocity v and a radial acceleration a is given by

$$\phi_l = \frac{2}{c}(R + vl\Delta t_{\min} + \frac{1}{2}al^2\Delta t_{\min}^2)(f_0 + l\Delta f), \text{ for } l = 0, \ldots, L \tag{3.22}$$

if we use the Linear Waveform. In general, ϕ is a polynomial of degree 3 in l. Clearly, the Fourier Transform of a complex sequence with this phase behaviour gives a broadened peak because the frequency is non-constant. To find the effects of a non-zero v and a we use Equations 3.22 and 3.11. We find:

$$\phi_l^0 = \frac{2f_0R}{c}(1+\frac{\gamma}{L}l) \tag{3.23}$$

$$\mathcal{V}_l = \frac{2\Delta t_{\min} f_0}{c}(l+\frac{\gamma}{L}l^2) \tag{3.24}$$

$$\mathcal{A}_l = \frac{\Delta t_{\min}^2 f_0}{c}(l^2+\frac{\gamma}{L}l^3) \tag{3.25}$$

where γ is the fractional bandwidth, B/f_0. The first function is linear in l, and thus, as we saw before, causes no degrading effects on the HRR profile. For the second and third term we compute the deviations (Equation 3.16) and find the closed-form solutions:

$$\mathcal{D}(\Delta\mathcal{V}) = \frac{\Delta t_{\min} f_0 \gamma}{c}\sqrt{\frac{4L}{3(L+1)}}$$

$$\mathcal{D}(\Delta\mathcal{A}) = \frac{\Delta t_{\min}^2 f_0}{c} \times$$

$$\times\sqrt{\frac{(L \Leftrightarrow 1)(5+10L+5L^2+15L\gamma+15L^2\gamma \Leftrightarrow 3\gamma^2+12L^2\gamma^2)}{15(L+1)}}$$

In practical applications, $L+1 > 100$ and $\gamma < 0.25$, so good approximations of the above equations are:

$$\mathcal{D}(\Delta\mathcal{V}) \approx \frac{2}{\sqrt{3}}\frac{\Delta t_{\min} f_0 \gamma}{c} \tag{3.26}$$

$$\mathcal{D}(\Delta\mathcal{A}) \approx \sqrt{\frac{1+3\gamma}{3}}\frac{\Delta t_{\min}^2 f_0 (L+1)}{c} \tag{3.27}$$

The width $\mathcal{W}$ of a point scatterer is therefore (using Equation 3.21):

$$\mathcal{W}_{LW} = \frac{Lc}{2\gamma f_0}|\mathcal{D}(\Delta\phi)| \tag{3.28}$$

$$\approx \left| v\Delta t_{\min} L/\sqrt{3} + \frac{1}{2}a\Delta t_{\min}^2 L^2 \sqrt{\frac{1+3\gamma}{3\gamma^2}} \right| \tag{3.29}$$

$$\approx \left| \frac{1}{3}\sqrt{3}vT + \frac{1}{2}aT^2\sqrt{\frac{1+3\gamma}{3\gamma^2}} \right| \tag{3.30}$$

where $T = L\Delta t_{\min}$ is the waveform duration. Note that it is possible that the two terms cancel for specific values of a and v. In order to avoid distortions of the range profiles, $\mathcal{W}$ should be smaller than the size of one range cell:

$$\mathcal{W} < \Delta R_N \tag{3.31}$$

So a convenient measure for the distortion of the range profile due to the motions is:

$$\beta \equiv \mathcal{W}/\Delta R_N \tag{3.32}$$

As long as β is smaller than one, the profile will not significantly be distorted by the motions of the aircraft.

Let us now apply the resulting equations on the ORFEO parameters from Table 3.1. The plot on the left hand side from Figure 3.8 shows the waveform. The figure on the right hand side shows for which combinations of velocity and acceleration the parameter β is smaller than one. In this region the range profiles are expected to be focussed. The figure shows that only in a small region in the (a, v) plane focussed profiles result.

3.5.3 Velocity Tolerant Waveform

The *Velocity Tolerant Waveform* is known from the sonar literature as the *Hyperbolic Frequency Modulated* waveform — see, e.g. , [29]. However, it can also be used in stepped frequency radar waveforms [2]. The idea is that the times at which the pulses are transmitted are chosen such that the resulting range profiles are focused irrespective of the velocity.

Derivation

Assume that the point scatterer moves with constant velocity and zero acceleration and has an initial distance R. The phase is then given by:

$$\phi_l = \frac{2}{c}(R + vt_l)(f_0 + l\Delta f) \tag{3.33}$$

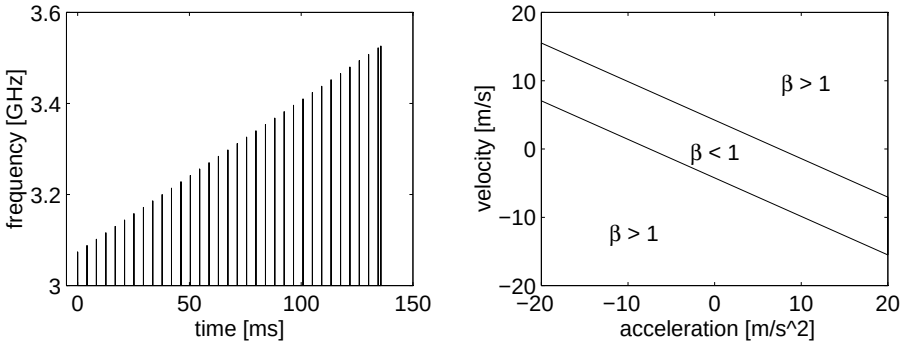

Figure 3.8: *Linear Waveform using the ORFEO parameters. Left: waveform, only the pulses* $0, 10, 20, \ldots, 320$ *and the last pulse* 323 *are shown. Right: regions where the profiles will* ($\beta < 1$) *and will not* ($\beta > 1$) *be focussed.*

Choosing $t_0 = 0$ we write:

$$t_{l+1} = t_l + \Delta t_l \tag{3.34}$$

and

$$t_{l+2} = t_l + \Delta t_l + \Delta t_{l+1} \tag{3.35}$$

Now we demand that the differences in phase are constant

$$\phi(t_{l+2}) \Leftrightarrow \phi(t_{l+1}) = \phi(t_{l+1}) \Leftrightarrow \phi(t_l) \tag{3.36}$$

If we insert Equations 3.33, 3.34 and 3.35 into 3.36 it can be derived formally that the time instants at which the pulses should be emitted are:

$$t_l = \Delta t_0 \frac{l(\gamma + L)}{\gamma l + L} \tag{3.37}$$

Note that this solution for t_l, defined as the *Velocity Tolerant Waveform*, makes the product $t_l f_l = l f_0 \Delta t_0 \frac{\gamma + L}{L}$ linear in l. The unknown constant Δt_0, which is the delay between the first two pulses, can be found from the requirement that the smallest timestep possible should

equal $\Delta t_{\min}$. These steps decrease as l increases, so $\Delta t_{\min}$ should equal the delay between the two final pulses:

$$\Delta t_{\min} = t_L \Leftrightarrow t_{L-1} \tag{3.38}$$

We find:

$$t_l = \Delta t_{\min} \frac{l(1+\gamma)(L \Leftrightarrow \gamma + L\gamma)}{\gamma l + L} \tag{3.39}$$

See also Figure 3.9. The total duration of the waveform is now:

$$T = \Delta t_{\min}(L \Leftrightarrow \gamma + L\gamma) \tag{3.40}$$

The following inequality may be of interest:

$$T = \Delta t_{\min}(L(1+\gamma) \Leftrightarrow \gamma) \leq \Delta t_{\min}(2L \Leftrightarrow \gamma) < 2\Delta t_{\min}(L+1) \tag{3.41}$$

which shows that, using the same radar parameters, the total duration of the VTW is always less than twice the duration of the LW.

Another positive side effect using the velocity tolerant waveform instead of the linear waveform can be mentioned. In the previous chapter we saw that JEM, Jet Engine Modulation, can be used to identify aircraft. For measuring range profiles with an LW, i.e. with equal time delays between the pulses, the JEM lines are a nuisance because they cause disturbances throughout the range profile and the positions of these disturbances change with the engine shaft rotation speed. In the velocity tolerant the time delays are not equal, but vary smoothly. The result is that the JEM lines are effectively decorrelated [2] .

Effect of acceleration on VTW

If we start from Equation 3.20 we find for the phase:

$$\phi_l = a\mathcal{A}_l + \mathrm{O}[l^1] + \mathrm{O}[l^0] = a\frac{f_0 \Delta t_0^2}{c} \frac{l^2(L+\gamma)^2}{L(L+l\gamma)} + \mathrm{O}[l^1] + \mathrm{O}[l^0] \tag{3.42}$$

It appeared not to be possible to find a closed-form expression for $\mathcal{D}(\Delta\mathcal{A})$ directly with Equation 3.17, but we approximated the summations by integrals and used the fact that in practical applications L

is large and γ is small. We found a useful approximation for $\mathcal{W}_{VTW}$ (See also 3.21):

$$\mathcal{W}_{VTW} \approx \left| \frac{1}{2} a \Delta t_{\min}^2 L^2 \sqrt{\frac{1+\gamma}{3\gamma^2}} \right| \tag{3.43}$$

Again we apply the equations to the ORFEO parameters from Table 3.1. The plot on the left hand side from Figure 3.9 shows the waveform. The figure on the right hand side shows the regions in which the range profiles will be well-focussed.

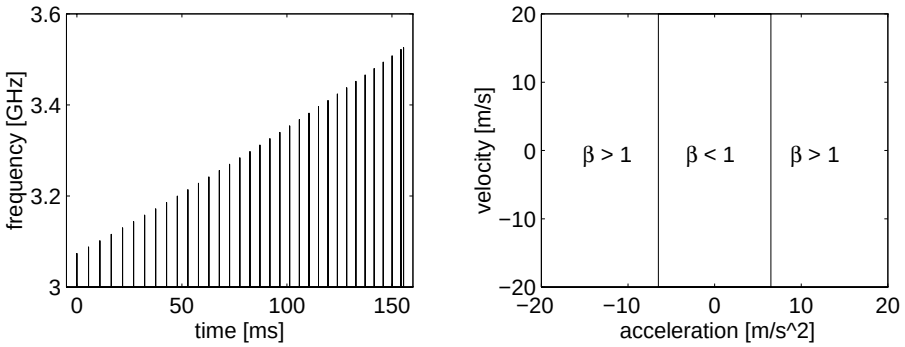

Figure 3.9: *Velocity Tolerant Waveform using the ORFEO parameters. Left: waveform, pulses* $0, 10, 20, \ldots, 320$ *and the last, rightmost pulse* 323*. The change in delay times is present but hardly noticeable. Right: region where* $\beta < 1$ *so that the profiles will be focussed.*

3.5.4 Acceleration Tolerant Waveform

Derivation

The essence of the velocity tolerant waveform is that the product $t_l f_l$ is linear in l. Analogously we can see that for acceleration compensation it is sufficient to demand that $t_l^2 f_l$ is linear.

We thus define for the Acceleration Tolerant Waveform:

$$t_l = \Delta t_0 \sqrt{\frac{l(\gamma + L)}{\gamma l + L}} \text{for } l = 0, \ldots, L \Leftrightarrow 1 \tag{3.44}$$

Where Δt_0 can be found using the requirement:

$$t_L \Leftrightarrow t_{L-1} = \Delta t_{\min} \tag{3.45}$$

So that:

$$\Delta t_0 = \frac{\Delta t_{\min}}{\sqrt{\frac{\gamma+L}{1+\gamma}} \Leftrightarrow \sqrt{\frac{(L-1)(\gamma+L)}{L+\gamma L-\gamma}}} \tag{3.46}$$

See also Figure 3.10.

Like the Velocity Tolerant Waveform, the ATW is not very sensitive for JEM interference as the time instants at which the pulses are emitted are not equidistant.

Effect of velocity on ATW

In this case, both ϕ^0 and $\mathcal{A}$ are linear in l. A useful expression for $\mathcal{W}_{ATW}$ (eq. 3.21) as a function of v is not available so we have to resort to a numerical evaluation of $\mathcal{D}(\Delta\mathcal{V})$. $\mathcal{V}_l$ equals

$$\mathcal{V}_l = \frac{2\Delta t_0 f_0}{cL}\sqrt{l(\gamma+L)(\gamma l+L)} \tag{3.47}$$

where Δt_0 is given by Equation 3.46. The values $\mathcal{V}_l$ can be used to find the differences $(\Delta\mathcal{V}) \equiv \mathcal{V}_l \Leftrightarrow \mathcal{V}_{l-1}$, for $l = 1, \ldots, L$. To find $\mathcal{D}(\Delta\mathcal{V})$ and $\mathcal{W}_{ATW}$ is straightforward from equations 3.17 and 3.21.

Finally, we also apply the results of the acceleration tolerant waveforms to the ORFEO parameters from Table 3.1. Again, the plot on the left hand side from Figure 3.9 shows the waveform (only the pulses $0, 10, 20, \ldots, 320$ and 323 are shown). The figure on the right hand side shows the regions in which the range profiles will be well-focussed.

3.5.5 Waveform choice

If we compare the right hand side plots of the Figures 3.8, 3.9 and 3.10 we see that the Velocity Tolerant Waveform has the best tolerance for both radial velocity (any) and for radial accelerations (up to 6.3 ms^{-2}). This amount of acceleration is larger than civil airliners usually exhibit. In the ORFEO measurement campain, we have seen these accelerations in only very few cases [33].

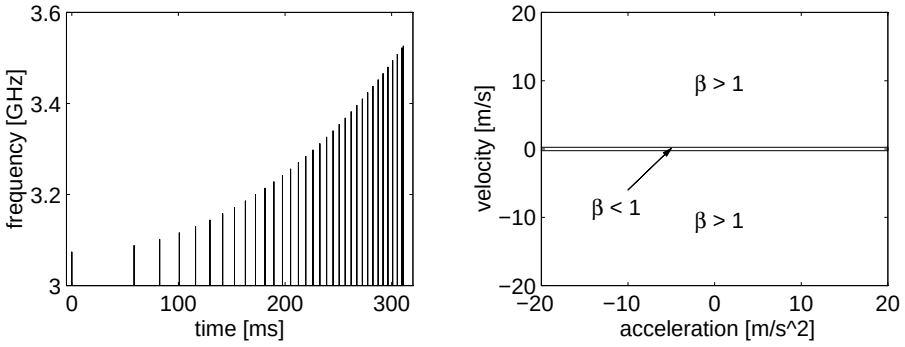

Figure 3.10: *Acceleration Tolerant Waveform using the ORFEO parameters. Left: waveform, pulses* $0, 10, 20, \ldots, 320$ *and* 323*. Right: in only a very small region, the profiles are expected to be focussed* ($\beta < 1$)*.*

The acceleration tolerant waveform is very sensitive for velocity and the linear waveform produces focussed range profiles only for a small range of combinations of velocity and acceleration.

Keeping in mind that the typical velocity of a large civil airliner is 300 ms^{-1}, we have therefore chosen the Velocity Tolerant Waveform for range profile generation in the ORFEO measurements.

3.6 Variability of range profiles

3.6.1 Sources of variability

The following parameters have an influence on the shape of a range profile.

1. *Radar parameters*

   We mention carrier frequency, bandwidth, polarisation and waveform choice.

2. *Absolute radar calibration*

3. *Additive noise*

Thermal noise and quantisation noise.

4. *Aircraft geometry*

   This dependence obviously makes range profiles suitable for recognition purposes. However the recognition problem for, in particular, fighter aircraft is made more complicated by factors such as scanning radar antennas, external stores that may or may not be present, and removable RAM tiles for the reduction of the total amount of reflection[2].

5. *Aircraft distance*

   The aircraft distance defines

   - the positioning of the aircraft within the range profile and
   - the energy content of the range profile elements.

6. *Aircraft motion*

   The range profiles we use in this thesis are not distorted by aircraft motion.

7. *Aircraft aspect angle*

   The dependence on aspect angle is one of the most important issues in range profile classification as the profile shows an extreme variability as a function of this variable. It is decisive for the design of a classification system.

The next three subsections deal with the sources of variability (items 2, 5, 6 and 7 from the aforementioned list) that introduce fluctuations of range profiles that have a direct influence on the design of a range profile classification system.

3.6.2 Variability in amplitude scaling

Calibration (item 2 from 3.6) must be known accurately to scale the values of the range profile elements to units of energy. However, we

[2]RAM = Radar Absorbing Material

do not want to rely on a, possibly inaccurate, estimate for the calibration. Throughout this thesis we restrict ourselves to relative energies by making the sum of energies equal to one (normalisation, specified in 5.3.3).

3.6.3 Variability of aircraft location in profile

The exact motion parameters define the position of the set of aircraft scatterers in the range profile. These parameters must be known accurately to place the scatterers at an objective position: it should be known within the accuracy of the size of a range resolution cell. (Recall that the range resolution for the ORFEO measurement is 0.33 m.)

Again, we do not want to rely on possibly inaccurate estimates. Therefore *shift-invariant* classification techniques are used, see Chapter 5.

3.6.4 Variability due to aspect angle

To acquire a feeling for the sensitivity to aspect angle, the following example from the ORFEO measurements is taken.

Target	Boeing 747-400		
Target height (landing gear stowed)	T_H	=	17 m
Target wing-span ($\approx$ target length)	T_D	=	64 m
Length target wing leading edge	T_W	=	32 m
Nominal Range Resolution	ΔR	=	0.33 m
Wavelength	λ	=	0.1 m

Table 3.2: *Example from ORFEO measurements.*

As the waveform is emitted within several tenths of seconds we will neglect the change in aspect angle during the measurement of a single range profile. We believe the errors on the estimated aspect angles are less than 5°. Five effects are responsible for the dissimilarities between two profiles measured from the same aircraft at different aspect angles: speckle, rotational range migration, shadowing effects, specular reflections and polarimetric effects.

Speckle

If in a single resolution cell two distinct scatterers are present, only a slight rotation of the aircraft in aspect azimuth can be enough to change the differential path length to the radar over half the wavelength. See Figure 3.11.

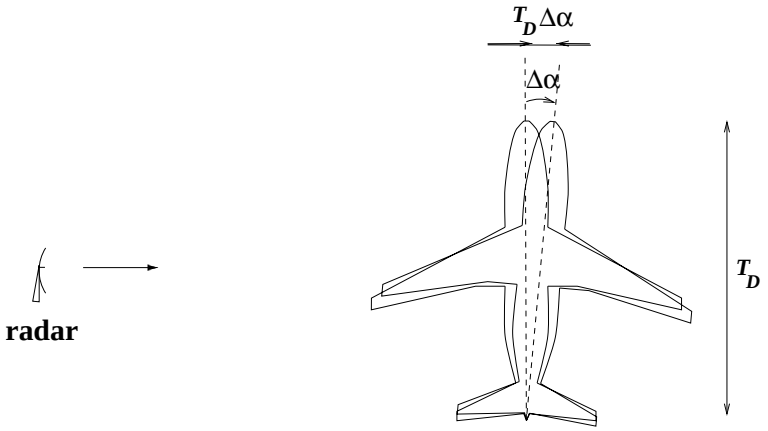

Figure 3.11: *The aircraft rotates over an aspect azimuth of $\Delta\alpha$. The two outermost scatterers are separated by a distance T_D and are located within the same resolution cell. These scatterers change their relative path length over $T_D\Delta\alpha$ during this rotation.*

This effect, *speckle*, causes the sum of the two scatter contributions to turn from constructive to destructive interference within a very small change of aspect azimuth.

If T_D is the maximum distance between two scatterers within one resolution cell, the change in aspect angle $\Delta\alpha$ between two profiles must be much smaller than

$$\Delta\alpha_{\text{speckle}} \equiv \frac{\lambda}{4T_D}[\text{rad}] \tag{3.48}$$

to avoid a difference between the two profiles due to speckle [41].

If we view the Boeing 747 from broadside and if we assume that a scatterer is present on both the tail and the nose, we find that the change in aspect azimuth $\Delta\alpha$ should satisfy

$$\Delta\alpha \ll \Delta\alpha_{\text{speckle}} = 0.022^\circ \tag{3.49}$$

to avoid differences due to speckle between two range profiles.

The result of speckle is that the amplitudes of the range profile elements that contain multiple scatterers vary rapidly if a sequence of subsequently measured range profiles is considered — the change in aspect angle is due to small aircraft yaw motions during the recording time. The peak positions do not alter. See Figure 3.12.

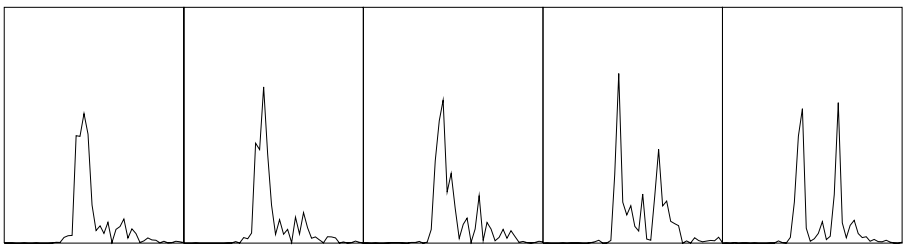

Figure 3.12: *Range profile variations due to speckle. The peak positions are unaltered, the large variations of the peak magnitudes are caused by speckle.*

For three specific, idealised cases it is possible to analytically derive the statistical behaviour of range profile elements as a function of aspect angle:

1. if the rotations have a normal distribution with a standard deviation much smaller than $\Delta\alpha_{\text{speckle}}$,

2. if a range profile element contains exactly two scatterers,

3. if an infinite number of independent scatterers is present in the range cell.

In practice, the first situation never occurs. The second case is not usable either, as the number of scatterers seldomly equals two. Even if this were the case at some time instant, it would change due to the migration of the scatterers from one cell to the other.

The last case is based on the central limit theorem. It may be of interest as for a relatively small number of scatterers an approximation holds. It is often used in SAR imaging and results in an exponential

distribution of energy. For a derivation and detailed further discussion we refer to Ulaby [74].

In general we may state that the speckle fluctuations are mainly multiplicative as the amount of variation of a particular range cell increases with its average value.

The change in aspect *elevation* may also cause speckle. If we consider an aircraft at zero aspect elevation, small roll motions may cause scatterers within the same resolution cell change their relative pathlength to the radar.

For zero elevation we may simply write, equivalent to equation 3.48:

$$\Delta\theta_{\text{speckle}} \equiv \frac{\lambda}{4T_H}[\text{rad}] \tag{3.50}$$

where T_H is the height of the aircraft. For non-zero elevations, T_H should be replaced by $T_H \cos\theta$

From Table 3.2, we find for the FELSTAR measurements of the Boeing-747:

$$\Delta\theta \ll \Delta\theta_{\text{speckle}} \equiv 0.08^\circ \tag{3.51}$$

It shows that this effect is of much less importance than speckle due to the change in aspect azimuth.

Note that both the value for $\Delta\theta_{\text{speckle}}$ and for $\Delta\alpha_{\text{speckle}}$ are lower bounds as we inserted the maximum dimensions of the aircraft. Scatterers are not necessarily present on these outer parts so usually the true separations between multiple scatterers are less than the "optical" dimensions.

Rotational Range Migration

If an aircraft rotates over a large azimuth angle, such that the outermost scatterers move from one resolution cell to the other, the measured range profiles during this rotation suffer from Rotational Range Migration (RRM).

If two profiles have aspect angles whose difference is less than

$$\Delta\alpha_{\text{RRM}} \equiv \frac{\Delta R}{T_D}\ [\text{rad}] \tag{3.52}$$

then the profiles do not differ due to Rotational Range Migration. The example gives $\Delta\alpha_{\mathrm{RRM}} = 0.3°$. The angle changes over which RRM occurs is usually one or two orders of magnitude larger than the changes over which speckle results.

Rotational range migration also occurs if the aspect elevation changes over a large angle. See Figure 3.13. Due to the projection effect, the

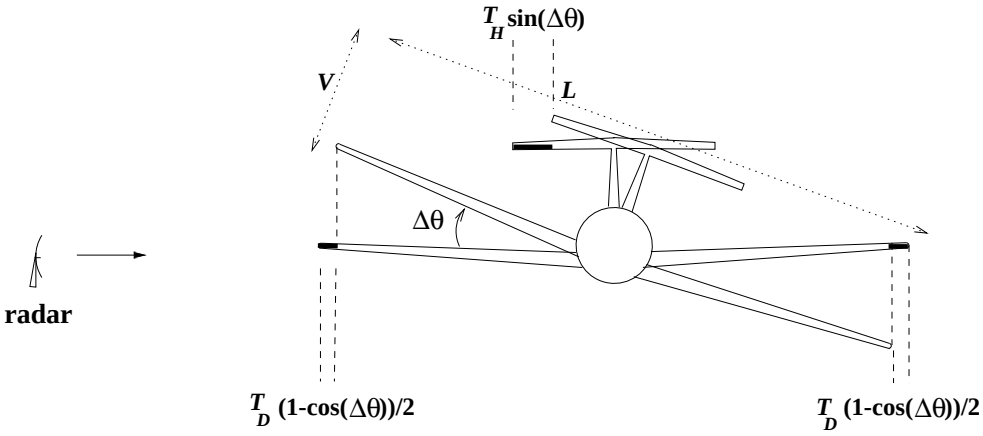

Figure 3.13: *Differential path length due to change in aspect elevation. In this geometry, the aircraft has its nose pointed perpendicular to the paper (towards the reader) and makes a rotation in elevation of $\Delta\theta$ (roll motion).*

change in path-length is a second order effect. The maximum change in aspect elevation that does not imply a difference due to rotational range migration is:

$$\Delta\theta < \Delta\theta_{\mathrm{RRM}} \equiv \min(\arccos(1 \Leftrightarrow \frac{\Delta R}{T_D}), \arcsin(\frac{\Delta R}{T_H})) \tag{3.53}$$

For the example we find $\Delta\theta_{\mathrm{RRM}} = 1.1°$.

Thus the following view may be adopted: during the measurement of a sequence of range profiles in real flight, the aircraft rotates with respect to the radar. Over large rotations the relative peak positions in the profiles change due to Rotational Range Migration. A small sector may be defined where RRM does not occur, which is called an *RRM-sector* .

Within this region the statistics of the range profile elements are governed by speckle. This speckle phenomenon may be regarded as a noise process, which is mainly multiplicative.

A common situation (in e.g. *surveillance*) is where the aircraft is flying level at a fairly large distance. Consequently, the aspect elevation is close to zero. Now the change in orientation as seen from the radar is largely on account of a net yaw rotation. This rotation is caused by both aspect azimuth change due to target translation, and the inherent target yaw motions induced by instability or external disturbances such as cross winds or turbulence.

Shadowing

Due to the change in aspect angle of an aircraft, either in azimuth or elevation, scatterers may appear or disappear due to obscuration. An example is where the aircraft is looked upon from broadside: scatterers may emerge abruptly from the engines on the far wing if the aircraft executes a roll or increases its height.

The aircraft geometry, aspect angle and change in aspect angle determine the range profile variation due to shadowing effects. The effect of shadowing is an abrubtly increased or decreased value of one or more range cells.

Shadowing effects are expected to be highly discriminative for a specific aircraft type if a *sequence* of range profiles is considered.

If during such a sequence a part of the object appears induced by a change of aspect angle, a sudden increase in one or more range cells is observed. Then, if in a new measurement under the same conditions a similar dynamical behaviour is seen, the target is probably of the same type as in the first measurement. The trouble for employing this effect is the sharp dependence on aspect angle: the effect will only occur again if the aircraft changes its orientation in the same way. Another problem that makes this feature not useful for recognition is that the errors on the estimated aspect angles are usually several orders of magnitude larger than the aspect angle changes associated with shadowing effects.

Specular reflections

If a flat part on the aircraft happens to be pointed perpendicular to the radar line of sight, a large reflection from this part is returned. Such a flash is highly aspect angle dependent.

The following example may clarify this: suppose the leading edge of a wing has a length of T_W meters and the radar is looking perpendicular on it. See Figure 3.14. Then this edge will behave like an antenna and

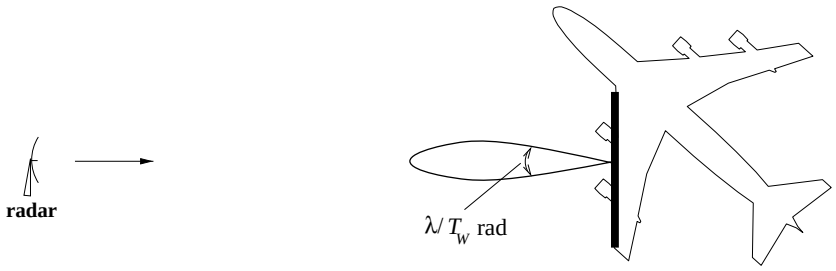

Figure 3.14: *Specular reflection on edge of wing.*

we will see the specular reflection over an aspect angle range of approximately λ/T_W [rad]. For our example we find that over roughly 0.25° this reflection will be seen. It clearly follows that for a minor change in aspect angle, the amount of reflected energy changes dramatically.

Similar to the shadowing effect, specular reflections may be utilised for classification if the dynamic behaviour of a sequence of range profiles is considered. The specific aspect angles at which the specular reflections occur may be very discriminative for the target type. Again, however, the difficulty is the sharp dependence on aspect angle and the large errors in the estimated aspect angles.

If we do not use the dynamics of series of range profiles, this effect gives rise to *lower* classification rates because contributions from other parts of the aircraft are dwarfed if one the range cells contains a specular reflection. A range profile with only a single strong scatterer will look the same as profiles from other aircraft viewed at one of their specular aspect angles.

Polarimetric effects

Differences due to the polarimetric state of the radar are present if two profiles are measured from the same aircraft rotated around the line of sight. E.g. if we would use vertical polarisation on transmit and receive, vertical features in the aircraft geometry would have a much larger return than horizontal features. If the aircraft is rotated around the line of sight over 90 degrees, originally horizontal features would be more prominent.

For example, this effect occurs if the aircraft executes a roll when it is viewed from nose-on. In comparison to the previously mentioned sources of variability due to aspect angle, this effect will be negligible as very large changes in aspect angles are involved.

Given a radar that is able to provide the full scattering matrix, polarisation information can be exploited to improve recognition. E.g. a range cell that contains a single scatterer exhibits the polarisation characteristics of this scatterer — in this way, range profiles can be made more discriminative. A currently unanswered question is, however, if expensive inclusion of the scattering matrix capability in a radar is cost-effective.

3.6.5 Resulting requirements for a classification system

The assessment of the variability of range profiles has clarified what requirements a range profile classification system has to meet.

1. *Robustness for absolute amplitude*

   A classifier needs to be independent of the absolute amplitudes of the range profile. This can easily be done by normalising the range profile such that the total energy equals one.

2. *Robustness for shifts*

   A classifier must be shift-invariant as the position of the aircraft in the profile is ill-defined.

3. *Robustness for aspect angle errors*

The change in aspect angle over which a profile fluctuates is several orders of magnitude smaller than the errors in the aspect angle estimates. Therefore a classifier has to be trained using the information from profiles from an aspect angle range that is larger than the error in aspect angle.

3.7 Summary

In this chapter we have provided the groundwork for the production of radar range profiles. The physics of range profiles measured with stepped frequency waveforms, highlighting the fundamental parameters and trade-offs, were reviewed.

Several waveform choices were shown to be available for measurements, depending on the motion tolerance one desires. Given the radar parameters used in the ORFEO measurements with the FELSTAR radar and assumed target characteristics, the Velocity Tolerant Waveform proved to be the best option.

We have also reviewed the variability of range profiles in amplitude, target position, and due to aspect angle. This resulted in a specification of the demands on a range profile classification system: it should be independent for shifts and absolute amplitude and it should be robust for errors on estimated aspect angle.

Chapter 4

Range profile acquisition and experimental set-up

4.1 Introduction

For classification experiments on radar range profiles we need data to train and test a classifier. It is imperative to understand which assumptions and trade-offs are made when using range profile measurements for such tests.

In general, four sources of data can be identified: **1)** controlled measurements, **2)** non-controlled measurements, **3)** scale-model measurements and **4)** computer simulations.

For research purposes no doubt the first source is most valuable: the aircraft carry instrumentation such that accurate positions and attitudes of the target are known throughout the flight. Similar information can be acquired by hanging a real aircraft on cables and measure it from all desired directions — an extra advantage being that the aspect angles can be defined accurately in advance. Evidently, such data are available only for fully cooperative aircraft.

The second type of measurements is easier to acquire, as the objects do not need to carry instrumentation — they even do not have to be cooperative and can be 'targets of opportunity'. Clearly, for training and testing a classifier on such data one needs to have knowledge on the target type which can be provided by a secondary radar system (IFF). Testing a classifier with range profiles from this source simulates the operational situation most faithfully as the aspect angles are known by approximation only[1].

Scale model data are acquired by fixing a downsized replica of the object of interest onto a turntable and view it from all desired aspect angles. In such a set-up the radar will have to operate at frequencies that are multiplied by the downscaling factor of the model.

Finally, range profiles can be produced using accurate models of aircraft and computational electromagnetics.

Data from both the first and third source are relatively expensive to acquire as one requires the availability of a real aircraft or a detailed scale-model of an aircraft and additional measurement facilities. For the work in this thesis we shall therefore use data from the second and

[1] Evidently, the same type of data can be produced by adding noise to the aspect angles corresponding to the range profiles from the first source.

the fourth source.

The measured data originate from two different S-band radars. The simulated range profiles have been produced by a software package for computational electromagnetics in conjunction with accurate models of aircraft. A large part of this chapter is devoted to the background of these computations and a description of the aircraft models. For the simulations, as many profiles can be computed as desired. To limit the computational effort in both the electromagnetic simulations and the classification experiments, it needs to be examined carefully at which resolution in aspect angle the range profiles need to be sampled and still be to find adequate classification performance.

In the choices for selecting data in order to perform classification experiments, it is essential to understand which assumptions and trade-offs are made. The scarcity or even lack of measured data in classification experiments makes it tempting to accept undesired correlations and major simplifications resulting in too optimistic and/or ill-understood results.

The outline of this chapter is as follows: the next section will deal with the trade-off involved in selecting a training and test set from measured data only. Then, the first set of real data is presented, followed by a description of the so-called *ORFEO measurement campaign*. Subsequently, the background on the computational electromagnetics is reviewed, together with a description of the predicted range profiles.

To complete this introduction, let us make three remarks on the terminology used. In the radar community the results of the computational electromagnetics are commonly referred to as *predictions* and the software to do this is called a *Radar Cross Section (RCS)-prediction code*. We will also adopt the term *leg*, which is a collection of range profiles that is acquired in one continuous measurement from a single aircraft while the radar is tracking the object.

Finally, we will use here, and throughout this thesis, the terms *input*, *training* and *test* set. The input set contains a collection of range profiles with known class labels, from which a subset is chosen to construct a classifier. This subset is called the training set. We will choose the training set as a subset of the input set in order to study the influence of the size of the training set on classification performance. The test set contains independent range profiles, also with class labels.

Independence means that the profiles are acquired in measurements different from the input set. These profiles enable us to assess the performance of the classifier.

4.2 Philosophy of data usage

4.2.1 Options without availability of measured data

In many reports on range profile classification, techniques have been trained *and* tested on synthetic, computer-generated data [51, 64, 71] and scale-model measurements [44, 43, 27, 52].

Using this data for both training and testing obscures several problems that do need to be addressed for range profiles from in-flight aircraft [80]. It therefore implies that many of the classification techniques proposed in the forementioned references should be enhanced with additional preprocessing steps before they can be applied to classify real data, if they can be applied at all. We mention the treatment of uncertainties in the target position in the range profile, the total strength of the target return and errors in aspect angle. Additionally, the computer-generated data (but also the scale-model data if simple objects are used) have an intrinsically lower complexity than the measurements on real aircraft.

Clearly, utilising such data sources are the best one can do if no real data are available: the (usually tacit) assumption is that if the technique does not work for *this* data, then it will certainly not work for real data.

4.2.2 Training and testing on measured data

In two instances in the literature, measured data from targets of opportunity are taken for training and testing a classifier [41, 80]. We are also in the fortunate position that we have such data available, so that we approach more closely the eventually desired application, the classification of range profiles in an operational environment. Now, however,

we need to justify the selection of an input set and a test set from the full collection of range profiles, that are grouped in legs.

Two approaches are possible:

1. *Choose independent input and test sets.*

   This is done by using the profiles from a particular leg for either the input set or the test set.

   The drawback is that generally a low number of different aircraft types classes are left over in the sets: only for a small number of aircraft multiple measurements at approximately the same aspect angles are available. In [80], e.g. , only two aircraft are contained in the independent input and test sets.

2. *Allow profiles for input and testing to be taken from the same legs.*

   Usually, this leads to input and test sets with a large number of classes, as from each leg a portion can be used for the input set and another portion can be used for the test set. This is desirable, because in future real world applications, the number of classes will probably be large too.

   In this approach, one should leave a time-gap between profiles for input and testing to lower the dependence between input and test set. In [41] data are taken for both training and testing from the same legs and they use a gap of ≈ 2 s.

   The gap is a trade-off between two conflicting requirements.

   (a) It should not be chosen too large, as in that case the aspect angle changes too much between input and test profiles.

   (b) It should not be chosen too small, such that we look at the aircraft at essentially the same aspect angle without sufficient random rotation.

In the first approach we will have a too optimistic estimate of the classification error compared to the operational situation because the number of classes is low.

In case 2 there is always an undesired correlation between the profiles present; **1)** we look at exactly the same specimen of the particular

type of aircraft (this is especially important when looking at fighter aircraft which can have different external store configurations); **2)** we measure the aircraft in approximately the same direction at the same weather conditions, which means that clutter and multi-path contributions are comparable; **3)** the aircraft is measured at approximately the same distance, leading to similar signal-to-noise ratios. All these correlations, unknown in magnitude, give rise to overoptimistic classification results.

To narrow down the number of causes for too optimistic results we have chosen for the first option where two or more specimen of the same aircraft type are measured independently.

4.2.3 Hybrid use of range profile data

Many interesting results are expected to be found if the data used for training come from a different type of source than the test data. The most important combination is where the test set is filled with range profiles from non-cooperative, real targets and the training set with either scale-model data or predictions. If these classification experiments are successful, the need for large trials on cooperative targets is much less as the required data can be produced on turntables or with computers.

4.3 Data set *I*: S-band data of four aircraft

Data set I contains an input set of 357 profiles of four different aircraft that have been measured with an S-Band radar ($\lambda \approx 0.1$m). The targets are simply labeled by $1, \ldots, 4$. Independently, 339 profiles were obtained from the same four aircraft types. These profiles made up a fully independent test set. The number of elements of the profiles, i.e. the dimension d, equals 128.

Figure 4.1 shows some examples of range profiles from this data set.

For each profile, an approximate aspect angle is available. The absolute aspect azimuths (we assume that the aircraft is symmetric around nose-on aspect angle $\alpha = 0°$) are in the range from $0°$ to $20°$. The errors

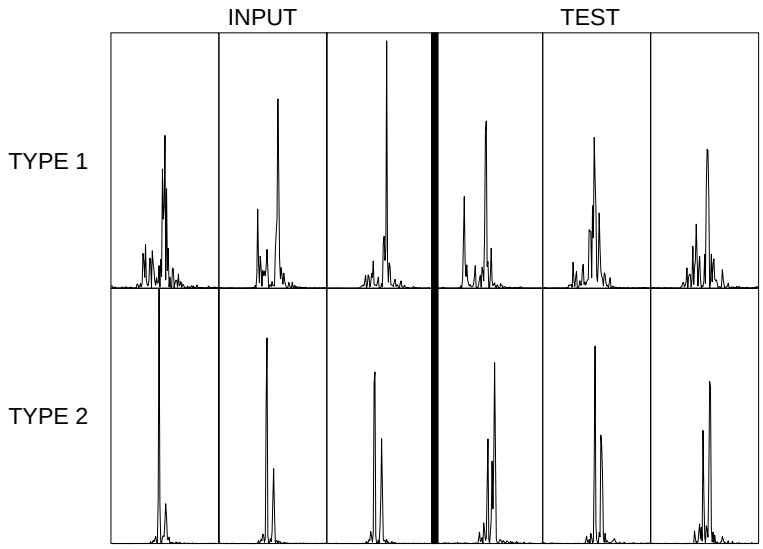

Figure 4.1: *Examples of profiles from Data set I.*

on the azimuths angles are believed to be within 5°. All aircraft have flown level on relatively large distances, therefore the aspect elevation θ is close to 0°.

4.4 ORFEO measurements

Data sets *II* and *III* together consist of 11 legs of range profiles from civil targets of opportunity, measured with the FELSTAR S-band radar at TNO-FEL. The range profile collection was done in the *ORFEO measurement campaign*, carried out for, partly, the research in this thesis. In this section we will summarise additional relevant information — details on the FELSTAR radar, measurement procedures, the processing of the raw radar data and the tracking data can be found in a technical report [33].

4.4.1 Waveform and profile processing

The waveform used for these measurements, the Velocity Tolerant Waveform, was discussed in Section 3.5. Application of this waveform assures

that the profiles are free of distorting influences from radial velocity. The maximum acceleration that can be tolerated without expecting a distortion of the profile is 6.3 m/s^2: as the legs we consider in this thesis were selected from pieces of the trajectory where no turns were performed, the acceleration will therefore be well below this limit and hence no additional acceleration compensation is required.

Following Section 3.3, the profiles were Hamming weighted and zero-padded by a factor two to (partly) reveal spectral contributions that are within grid-points.

4.4.2 Information on aircraft type

During the measurements information from an IFF system was available that provided the aircraft type. This type is given in a four-digit string like `FK10`, `EA31` or `B747`. Different series of the same aircraft may however have the same name: for our data we have ambiguities for two different aircraft labels, the `B73S` and the `MD80`.

1. The designation `B73S` is used to represent a Boeing 737 from both the 300 and the 500 series. It is therefore not possible to tell which of the types is actually measured. The two aircraft types are identical apart from the length of the fuselage, being 33.4 m and 31.0 m for the 300 and the 500 series, respectively.

2. The code `MD80` is used to represent all aircraft from the McDonnell Douglas MD-80 series. Externally these are all identical — only the MD-87 differs as it has a five meter shorter fuselage.

Note that these two equivocalities are of negligible influence when the aircraft is looked upon from broadside.

4.4.3 Target aspect angle

For each range profile, we estimated the target aspect angle (α, θ) from the tracking data, taking into account the target position, motion and roll-angles. We have taken care to measure the aircraft a considerable time before landing and after take-off to make sure that the landing

gear is stowed. The legs are to a good approximation straight lines: it is very improbable, therefore, to expect large roll-angles of the aircraft.

We unfortunately have no firm estimate of the errors in the aspect angle coordinates. The differences in aspect angles will be fairly accurate — a bias on both aspect azimuth and elevation for the whole leg could nevertheless be present.

We are not certain about the maximum bias to choose: a worst-case analysis on the errors on aspect azimuth (through evaluation of the maximum uncertainties in radar azimuth and range) gave an error of maximally four degrees. Generally the uncertainties due to these measurement errors will be much lower than four degrees, for most cases below two degrees. However, the main issue is that this analysis is done under the assumption that the aircraft moves in the direction of its nose. This is not the case in general: the presence of wind makes the aircraft fly at so-called slip-angles. We unfortunately do not have any information of the wind conditions during the measurements available. We shall therefore take a maximum angle error considerably larger than the errors expected from the angle estimation procedure — we assume that the errors on aspect angles do not exceed 5 degrees for the legs we consider, as is indicated by the results in Chapters 7 and 8.

As the aircraft flies level for each of the legs and does not perform any manoeuvring we do not expect large errors in aspect elevation. A maximum error of 5 degrees in aspect elevation is therefore a conservative estimate.

We thus assume that the bias for a particular leg will not exceed 5 degrees for both α and θ. For simplicity, we will assume a uniform distribution in this cone arount the estimated aspect angle. If (α, θ) and $(\langle\alpha\rangle, \langle\theta\rangle)$ are the true and the estimated aspect angles, respectively, we thus assume the following probability density function for a single range profile.

$$p(\alpha,\theta) = \begin{cases} \frac{1}{25\pi} & \text{if } (\alpha \Leftrightarrow \langle\alpha\rangle)^2 + (\theta \Leftrightarrow \langle\theta\rangle)^2 \leq 25 \text{ deg}^2 \\ 0 & \text{elsewhere} \end{cases} \tag{4.1}$$

4.4.4 Data set *II*

Data set *II* is used for a *comparison* between measurements of three legs of the Boeing 737-500, (code `B73S`), and synthetic data. These legs, measured from three different individual aircraft seen at near nose-on, broadside and tail-on are numbered 1, 2 and 3, respectively.

Figure 4.2 shows the aspect angles of the three legs.

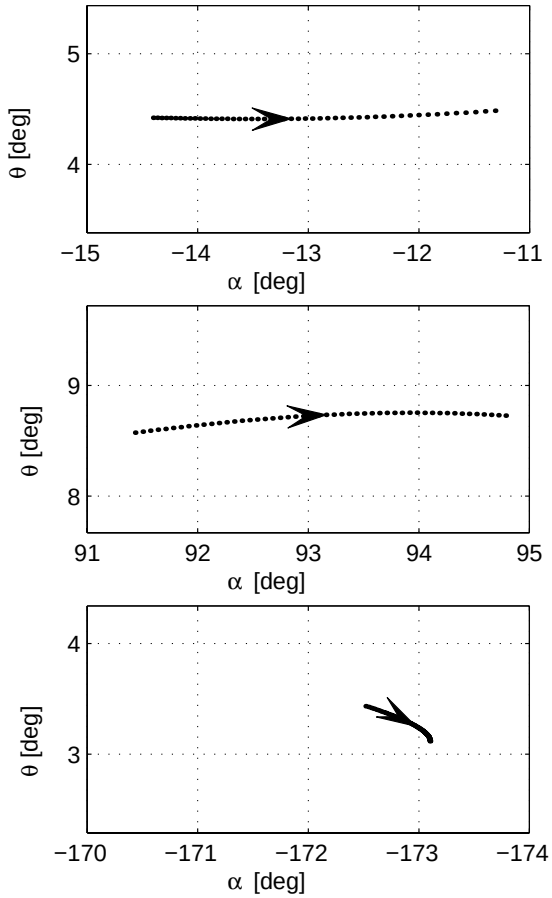

Figure 4.2: *Aspect angles of measured range profiles from* `B73S` *used in comparison (Data set II).*

4.4.5 Data set *III*

The third data set, Table 4.1, contains eight legs from five different aircraft, measured at approximately broadside aspect angles and numbered $1, 2, \ldots, 8$. On this set *classification* experiments will be performed: it shall be taken as a test set while synthetic data are employed for training.

Leg	*Code*	*Name*	$\overline{\alpha}$	$\overline{\theta}$
1	B73S	Boeing 737-300/500 series	93.3	8.7
2	B73S	"	94.2	4.5
3	B73S	"	93.5	6.2
4	B74F	Boeing 747-400	89.6	4.0
5	EA31	Airbus Industry, A310	90.8	5.6
6	FK10	Fokker 100	92.9	2.1
7	MD80	McDonnell-Douglas, MD-80/81/82/83/87/88	93.8	5.9
8	MD80	"	94.0	10.6

Table 4.1: *Data set III: eight sequences collected by FELSTAR. The second-last and last column are the average aspect azimuth and average aspect elevation, respectively, computed from the radar tracking data.*

Figure 4.3 shows the aspect angles of all the range profiles from Data set *III*, and a range profile from each of the legs.

Using equations 3.52 and 3.53 let us now compute $\Delta\alpha_{\mathrm{RRM}}$ and $\Delta\theta_{\mathrm{RRM}}$, the angles over which the aircraft can rotate without causing differences in the range profiles due to Rotational Range Migration. Table 4.2 shows the results. Recall that these angles are lower limits: for their computation, the aircraft length and the aircraft height, stowed landing gear, are taken. These are the maximum dimensions of the aircraft even though it is unlikely that radar scatterers are present on the outermost parts of the aircraft.

From the table we find that the aspect angle changes associated with Rotational Range Migration are minimally 0.35 and 1.3 degrees, obviously corresponding to the largest aircraft, the Boeing 747. These numbers are important for the choices of the aspect angles at which

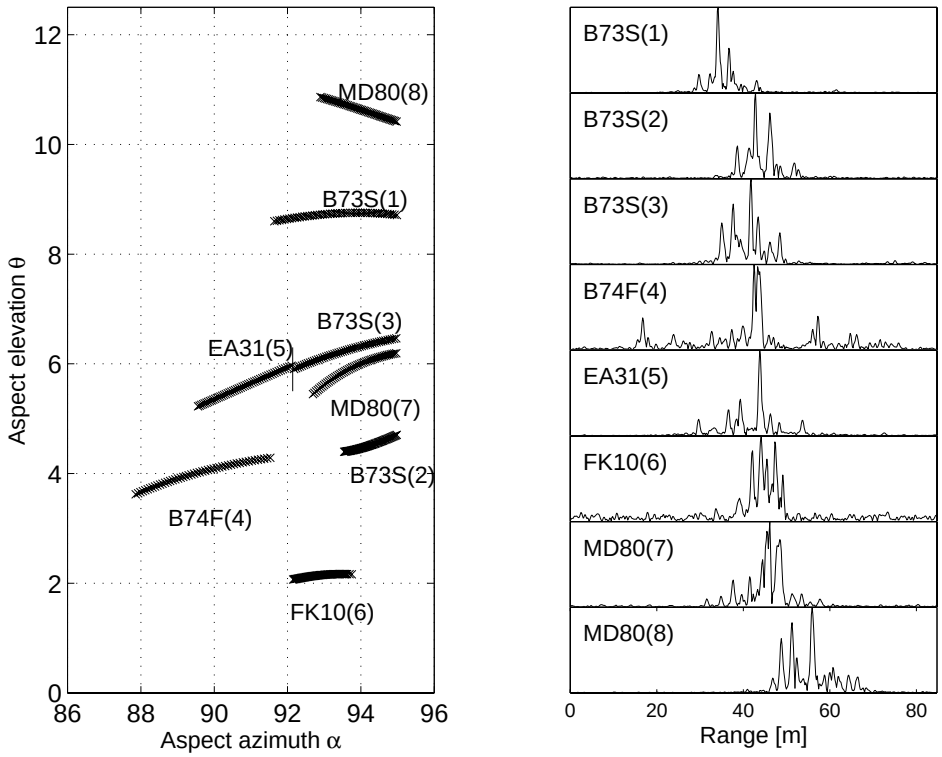

Figure 4.3: *Left: estimated aspect angles of measured range profiles. Each cross represents a range profile. The numbers right from the aircraft codes are the leg numbers. The figure on the right hand side shows for each leg the profile with the largest aspect azimuth (Data set III).*

Code	*Length [m]*	*Height [m]*	$\Delta\alpha_{\mathrm{RMM}}$ *[deg]*	$\Delta\theta_{\mathrm{RRM}}$ *[deg]*
B73S	31.0	11.3	0.79	2.2
B74F	70.7	19.3	0.35	1.3
EA31	46.7	15.8	0.53	1.6
FK10	35.5	8.5	0.69	2.9
MD80	45.5	7.7	0.54	3.2

Table 4.2: *Aspect angle changes in azimuth and elevation that can be exerted without the range profiles being different due to Rotational Range Migration.*

the synthetic range profiles will be calculated as they indicate at which angular resolution they should be sampled.

4.5 RCS-predictions

An important characteristic of radar range profiles is their strong dependence on aspect angle. It implies that we need an extensive database that contains profiles viewed at a very large number of aspect angles from all aircraft that we expect to see.

To this goal several trials, such as ORFEO, have been carried out that have provided fairly large amounts of data. However, countless more measurements should be carried out to give a *full* aspect angle coverage for all desired aircraft.

A possible way out is to use Radar Cross Section (RCS) prediction techniques. At TNO-FEL the numerical scheme RAPPORT is developed (Radar signature Analysis and Prediction by Physical Optics and Ray Tracing). Devised originally for computing the total amount of radar reflection of the object, it can also provide an amplitude and phase "measurement" of any aircraft model at any frequency and aspect angle. Thus, it can mimic a coherent radar used for the measurement of radar range profiles and 2D-ISAR images.

RAPPORT requires a geometrical description of the aircraft, in which the surface of the aircraft is approximated as a collection of facets. Such a description can, for example, be a Computer Aided Design (CAD) model of the aircraft, see for example Figure 4.4.

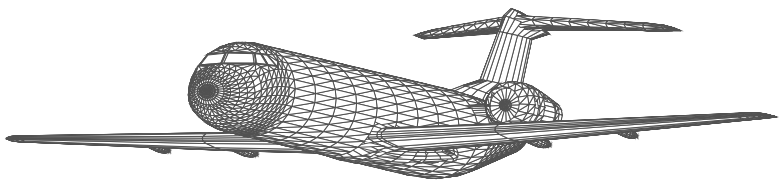

Figure 4.4: *CAD-model of Fokker 100. (Model made by Viewpoint Datalabs Intl, Orem, Utah, USA.)*

Given an accurate geometrical characterisation of the aircraft, good

knowledge of the scatter processes and sufficient computer power, a range profile database can be constructed at an arbitrarily dense coverage of aspect angle.

4.5.1 High-frequency RCS-predictions codes

RCS-prediction techniques are based on the following two principles.

- When an electromagnetic wave hits the surface of a perfectly conducting, infinitely large surface, it is reflected according to Snell's principle: the angle of incidence is the angle of reflection.
- When an electromagnetic wave hits the surface of a perfectly conducting surface of *finite* size, we must take into account the spatial radiation pattern caused by the induced surface currents. A fraction of the energy is still reflected according to Snell's law, but the rest is scattered to other directions [67]. The larger the surface, the more energy will be scattered following Snell's law.

In the radar community the way to deal with the first, specular, type of reflection is called the ray tracing approximation (Figure 4.5, right hand side, first reflection). The approach to computationally handle the second, diffuse, type of process is referred to as the Physical Optics (PO) approximation (Figure 4.5, left hand side). Looking with a radar at a certain object, ray tracing is used to determine which part of the object is visible to the radar. This gives the incident field on each of the individual flat surfaces, also called *facets*. The reflection on the facets is computed by PO and all the individual contributions towards the radar are summed. This is the procedure for a single reflection.

In order to find the reflected energy that will again hit the object, for each of the facets it should be computed in which directions the energy of the *full* spatial radiation pattern is scattered. Obviously, it requires a tremendous computational effort to compute in this way the result of two or even more reflections. It has been shown, however [45, 46, 79] that in the high frequency approximation (where the object or parts of the object are larger than typically five wavelengths) it is sufficiently accurate to use the PO approximation *only* for the computation of the reflected energy towards the radar. The contribution of

the energy to further reflections can be done sufficiently accurate by ray tracing alone. The assumption is that for object-to-object scattering the diffuse reflection may be replaced by a specular reflection. Additional reflections are done until there are no reflections left, or until the number of reflections exceeds a certain predefined limit.

The full procedure can be summarised as follows:

1. illuminate object from a certain direction;
2. find illuminated patches with ray tracing;
3. for each illuminated patch
   (a) compute returned energy in direction of radar with PO;
   (b) compute direction of reflected energy with ray tracing;
   (c) Goto 1 if this ray hits another part of the object and if the chosen maximum number of reflections is not reached.

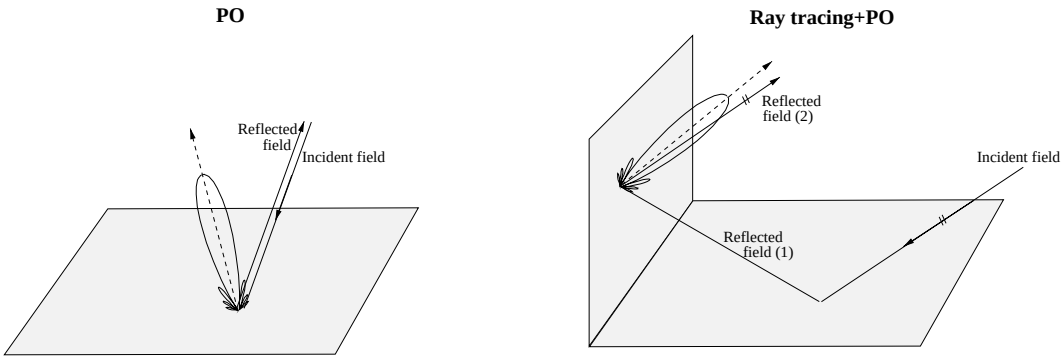

Figure 4.5: *Physical Optics approximation, left. The figure on the right hand side shows the contribution of RCS as a result of a double reflection. The field bounces on the horizontal plane, gives a contribution to RCS (as shown in the figure on the left), but is also reflected, according to ray tracing, to constitute the incident field for the second reflection. Again with PO, now applied to the second reflection, the contribution to RCS is computed.*

In most techniques the ray tracing implementation is based upon (a variant of) the *shooting and bouncing ray* (SBR) technique [53]: a

dense grid of rays is shot from the incident direction towards the target. Rays are traced as they bounce around the target. At the exit point of each ray, a ray tube integration is performed to sum up its contribution to the total scattered field. A well-known example of a such a code is XPATCH [3]. The SBR method has the disadvantage that a sampling density of ten rays per wavelength has to be used to obtain accurate results.

RAPPORT, the RCS-pediction program available for the research in this thesis, is similar to most other high-frequency electromagnetic scattering codes; see Brand [12] for a description. RAPPORT contains, however, a fundamental advantage compared to most other codes that lies in the ray tracing implementation: the illuminated area on the object is reconstructed explicitly with a certain accuracy, using a non-uniform or *backward* ray tracing algorithm. The backward ray towards the radar from a particular facet is constructed to check whether this facet is illuminated or blocked by another part of the object. Once the area is known for a certain aspect angle and object, the RCS can be calculated for any desired frequency. RAPPORT is computationally more efficient than SBR techniques as the ray density to obtain the same accuracy is far less.

The number of multiple reflections that has a significant contribution to the RCS is determined by taking an increasing number of reflections into account until the total RCS converges.

Note that several reflective processes are not included in RAPPORT: no edge diffraction is included nor are dielectric materials taken into account.

4.5.2 Geometrical models

The five models that are used in this thesis have been purchased at Viewpoint Datalabs Intl., Orem, Utah, USA[2]. Figure 4.6 shows a rendered representation of the five models at approximately broadside aspect angles. For each of these models the predictions are carried out using RAPPORT.

The object description for each object compares very well with the

[2]See http://www.viewpoint.com

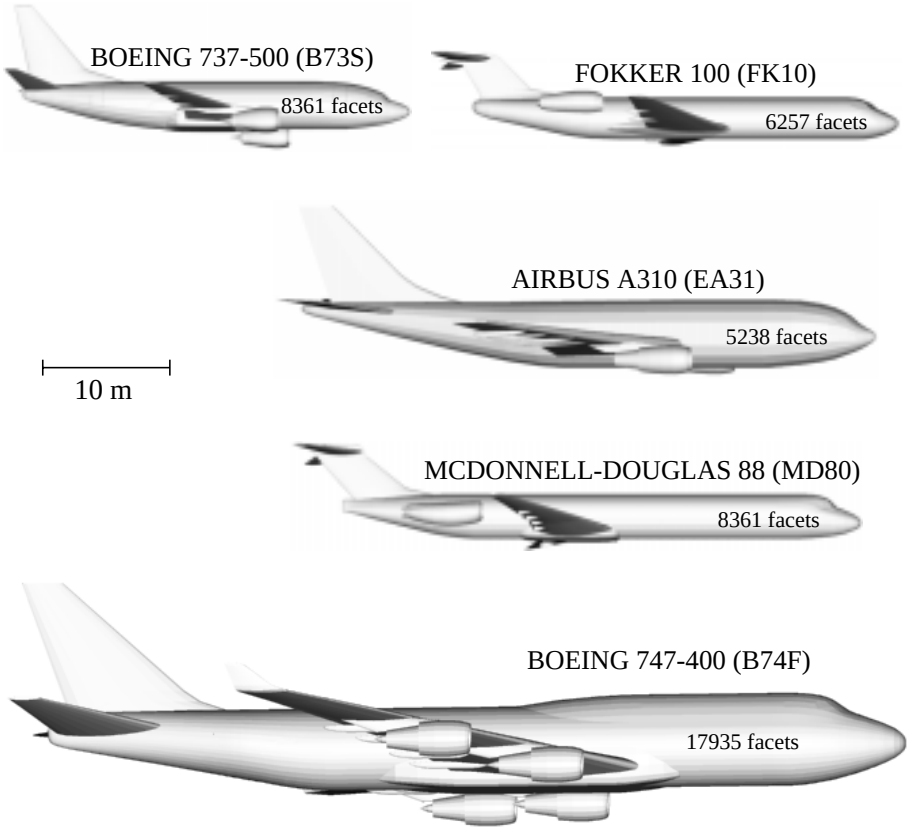

Figure 4.6: *The five models available for RCS-predictions. Above each model the name of the aircraft type is shown. Between brackets, the code-name as used by secondary radar systems is given.*

real object with respect to the external dimensions. Observed exceptions are:

1. *Engines*

   For all models the engines are closed near the front entrance.

   Also, there are no arrangements, nor in the model nor in the RCS-prediction code to produce contributions from the rotating parts in the engine.

   Note that this simplification will not influence results on range profiles computed at broadside aspect angles.

2. *Number of facets*

   The number of flat facets ranges from 5,238 for the Airbus to 17,935 for the Boeing 747. This demonstrates that the number of facets for the Airbus is relatively low. From the previous figure the rougher surface of the Airbus can actually be observed. Clearly, the predictions for the Airbus will therefore be less accurate than for the other aircraft types.

3. *Landing gear bay*

   The *bulge* (the part of the fuselage where the wings are attached and that contains the landing gear bay) of the Boeing 737-500 model is rather 'boxy'. It should in reality, however, have quite smooth edges.

   In Chapter 7, we use this model to compare the measured range profiles with the predictions. It shows that the comparison was very good on broadside aspect angles — apparently for the predictions this model imperfection is not crucial.

Note that for recognition on radar images of civil aircraft it is not sensible to model details that differ from aircraft to aircraft *from the same type*. Such efforts will not pay off, as the model has to be useful for a *class* of aircraft, not a particular example.

Incidentally, for fighter aircraft the requirements could however be different — radar images may look quite different for various configurations of the same aircraft. The configuration itself (weapons, external fuel tanks) could also be of interest for a threat evaluation.

4.5.3 Data set *IV*

We used RAPPORT for the prediction at exactly the same 324 frequencies as at which the ORFEO measurements in were performed.

For our experiments we compute profiles of the model of the Boeing 737-500 at each of the estimated aspect angles from Data set *II* (Figure 4.2). We thus mimicked the measurement of a stepped frequency waveform. The predicted radar data were processed in the same way as the real data, i.e. Hamming weighting, zero-padding, Fourier Transforming and taking the absolute values. We thus produced 150 synthetic range profiles. These will be used for a comparison between measurements and predictions.

To enable a further correlation experiment on leg 2 of this data set, we computed range profiles on a grid around the estimated aspect angles. As discussed earlier, we settle for a sampling in order to avoid Rotational Range Migration. We chose steps of 0.6 degrees in aspect azimuth and steps of 2.5 degrees in aspect elevation. See Figure 4.7.

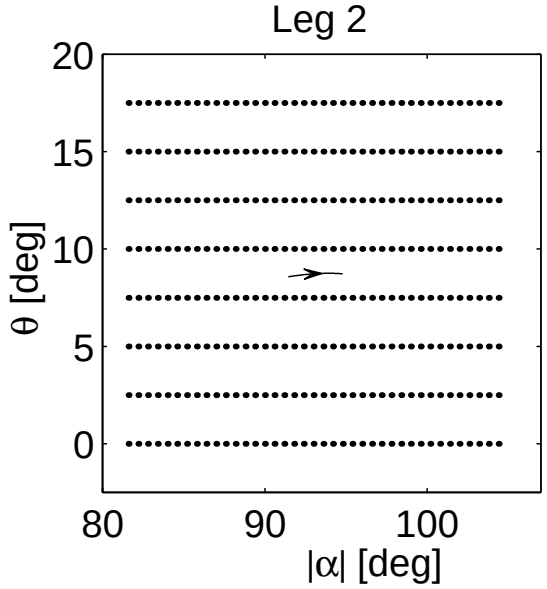

Figure 4.7: *Grid at which additional range profiles were computed (dots) around the aspect angles of leg 2 (solid).*

As an illustration of the nature of the data, we display in Figure 4.8 an image of a stack of computed range profiles viewing the Boeing 737-500 model at aspect azimuths of $0°, 1°, \ldots, 180°$, aspect elevation zero. It demonstrates the displacement of scatterers in the profiles as a function of aspect azimuth. E.g., the line that starts in the lower left corner is the aircraft nose. Until this scatterer becomes shadowed (at about 90°) it describes a $\cos(\alpha)$ path.

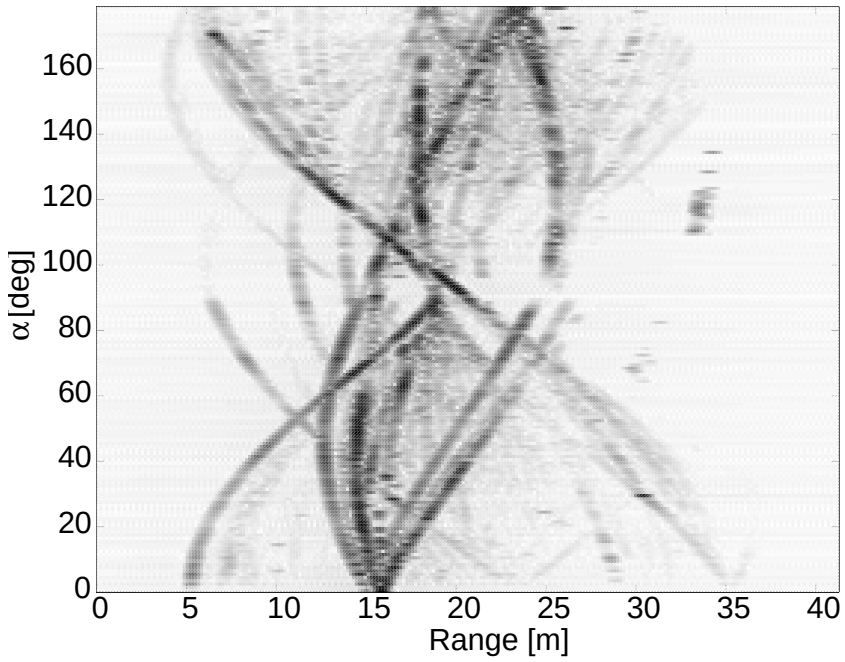

Figure 4.8: *Stack of computed range profiles from the Boeing 737-500 model for steps in aspect azimuth of* 1°*, zero elevation. The lowest amplitude is white, the largest black. The radar is situated on the left hand side.*

4.5.4 Data set V

For classification experiments we compute range profiles at a *grid* of aspect angles around the eight collected legs of the five different aircraft. The maximum number of reflections chosen in RAPPORT for

the calculation on these aircraft was three. For the Airbus, however, we needed to add another reflection before convergence was reached.

For the choice of a grid it is preferable to obey the limits given by the Rotational Range Migration, given in Table 4.2. We would thus like to produce a few range profiles per 0.35 degrees in azimuth and a few per 1.3 degrees in elevation. Given the number of computations for RCS-predictions and the number of classification experiments we want to perform, we have however settled for a lower sampling rate to limit the computational load: in the order of one RRM-sector in azimuth (0.4 degrees) and in the order of two RRM-sectors in elevation (2.5 degrees).

Even though the values for $\Delta\alpha_{\mathrm{RRM}}$ and $\Delta\theta_{\mathrm{RRM}}$ are lower estimates for the largest aircraft, it is possible that the classification results (Chapter 8) can improve at the cost of an increased computational load if the sampling is more dense. Figure 4.9 displays a number of PRPs of the Boeing 747-400, by far the largest aircraft in the data base. It shows that for a constant aspect elevation (figure at the left hand side), the profiles appear quite similar, showing that undersampling is not very important. For a sequence of range profiles at a constant aspect azimuth of 92 degrees, the profiles show much more variation, indicating that a more dense sampling in aspect elevation for this large aircraft is probably worthwhile. Additionally, shadowing effects can play an important role at aspect elevations near zero.

As we have assumed that the errors in both α and θ are smaller than five degrees, we do not have to compute predictions outside five degrees from any MRP. We have however further limited the number of computations by omitting the predictions at aspect elevations smaller than zero and larger than 12.5 degrees — it is unlikely that an aircraft has an aspect elevation outside these limits, as all legs are, in a good approximation, without curves. Concludingly, we have produced 215 range profiles per aircraft.

Figure 4.10 shows the aspect angles of each of the predictions that were computed.

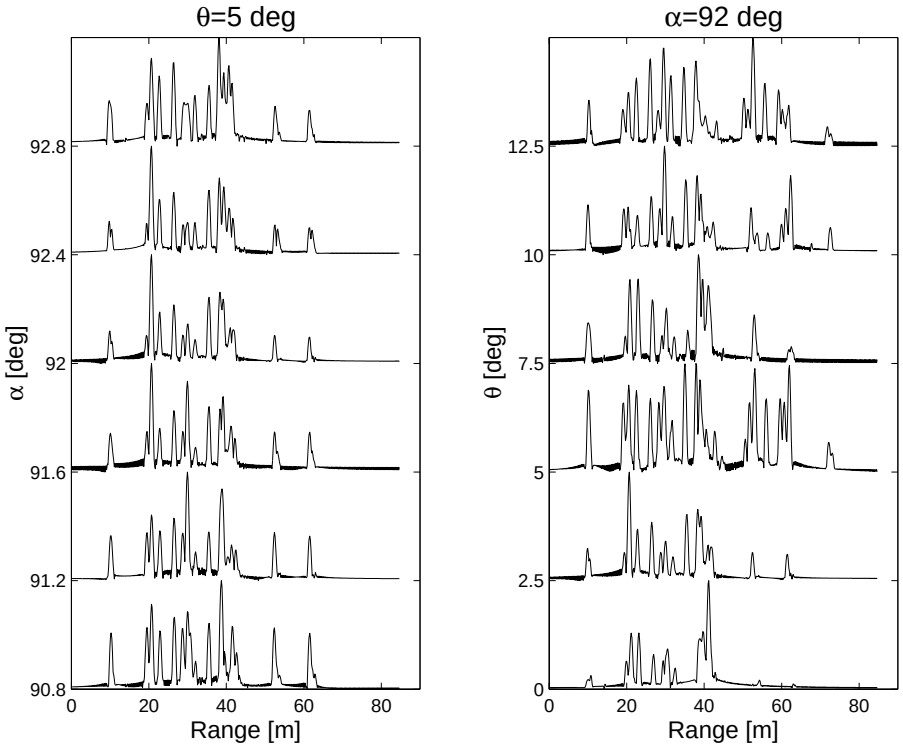

Figure 4.9: *Examples of predicted range profiles from the model of the Boeing 747-400. Left hand side: constant elevation, right hand side: constant azimuth.*

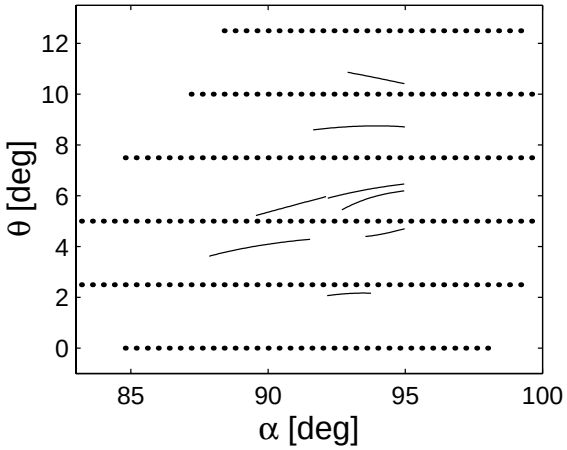

Figure 4.10: *Aspect angles of predicted range profiles for Data set* V*, each dot represents a range profile. For completeness, also the measured legs are drawn.*

4.6 Summary

In this chapter we have prepared the data that will be used in the classification and comparison experiments in the next four chapters. Table 4.3 provides an overview. We have elaborated on the criteria for data selection and have chosen to run our classifiers on fully independent training and test sets at the cost of a low number of classes. Subsequently, the background, the size and properties of the profile sets that contain measured data were described. Then, we reviewed the approach to produce synthetic range profiles through computational electromagnetics and aircraft models. We have sought to pragmatically limit the number of profiles using the information on the error in aspect angle and the knowledge on variability of range profiles through aspect angle variations.

set	*source*	*nature*	*purpose*	*# classes*	*# legs*	*# profiles*
I	S-band radar	measured	Classification	4	11	696
II	FELSTAR	measured	Comparison	1	3	150
III	FELSTAR	measured	Classification	5	8	400
IV	RAPPORT	synthetic	Comparison	1	3	1086
V	RAPPORT	synthetic	Classification	5	8	1075

Table 4.3: *Overview of the five data sets.*

Chapter 5

Classification of radar range profiles

5.1 Introduction

[1]The classifier forms the heart of an aircraft recognition system. It assigns a profile of unknown origin to an aircraft type in the database. Possibly, when the classifier is allowed to do so, also the declaration "not in data base" may be produced.

The field of pattern recognition, that hopefully provides a solution for our profile classification problem, covers a very wide research area. Although numerous techniques are available, they are based on a relatively small number of different concepts. Several properties of range profiles (e.g. , their variability and the shift-invariance they require) do not allow us to apply most of the pattern recognition techniques to our classification problem in a straightforward way — if applicable at all. This forced us to narrow down the number of techniques considerably culminating in the choice for methods that work only on similarities between range profiles. Two of them are based on the Nearest Neighbour rule, the other two are implementations of a Radial Basis Functions network.

We pay substantial attention to the comparison of classifiers. Clearly, an important criterion is the error on an independent test set. In view of future applications we believe that it is important to include the computational burden in the comparison as well.

The organisation of this chapter is as follows: the next section presents a general introduction on pattern recognition and relates that to the application of radar range profiles. Section 5.3 describes the used distance metric and the four classification techniques. Then, Section 5.4 establishes the approach to compare classification techniques. Section 5.5 shows the results. Discussion of the outcomes are covered by Sections 5.6 and 5.7, respectively.

[1]Parts of this chapter were published in [39, 36].

5.2 Pattern recognition view on radar range profiles

In the next subsection we start with a compact overview of the field of pattern recognition (PR) using the excellent book by Duda and Hart [22] as the main (in a few instances literal) source. We will then proceed with looking at radar range profiles from a PR point of view.

5.2.1 A bird's-eye view of pattern recognition

The commonly used, abstracted model for a PR system consists of three basic elements, a sensor, a feature extractor and a classifier. The sensor measures or "sees" the object and converts it to a form suitable for machine processing. The feature extractor selects relevant information from the sensor data. The classifier uses this information to assign the object to one of a number of categories.

Sensors could be a camera, a bar-code reader or a radar. The feature extractor reduces the data (an image, a binary pattern or a radar signature) by measuring properties that distinguish objects from different categories. The collection of such properties is stored in a *feature vector*[2]. Finally, the classifier partitions the feature space (in which the feature vectors are defined) into regions, each of which is assigned to a category. In the ideal case, this partioning could be done such that the regions do not overlap, and that none of the decisions is ever wrong. Generally, however, the best thing one can do is to choose the regions such that the average error is minimised, or, if one type of error is more costly than the other, to minimise the average cost.

The most fundamental approach to PR is through the Bayes decision theory. Application of this approach requires that the problem is fully posed in probabilistic terms. In its simplest form, where all classes have an equal *a priori* probability of occurrence, it proclaims that the unknown feature vector is assigned to that class that has, of all classes that are present, the highest probability to produce it. It can be rigourously shown that this assignment gives the lowest average

[2]For some applications it is beneficial to discard the feature selection and classify the raw samples directly, this is called *featureless* classification.

error probability possible. In a more general treatment, *a priori* probabilities and cost functions can be introduced. In the latter case the cost is minimised instead of the average error.

The requirement of availability of the full probabilistic knowledge on the problem is a very restrictive one. Only in very idealised cases such will be the situation. In the best case, we would have some vague, general knowledge about the probability density functions (PDFs) of the classes, together with a collection of samples from each class, the training samples. The problem is then to find a way to find the PDFs using this information.

The PDFs can be estimated in two different ways. Firstly, a functional form can be assumed with several parameters left to be estimated. The availability of the training samples allows us to find these estimates. This can be performed using the maximum likelihood approach, where the parameters are chosen such that the likelihood to produce the training samples is maximised. For example, if we assume that the samples are distributed normally with unknown mean, the maximum likelihood estimate of the mean can be found to be simply the arithmetic average of the samples. After we have found the parameters, and thus the PDFs, we can proceed with classification using the Bayes decision theory.

This first approach assumes that the PDFs are known or, at least, the functional forms of the PDFs are available. In many cases such assumptions are suspect as the functional forms rarely fit the densities actually encountered. So, a second way to find the PDFs is to fully estimate them from the training samples using a density estimator like, e.g. , *Parzen windows*. If we have confidence that this is done sufficiently accurate, the found PDFs can be inserted into a Bayes decision process.

A conceptually different idea is to forget about PDFs and Bayes decision theory, and use the training samples straight away for the classification of unknown samples. This can be done by using the *nearest neighbour* method that assigns an unknown sample to the class of the nearest sample in the training set. This method is sub-optimal; it will lead to an error generally greater than the Bayes error. However, with an unlimited number of samples, the error rate is never worse than twice the Bayes rate. This property, a simple technique with often reasonable

performance without foreknowledge, makes the nearest neighbour rule a very attractive method in itself but above all a convenient baseline for other methods.

Another avenue to avoid estimating the PDFs is the use of discriminant functions. This approach requires no assumptions about the PDFs, but instead of the functions that separate the classes. The parameters of these decision boundaries are estimated from the training samples. A famous example is the *Fisher Linear Discriminant* [25], where the parameters of a linear boundary between two classes are estimated from the training samples.

Essentially the same approach is followed in the use of neural networks for PR; the decision boundaries are however formed implicitly with a minimum of foreknowledge and can have (when the network is large enough) any shape. A neural network is basically a non-linear mapping from an input space (of arbitrary dimension) to an output space (of the same or lower dimension). It can be proven, given a sufficiently large network, that any non-linear continuous input-output relation can be constructed [47].

The mapping is found by examples from the input-output relation. Usually, these are offered repeatedly to the network — using a certain learning rule the (often huge amount of) parameters contained in the network hopefully converge.

The resulting network should be able not only to reproduce the input-output relation (to a degree depending on the noise on the samples and the application), but is also required to produce sensible outputs when a new input is offered: it should be able to *generalise*.

For PR these properties are beneficial. The network can be trained on the input-output relation that exists between feature vectors (input) and class labels (output). When a new feature vector is presented to the input, the network produces an output that is hopefully nearest to (the vector that corresponds to) the correct class label.

Numerous architectures of neural networks can be designed (see e.g. [40]) such as a feed-forward network (information flows in one direction through the network only) a recurrent network (contains a feedback loop) or a Radial Basis Function network (inputs are distances between samples).

After this general introduction on PR let us now explore the appli-

cation to range profiles.

5.2.2 Application on range profiles

Probably the first question a researcher asks himself when he considers the problem of range profile classification is: "Can I use a Bayesian classifier?" As mentioned before, if the statistical properties of the problem are known, this approach is optimal in the sense that it gives the lowest classification error. The second one might be on feature extraction: "How can I lower the dimensionality of the profiles without removing information that is discriminating between target types?". After all, classification problems become easier to deal with when the dimensionality of the features is low. Thirdly the issue "How can I classify these features?" could be raised. Finally, maybe he also considers featureless classification: "Can I do without feature selection and classify a profile as a whole?".

We will use this collection of issues as a guideline for a journey through the relevant range profile classification literature, ending with a justification of the approach we shall pursue and an account of the open-literature additions they represent.

Bayes decision theory

The way to attack the range profile classification problem with this technique is to use the aspect angles that correspond to the range profiles to assign each of the profiles to a Rotational Range Migration (RRM)-sector (recall that this is the angle interval in which no variability due to the migration of scatterers from one range cell to the other is expected, see page 48). Then the variations of the profiles due to speckle may be modelled by estimating the probability density functions (PDFs) of the range profile elements. Once these PDFs are available, a Bayes classifier could be used. Under special circumstances, the range profile elements have an exponential distribution. For this to be true, a range resolution cell should contain a large number of independent scatterers so that the central limit theorem holds. Then the resulting distributions of both the real and imaginary parts of the complex-valued sum of scatter contributions are normal. Hence the energy has an exponential distribution and a maximum likelihood approach can be used to find

its parameter. (Note that we have touched upon the statistics of range profile elements earlier, in Chapter 3, page 46.)

However, a cell hardly ever contains three or more scatterers and they are not independent as they are situated on a rigid body. For this general case, we do not have a model for the distribution. It means that we cannot estimate the PDFs parametrically, but we have to fully estimate the PDF.

In practice, such is not feasible either because accurate estimates of the aspect angles would have to be available to assign profiles to a certain RRM-sector. Unfortunately, the error in aspect angle is generally much larger than the size of an RRM-sector. Even if the aspect angle estimate would be accurate, only a limited number of range profiles per RRM-sector are present which makes the estimate of the PDFs a precarious procedure. Even though a few authors present some initiating work on Bayes-classification [24, 64], the problems we envisage in finding the PDFs make it impossible to follow this route.

Feature extraction and their classification

Feature extraction is about selecting (or computing) low-dimensional characteristics from samples that are maximally discriminative between different classes. This is done to make the problem more easy to deal with: less computation time and storage capacity is needed. Also, however, it is known that with a constant amount of training samples, the classification performance detoriates when the number of features increases: this effect is known as the "curse of dimensionality".

An important parameter in feature extraction is the *intrinsic dimension*: essentially it is the dimension of the subspace in which all variability occurs. The intrinsic dimension of Ku-band range profiles (radar frequency $\approx$ 18 GHz) was estimated using the Karhunen-Loève technique, also known as *principal component analysis* by Leushacke [50]. He finds a value of $10, \ldots, 30$

Spectral components of range profiles can be used as features. This has been described, in several variants, by Garber *et al* [27], DeWitt [21] and Kouba [48]. Classification of the feature vectors is carried out by a nearest neighbour rule, a *Hidden Markov Model* and a recurrent neural network, respectively. The latter two also have the interesting ability to process *legs* of range profiles.

Another feature extraction technique uses *wavelets*. Range profiles are represented at different resolution levels and the classification is done hierarchically: progressing from gross characteristics to finer details, more candidate classes to match an unknown profile are eliminated until one class remains. Such an approach was reported by Baras and Wolk on range profiles from ships [7].

Zyweck and Bogner [80] use data of two different aircraft, align the range profiles by a correlation technique and find the optimal *one*-dimensional subspace using a Fisher linear discriminant [25].

Close inspection of the highly variable range profile data did not give us hope for the selection of robust, discriminative features based on *a priori* knowledge. The problems we expect using feature selection is to fulfil the required shift invariance and to select sufficiently stable features despite the aspect angle dependency. This view was supported by a lack of literature on feature extraction on aircraft radar range profiles. It made us decide to investigate in this dissertation approaches other than those based on feature extraction.

It is interesting to see, therefore, that the feature extraction in the given references where aircraft range profiles are concerned, apart from the last one, is done on either scale-model measurements or computer simulations: in such cases shift-invariance is not required as the position of the aircraft in the profile is known exactly. This, however, does not mean that we fully abandon the extraction and classification of profile features for future investigation.

Featureless classification

It is attractive to bypass feature extraction and avoid the "curse of dimensionality" by classifying the profiles based on a similarity measure. A nearest neighbour classifier or a Radial Basis Function network are then a few of the options to assign samples to one of the categories. In featureless classification the class declarations are made on distances between samples only. Recently, this field has received renewed interest through the use of so-called *support vectors* [18, 76, 23].

A suitable distance metric is easily found: Novak [58] uses a metric that is shift invariant. We will call it the *Sliding Euclidean Distance*. The classification is done with a simple *nearest neighbour* rule. Similarly, Li *et al* [51] use the *matching score* to classify computed range

profiles — this measure is the maximum value of the correlation function between two range profiles.

Another solution to accomplish shift invariance is the use of *correlation filters*, presented by Hudson and Psaltis [41]. This paper shows results on a large number of range profiles from in-flight aircraft. As mentioned in the previous chapter, the use of the data is slightly incestuous: data from the same legs are used for input and testing probably giving a too optimistic estimate of the classification rate. The results are that foreknowledge of aspect angle makes the error rate drop from 43% to 21 % (bins of 20°). Using a majority vote on a sequence of eight profiles, made the average classification error drop to 16%. A majority vote on 48 profiles gave a perfect score.

Our approach and contributions

Because of the robustness and ease of computation of a distance metric, we have chosen to discard feature extraction and to classify the range profiles using featureless techniques. Instead of finding proper features, we now have come up with an appropriate similarity measure: much care is taken to find one. The manipulation of the distances to arrive at a class declaration is done in four different ways: two statistical and two neural procedures. One of the methods we shall employ is a nearest neighbour technique; additionally we shall apply a technique to reduce the size of the training set.

The reason for choosing the nearest neighbour techniques for our purpose is that it is the simplest classifier that operates on similarities between samples. Additionally, the achievements for large training set sizes give an indication for the best possible (Bayesian) performance which allows us to draw conclusions on the usefulness of range profiles for aircraft recognition.

Next to a nearest neighbour classifier we decided to design a neural classifier to investigate whether the alleged generalisation properties were also beneficial for our classification problem. A difficulty was to make the network suitable for the profiles, again to achieve shift-invariance. A good solution was found by using Radial Basis Functions: in the approach that is chosen most often, the only inputs are distances between pairs of profiles (see, e.g. , [40], pages 248-250 and references therein). We can hence employ the distance metric that is also used

for the nearest neighbour technique.

This chapter provides a *comparison* between these techniques, not only on classification error, but also on classification *effort*. Also, an attempt is made to assign the most appropriate classifier to combinations of four applications and two scenarios.

5.3 Range profile classification techniques

5.3.1 Notation

As stated earlier, we assume we have available two sets of independently measured range profiles: the *input set* and the *test set*. A subset of the input set is used for training, and is obviously called the *training* set. We have the latter subset to study the trade-offs between classifier size, accuracy and effort — these are found by varying the size of this set.

The profiles in the input, training, and test set are randomly ordered and are named $\boldsymbol{r}_i, i = 1, \ldots, N_{\text{input}}$, $\boldsymbol{p}_i, i = 1, \ldots, N_{\text{train}}$ and $\boldsymbol{q}_i, i = 1, \ldots, N_{\text{test}}$ respectively (see Figure 5.1). Boldface symbols denote vectors; where we use $\boldsymbol{x}_i$ for a generic range profile, it can be substituted by either $\boldsymbol{r}_i$, $\boldsymbol{p}_i$ or $\boldsymbol{q}_i$.

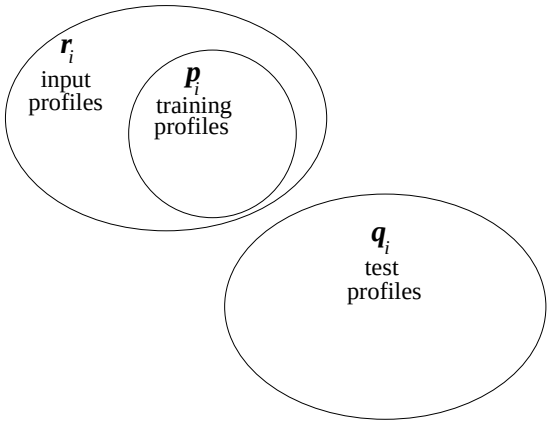

Figure 5.1: *Two given sets of range profiles. The input set and the test set. A subset of the input set is used for training a classifier.*

The profile elements of $\boldsymbol{x}_i$ are $x_i(1), \ldots, x_i(d)$ — all profiles can thus

be viewed as vectors in an d-dimensional space. For each $\boldsymbol{r}_i$ the class $C(\boldsymbol{r}_i)$ is known. C can take the values $1, 2, 3, \ldots, N_C$ only, where N_C is the total number of classes.

A *classifier* is fully determined by

1. the *classification technique* and
2. the *training set.*

Note that in item 2, not only the contents, but also the *order* of the profiles in the training set will (marginally, if well-trained on sufficient data) influence the classifier[3].

Suppose we construct a classifier with some classification technique X using the training set $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_{N_{\text{train}}}$. Then we denote the result of the classification with $X$of profile $\boldsymbol{q}_i$ by $C_{X,\boldsymbol{p}_1,\ldots,\boldsymbol{p}_{N_{\text{train}}}}(\boldsymbol{q}_i)$.

The true classes $C(\boldsymbol{q}_i)$ of the test profiles are also known. This enables us to assess the performance of a classifier. Clearly, in an operational stage a test set is not available.

5.3.2 Sliding Euclidean Distance

The absolute position of the aircraft response in the profile depends on the target range. As we do not know this accurately enough to place the scatterers on an objective position, we have to use shift-invariant classification techniques. They are all based on distances between profiles. If we apply a shift invariant distance metric, then classification will also be shift-invariant.

For two d-dimensional range profiles $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ we define the distance D as the minimum Euclidean distance over all shifts:

$$D(\boldsymbol{x}_1, \boldsymbol{x}_2) \equiv \min_{j=0,\ldots,d-1} \sqrt{\sum_{i=1}^{d} [x_1(i+j) \Leftrightarrow x_2(i)]^2}$$

[3]Strictly speaking, for some classifiers other than the ones that will be considered in this dissertation also the choice of a random seed may influence the classifier. E.g., when training a feed-forward neural network, the training samples are fed in random order to the network. If such a classifier is well-trained on sufficient data, however, the random seed should hardly influence the performance.

The shifts are cyclical (modulus d), that is $x_1(d+j) \equiv x_1(j)$.

If d can be factorised in small prime numbers (preferably factors of two), the use of a forward- and backward Fast Fourier Transform gives a considerable decrease of the computational load to calculate D.

5.3.3 Compression and normalisation of profiles

In Chapter 6 it will be shown that a compression or nonlinear scaling of the profiles with a power function is advantageous for the classification result. Here we have a chosen a fixed power of $\eta = 1/4$, in the next chapter we will further investigate this, so-called, *Box-Cox transformation*.

We also normalise the profiles to make the classification independent of the magnitudes of the range profile elements and thus insensitive to absolute radar calibration.

If $\boldsymbol{x}$ is the original profile with elements $x(1), \ldots, x(d)$, (which have the dimension of energy) the elements of the resulting compressed and normalised profile $\boldsymbol{x}'$ are:

$$x'(i) = \frac{x^{\eta}(i)}{\sqrt{\sum_{j=1}^{d} x^{2\eta}(j)}}, \text{ for } i = 1, \ldots, d \tag{5.1}$$

For the remainder of this chapter we omit the prime " $'$ "- it is tacitly assumed that the profiles are compressed and normalised.

5.3.4 Nearest Neighbour Algorithm

The Nearest Neighbour (NN) classification rule [22] assigns a test sample $\boldsymbol{q}$ to the class of its nearest neighbour in a set of training samples whose classes are known. Formally stated, it decides that the class $C_{NN,\boldsymbol{p}_1,\ldots,\boldsymbol{p}_{N_{\text{train}}}}(\boldsymbol{q})$ equals $C(\boldsymbol{p}_n)$ if

$$D(\boldsymbol{p}_i, \boldsymbol{q}) > D(\boldsymbol{p}_n, \boldsymbol{q}), \quad \text{for} \quad i = 1, \ldots, N_{\text{train}}, i \neq n \tag{5.2}$$

A straightforward extension is to search for $k > 1$ nearest neighbours, giving k class declarations. The class that is declared most often is assigned to the profile from the test set. In Chapter 8 we will further

extend the NN-rule by utilising the difference between the number of votes of the most favourite class, and the number of votes of the second-most favourite class. For now, however, we will use the single, i.e. $k = 1$, nearest-neighbour rule.

5.3.5 Condensed Nearest Neighbour

A drawback of the nearest neighbour technique is the large computational effort necessary for the classification. For each $\boldsymbol{q}_i$ for which classification is desired, we have to compute all distances to the profiles in the training set. This is even more problematic in our application, because the chosen distance measure is computationally expensive.

The technique we use to reduce the computational burden is based on the notion that a profile in the training set that is far from the decision boundary (the border between two classes) has, on average, far less influence on the outcome of the nearest neighbour classifier than a profile that is near the decision boundary. Therefore we might as well remove this profile and save the computation time.

In this way, we can limit the number of distances that has to be evaluated for a single classification. As the largest part of the total computation time for one classification is used for calculating the distances, it may thus bring down the total classification time considerably.

Generally, removing profiles also reduces the classification accuracy. This effect becomes stronger if the classification problem has a higher dimension. In case of the range profiles used in this chapter (dimension $d = 128$) the space is almost empty even if a fair amount of training profiles is available. In such a rarefied space removing a profile seldomly leaves the decision boundary intact.

We use the *condensing* algorithm given by Devijver and Kittler [20]. To arrive at the condensed subset of the training set, two complementary subsets of this set, named A and B, are defined. Place the first profile from the training set, $\boldsymbol{p}_1$, in A, the remaining profiles, $\boldsymbol{p}_2, \ldots, \boldsymbol{p}_{N_{\text{train}}}$, in B. The method proceeds as follows:

1. Use the nearest neighbour rule to classify the first profile in B using only the profile(s) in A as the training set. If it is classified correctly with the nearest neighbour rule, leave it in B, otherwise,

place it in A. Repeat this operation for all profiles that are left in B.

2. If not a single profile has been transferred from B to A in step 1, or if B is empty then terminate. Else return to step 1.

After termination, A contains a condensed subset, $\boldsymbol{p}_i^{\text{cond}}$, $i = 1, \ldots,$ N_{cond}, of the training set. See also Figure 5.2.

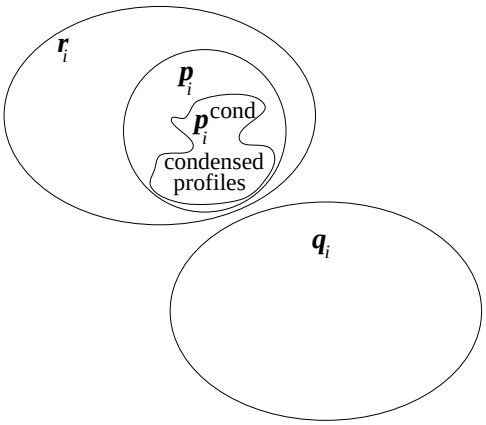

Figure 5.2: *A subset from the training set is the* condensed *set.*

The class of profile $\boldsymbol{q}$ determined by the condensed nearest neighbour technique is denoted by $C_{CNN,\boldsymbol{p}_1,\ldots,\boldsymbol{p}_{N_{\text{train}}}}(\boldsymbol{q})$. It decides that the class is $C(\boldsymbol{p}_n^{\text{cond}})$ if

$$D(\boldsymbol{p}_i^{\text{cond}}, \boldsymbol{q}) > D(\boldsymbol{p}_n^{\text{cond}}, \boldsymbol{q}), \tag{5.3}$$

for $i = 1, \ldots, N_{\text{cond}}, i \neq n$.

5.3.6 Radial Basis Functions

Radial Basis Functions (RBF) provide a way to construct a function that maps vectors from a high dimensional space onto a lower dimensional space [13]. As, for our application, the only inputs for this method are sample-to-sample distances, we can use the sliding Euclidean distance D between two range profiles to make the method

suitable for our classification purposes. The advantage of the RBF implementations compared to a nearest neighbour technique is that it generalises the training data providing a far more efficient classifier in terms of memory usage and classification speed.

The aim is to find a function $\boldsymbol{S}$ that maps the profiles $\boldsymbol{x}_i$ onto output vectors that depend on the class of the profile only. They are chosen as unit vectors in an N_C-dimensional space where the position of the non-zero element in the vector defines the class. That is, profiles that belong to class 1 are mapped onto output vector $\boldsymbol{e}_1 = (1, 0, 0, \ldots, 0)^T$, profiles that belong to class 2 are mapped onto output vector $\boldsymbol{e}_2 = (0, 1, 0, \ldots, 0)^T$, *et cetera*. In mathematical terms we write

$$\boldsymbol{e}_{C(\boldsymbol{x}_i)} \equiv (\delta_{1,C(\boldsymbol{x}_i)}, \delta_{2,C(\boldsymbol{x}_i)}, \ldots, \delta_{N_C,C(\boldsymbol{x}_i)}) \tag{5.4}$$

where δ_i is the Kronecker delta function:

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{5.5}$$

Once we have found $\boldsymbol{S}$ using $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_{N_{\text{train}}}$, we can classify a profile $\boldsymbol{q}_i$ from the test set by computing the output $\boldsymbol{S}(\boldsymbol{q}_i)$. We decide that the class corresponding to the class vector with the smallest Euclidean distance to this output is the class of $\boldsymbol{q}_i$.

The task is thus to find a function $\boldsymbol{S}$ of multidimensional mapping from an d-dimensional input space to an N_C-dimensional output space

$$\boldsymbol{S} : \mathbf{R}^d \to \mathbf{R}^{N_C} \qquad (N_C < d) \tag{5.6}$$

that satisfies

$$\boldsymbol{S}(\boldsymbol{p}_i) = \boldsymbol{e}_{C(\boldsymbol{p}_i)}, \text{for } i = 1, \ldots, N_{\text{train}} \tag{5.7}$$

In the commonly used Radial Basis Functions approach, $\boldsymbol{S}$ is restricted to a weighted superposition of functions that depend only on the distances $D(\boldsymbol{x}_i, \boldsymbol{c}_j)$ of the profile $\boldsymbol{x}_i$ to K fixed reference profiles $\boldsymbol{c}_j$. The latter profiles are called the *centres*. Hence the classifier does *not* act on the profiles directly but solely on the simularities between profiles. It thus allows us to exploit the robustness of featureless classification and, at the same time, to use neural techniques to adapt to the data. Although it is possible to find optimal centres in a training

stage [55, 60] or have the network configured using *genetic algorithms* [10], we will always choose them as a subset of the training set as in [13]. At this point we assume that the centres are known, the next two subsections describe the used selection methods. We may thus write for our training profiles:

$$\boldsymbol{S}(\boldsymbol{p}_i) = \boldsymbol{w}_0 + \sum_{j=1}^{K} \boldsymbol{w}_j \ \phi(D(\boldsymbol{p}_i, \boldsymbol{c}_j)) \tag{5.8}$$

The N_C-dimensional column vectors $\boldsymbol{w}_i$ are the free parameters — they are used to adapt $\boldsymbol{S}$ to represent the profiles. The functional form of the *kernel* ϕ is not crucial. A Gaussian kernel

$$\phi(D) = e^{-\frac{D^2}{2}} \tag{5.9}$$

proves to have good approximation capabilities, although also other functions could be chosen [17, 61]. Figure 5.3 shows the architecture of an RBF network.

For computing the weights $\boldsymbol{w}_i$ we combine equations 5.7 and 5.8 and find for each profile $\boldsymbol{p}_i$:

$$\boldsymbol{e}_{C(\boldsymbol{p}_i)} = \boldsymbol{w}_0 + \sum_{j=1}^{K} \boldsymbol{w}_j \ \phi(D(\boldsymbol{p}_i, \boldsymbol{c}_j)) \tag{5.10}$$

For N profiles a set of linear equations results:

$$[\boldsymbol{e}_{C(\boldsymbol{p}_1)}, \ldots, \boldsymbol{e}_{C(\boldsymbol{p}_N)}] = [\boldsymbol{w}_0, \boldsymbol{w}_1, \ldots, \boldsymbol{w}_K] \ \ \Phi \tag{5.11}$$

Where Φ equals:

$$\begin{pmatrix} 1 & \ldots & 1 \\ \phi(D(\boldsymbol{p}_1, \boldsymbol{c}_1)) & & \phi(D(\boldsymbol{p}_N, \boldsymbol{c}_1)) \\ \vdots & & \vdots \\ \phi(D(\boldsymbol{p}_1, \boldsymbol{c}_K)) & \ldots & \phi(D(\boldsymbol{p}_N, \boldsymbol{c}_K)) \end{pmatrix} \tag{5.12}$$

In matrix notation equation 5.11 becomes:

$$E = W\Phi \tag{5.13}$$

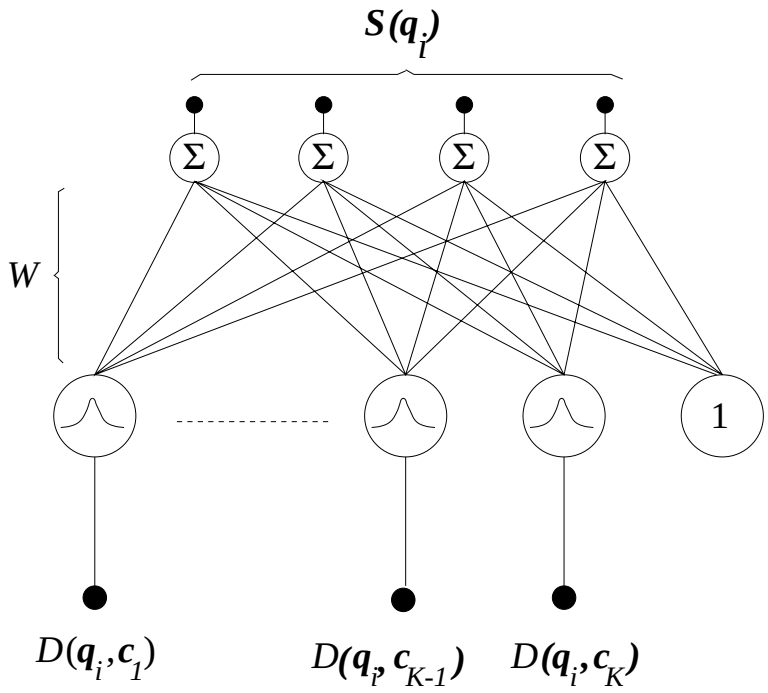

Figure 5.3: *Architecture of a four-class Radial Basis Function network. At the input units (black dots at the bottom) the distances between an unknown profile* $\boldsymbol{q}_i$ *and* K *centres enter the network. After a nonlinear transform, the distances and the bias are multiplied by a weight vector* W*. The four outputs form the vector* $\boldsymbol{S}(\boldsymbol{q}_i)$*.*

From these equations we see that each extra centre introduces a row in the matrix W. The number of degrees of freedom is $K+1$ with K the number of centres. The bias term gives the extra degree of freedom.

Now three cases can be distinguished:

$N < K+1$ The linear system is underdetermined and no unique solution exists.

$N = K+1$ All but one training profiles are used as centres. An exact solution can be found, and the equality sign of Equation 5.8 holds. Powell [61] has shown that in all practical cases the matrix Φ is non-singular.

$N > K+1$ The set of equations is overdetermined and we have to solve the system in a least squares sense [13]. The solution is

$$W = E\Phi^{+} \tag{5.14}$$

where Φ^{+} is the Moore-Penrose pseudo inverse [28]. This is the case in practically all real-world applications.

The latter case shows an attractive property of the Radial Basis Functions approach: although it is able to construct any nonlinear decision boundary, the weights can be found by solving a set of linear equations. The following two subsections deal with the selection of the centres.

5.3.7 Radial Basis Functions with Random Centre Selection

A good first choice for the centres is to select them randomly from the training profiles. One must be careful, however, about the number of centres to choose. Each extra centre adds an extra degree of freedom to fit the training profiles. If we take too few centres, the approximation will be too coarse. If we take too many centres (but less than N) also the noise on the profiles will be fitted ("overfitting"). In both cases, the generalisation capabilities of the classifier will be worse than with an intermediate number of centres.

To find the optimum number of centres we devised the following algorithm. Use the first half of the training set, $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_{N_{\text{train}}/2}$, as *design* set and the remaining profiles as *evaluation* set, see also Figure 5.4.

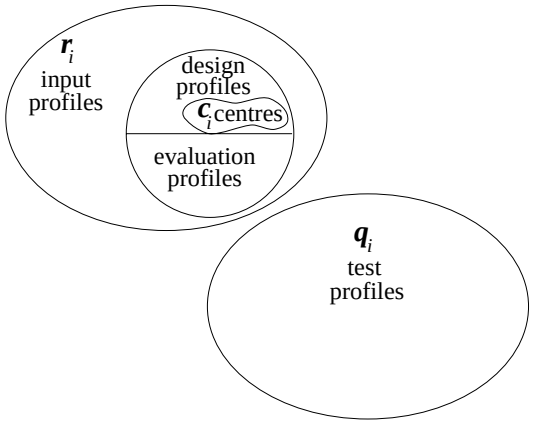

Figure 5.4: *Use of data in the Radial Basis Function approach. The training set is divided into an evaluation set and a design set. The design profiles are used to construct the function $\boldsymbol{S}$, the evaluation profiles are used to find the optimum number of centres. A subset of the design set is chosen as centre set.*

Do the following for $k = 1, \ldots, N_{\text{train}}/2 \Leftrightarrow 1$:

1. Use the profiles $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_k$ as centres[4]. Compute the weights W (eq. 5.14) and thus $\boldsymbol{S}_k$ with design set $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_{N_{\text{train}}/2}$.

2. Compute the outputs $\boldsymbol{S}_k(\boldsymbol{p}_i)$ for all profiles in the evaluation set ($i = N_{\text{train}}/2 + 1, \ldots, N_{\text{train}}$). Calculate the Euclidean distance between the resulting outputs and the class vectors corresponding to the actual classes. Sum these distances over all profiles in the evaluation set to find the total evaluation error ε_k:

$$\varepsilon_k = \sum_{i=N_{\text{train}}/2+1}^{N_{\text{train}}} ||\boldsymbol{S}_k(\boldsymbol{p}_i) \Leftrightarrow e_{C(\boldsymbol{p}_i)}|| \tag{5.15}$$

[4]Recall that the profiles in the training set are randomly ordered, (5.3.1, page 86). If we take the first k profiles, we therefore have a random profile choice.

It is instructive to compute the approximation error for the design profiles as well:

$$\delta_k = \sum_{i=1}^{N_{\text{train}}/2} ||\boldsymbol{S}_k(\boldsymbol{p}_i) \Leftrightarrow e_{C(\boldsymbol{p}_i)}|| \tag{5.16}$$

Although δ_l decreases monotonically as the number of centres increases, the evaluation error will reach a minimum for a certain number of centres such as shown in Figure 5.5.

The best number of centres is therefore at the $K = l$ at which ε_l reaches a minimum.

The classification function $C_{RR,\boldsymbol{p}_1,\ldots,\boldsymbol{p}_{N_{\text{train}}}}(\boldsymbol{q}_i)$ denotes the classification of profile $\boldsymbol{q}_i$ using the Radial basis functions with Random centre selection.

5.3.8 Radial Basis Functions with Gram-Schmidt Centre Selection

To find the most efficient centres from the design set a Gram-Schmidt Orthonormalisation technique can be used. It is based on the observation that each potential centre defines a row in the matrix Φ (expression 5.12) and can thus be viewed as a regressor that spans the output space. Now find the design profiles that, if used as centres, span the output space as efficient as possible.

As in the Random Centre Selection, the method is to use the first halve of the training set for designing the classifier and the other part for evaluation. Then we follow the technique devised by Chen *et al* [17] (for details we refer to the article):

> Start with a Radial Basis Function network without any centres. At each step, add this specific profile to the centre set that gives the best improvement of the least-squares approximation (eq. 5.16). Stop when $K = N_{\text{train}}/2 \Leftrightarrow 1$. For an efficient search, use a Gram-Schmidt orthonormalisation procedure.

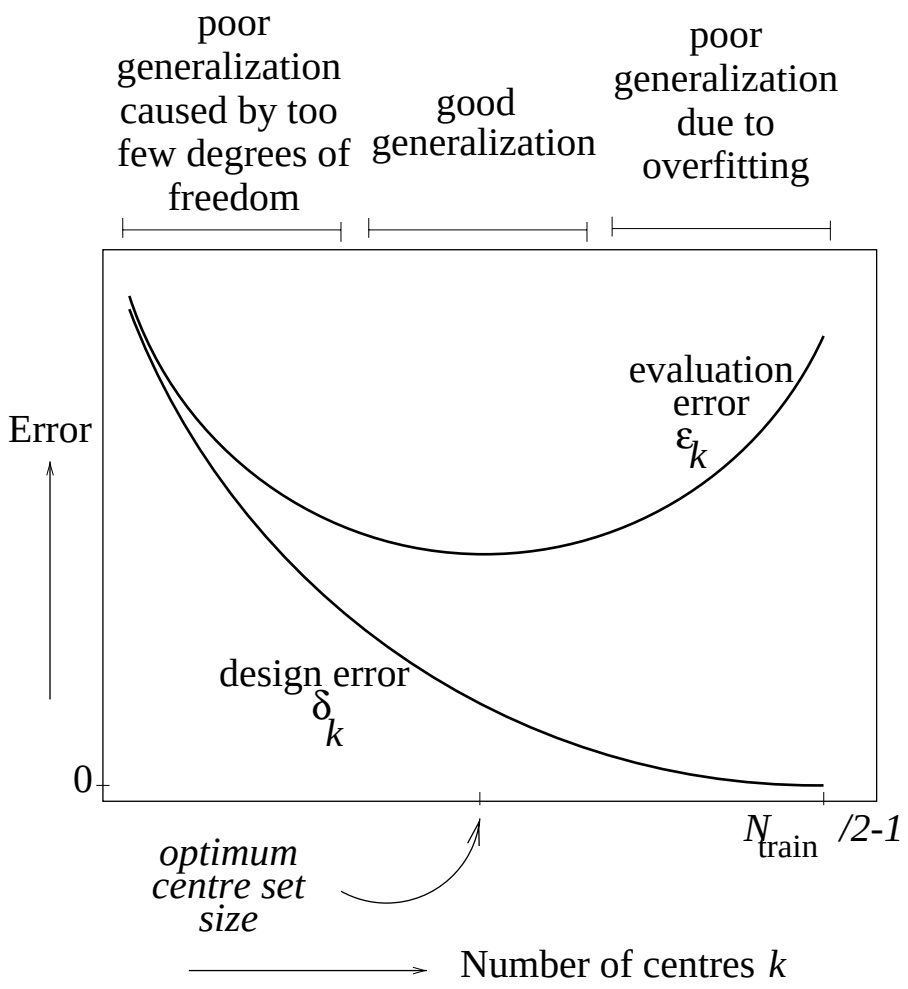

Figure 5.5: *As the number of centres increases the approximation function* $\boldsymbol{S}$ *is able to represent the design profiles better. In the extreme case,* $k = N_{\text{train}}/2 \Leftrightarrow 1$*, the representation error* δ_k *is zero. The true classification capability, however, is revealed by the error on an independent evaluation set* (ε_k)*.*

As in the random centre selection method, we also compute the classification error on the independent evaluation set and choose that number of centres where the evaluation error is lowest.

The classification function $C_{RGS,\boldsymbol{p}_1,\ldots,\boldsymbol{p}_{N_{\text{train}}}}(\boldsymbol{q}_i)$ denotes the classification of profile $\boldsymbol{q}_i$ using the Radial Basis Functions with Gram-Schmidt centre selection.

5.4 Comparison of classifiers

Often, classification techniques are compared in terms of the number of false classifications on a test set only. For most practical applications three more properties determine the usefulness of a classifier. Properties of interest are:

a_1 Classification error [% of false classifications on independent test set]

a_2 Computational effort needed for one classification [# floating point operations]

a_3 Computational effort needed for training the classifier [# floating point operations]

a_4 Memory capacity required to store the trained classifier [# of bytes].

In practice, for each classification technique there will be a trade-off between these four properties depending on application and scenario.

For example, let us consider a classifier that uses the nearest neighbour technique ($a_3 \equiv 0$) and a small sized training set. Then the classification error, a_1, will decrease if the training set increases. This also implies that more distances have to be computed and it thus results in larger values of a_2 and a_4.

To choose the proper classifier we would like to have weight functions w_i that express the cost of the properties a_i, so that the *cost function*

$$\sum_{i=1}^{4} \omega_i(a_i) \tag{5.17}$$

	scenario	*crisis*		*wartime*	
	classifier property	*correct class.*	*fast class.*	*correct class.*	*fast class.*
application	SHORAD	+	0	0	+
	HIMAD	+	−	0	0
	Fighter aircraft	+	0	0	+
	Surveillance	+	−	0	−

Table 5.1: *Relative importance of classification properties for application and scenario in terms of minus signs (less importance), zeroes (moderate importance) and plus signs (high importance). Here SHORAD means SHOrt Range Air Defense and HIMAD HIgh to Medium Air Defense (e.g. HAWK, PATRIOT).*

is minimised with respect to $a_1, \ldots, a_4$. Unfortunately, we do not have these functions available, but we can make a few simplifying but realistic assumptions to tackle the problem. The first one is that the most important parameters in most applications are a_1 and a_2. The time needed for training (a_3) is of much lesser importance, because it can be done off-line. The size of the trained classifier is generally also less significant because memory storage is cheap and very compact. Besides that, a_4 is (almost) linearly related to a_2 for the classification techniques we consider. Therefore we do not have to minimise a_4 by itself.

We do not make a choice for ω_1 and ω_2 either, but construct and test a large number of classifiers that each produce a combination of a_1 and a_2. From these results the classifiers that give the best combination given application and scenario can be selected.

At this point, we also want to stress that not only the application (e.g. surveillance or aircraft radar) is decisive for the classifier choice, but in a military context also the scenario (crisis or wartime). As an illustration, Table 5.1 shows roughly the importance of correct classification and fast classificaton as a function of the application and the scenario.

This table shows in qualitative terms that in times of crisis it is more important to have a reliable answer then to have a quick answer. In wartime it is of greatest important to have a fast answer ("quick on the trigger").

5.5 Results on real radar data

For testing the classification techniques, we use Data Set I (Section 4.3). The input set contains 357 profiles, the independent test set contains 339 profiles. The absolute aspect azimuths are within 20 degrees.

To investigate the properties of the classification techniques, we construct a large number of classifiers using the four techniques as described before and varying training sets as follows:

Carry out the following steps for $N_{\text{train}} = 8, 24, 40, \ldots, 152$:

1. Choose, randomly, $N_{\text{train}}/4$ profiles per class from the input set and use them as a training set.

2. Construct the classifiers

$C_{NN,\boldsymbol{p}_1,\ldots,\boldsymbol{p}_{N_{\text{train}}}}$	:	Nearest Neighbour
$C_{CNN,\boldsymbol{p}_1,\ldots,\boldsymbol{p}_{N_{\text{train}}}}$	:	Condensed Nearest Neighbour
$C_{RR,\boldsymbol{p}_1,\ldots,\boldsymbol{p}_{N_{\text{train}}}}$	:	Radial Basis Functions with Random Centre Selection
$C_{RGS,\boldsymbol{p}_1,\ldots,\boldsymbol{p}_{N_{\text{train}}}}$	:	Radial Basis Functions with Gram-Schmidt Centre Selection

   This gives, for each classifier, a_3 and a_4.

3. Classify all profiles in the test set using these classifiers. Compute the percentage of false classifications. This gives a_1. Also keep track of the number of floating point operations (flops) used for classifying a single profile (a_2).

4. Repeat steps 1-3 thirty times, and compute the averages and standard deviations of a_1 to a_4.

The results are shown in Figures 5.6 through 5.9. The length of the error bars denotes the 75% confidence level. That is, assumed that the errors on the average are normally distributed it shows the interval in which three quarters of the total number of realisations of the average would fall.

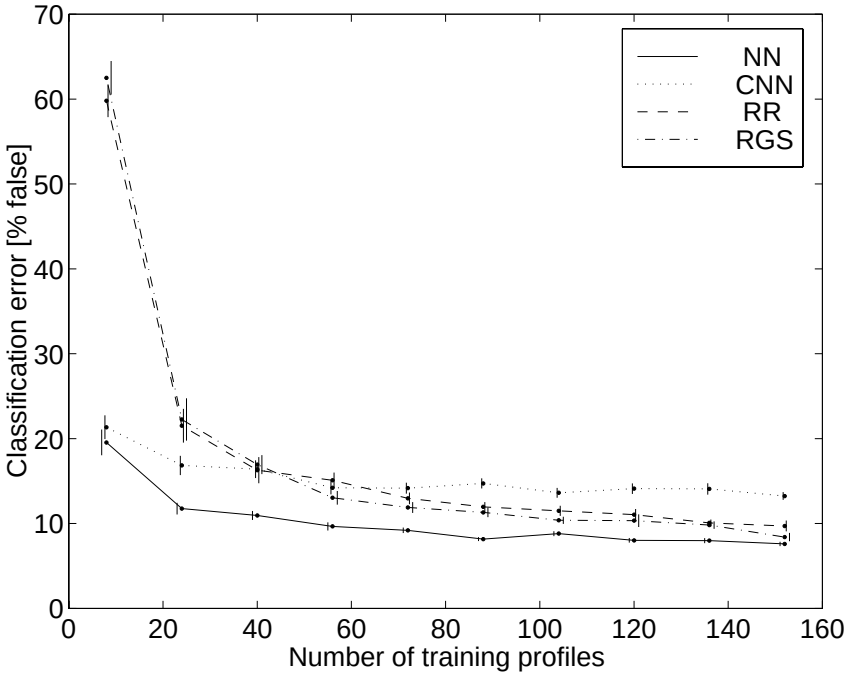

Figure 5.6: *Average error percentage (a_1) as a function of training set size for the four classification methods. The dots show the averages, the length of the error bars represents the 75% confidence levels. Note that the error bars contain a slight off-set with respect to the dots — this is done for display reasons in order to avoid overlap of the bars from the different curves.*

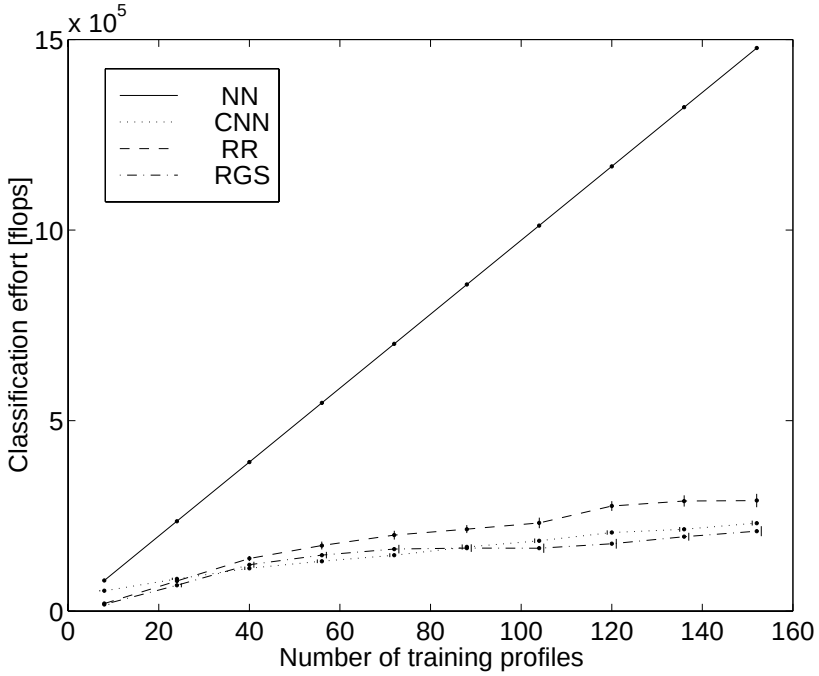

Figure 5.7: *For each of the four methods the average classification effort and its 75% confidence levels for a single test profile (a_2) as a function of training set size. The error bars contain a slight off-set with respect to the dots — this is done for display reasons in order to avoid overlap of the bars from the different curves.*

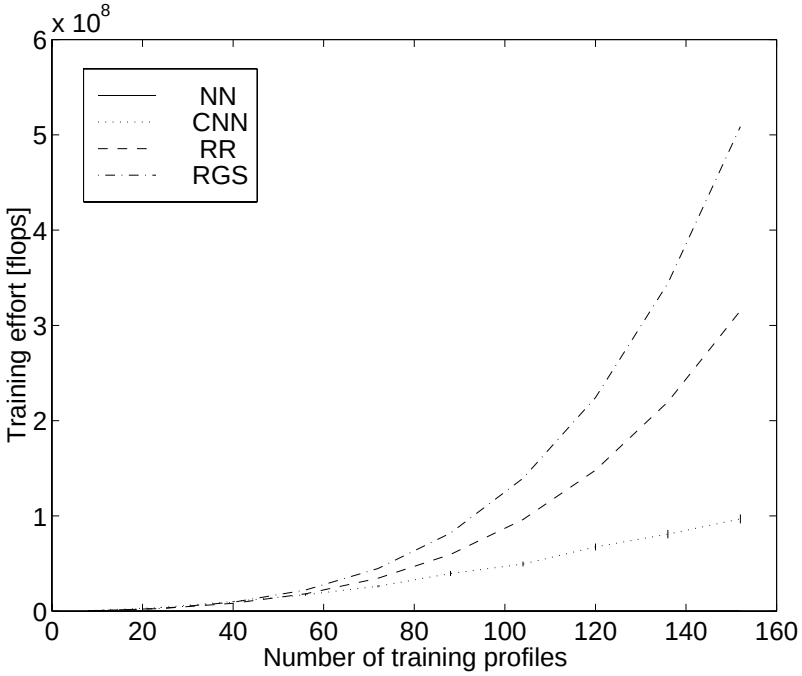

Figure 5.8: *Average and 75% confidence levels of training effort for classifier (a_3) as a function of training set size. The curve for the NN is identical to zero. The error bars contain a slight off-set with respect to the dots — this is done for display reasons in order to avoid overlap of the bars from the different curves.*

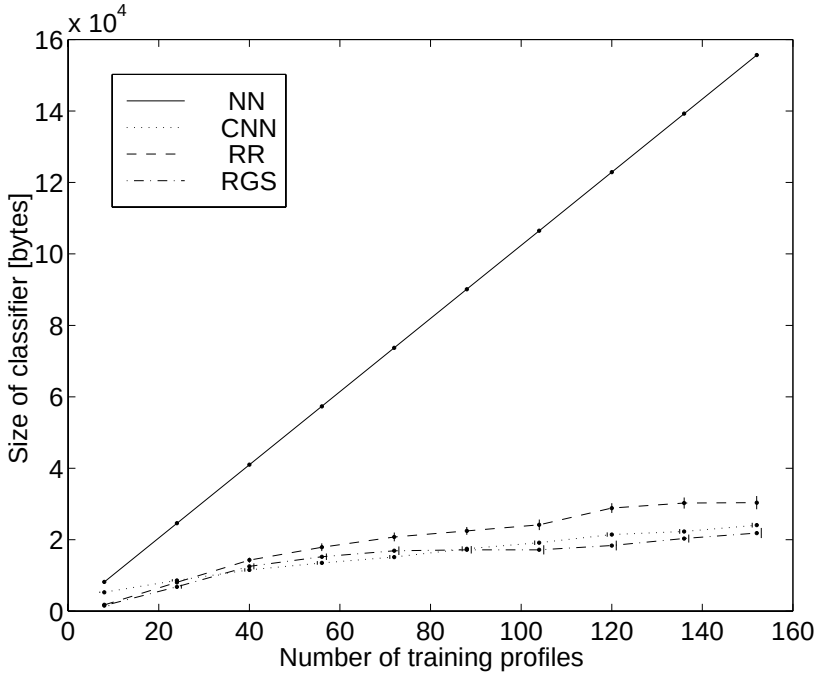

Figure 5.9: *Average and 75% confidence levels of the size of the resulting classifier (a_4) as a function of training set size for the four classification methods. The error bars contain a slight off-set with respect to the dots — this is done for display reasons in order to avoid overlap of the bars from the different curves.*

5.6 Discussion of results

1. *Training effort*

   Figure 5.8 demonstrates that the efforts for training the classifiers varies greatly for the different classification techniques. For the nearest neighbour no training is necessary. A moderate effort is required by the condensed nearest neighbour. Both RBF techniques require a considerable training time. The advantage of the RR technique compared to the RGS is that the computational load for training is much smaller.

2. *Classification error*

   Figure 5.6 shows that a fairly good classification rate can be achieved with only a small number of profiles per class in the training set. E.g., the nearest neighbour technique needs only 40 input profiles (on average one profile per class per two degrees) to achieve an error less than 11%. It indicates that a rather crude coverage of aspect angle (much larger than a coverage dictated by the speckle limit) suffices for reasonable classification. Again, however, one must be aware that the performance is favoured by the small number of classes.

   We observe that the NN error converges as the number of training profiles increases. Tentatively we conclude that the NN error has its asymptotic error at approximately 8% from which it follows that the optimum (Bayes) performance would achieve an error between roughly 4% and 8%.

   For small sizes of the training set, both RBF techniques have poor classification rates. This is because half of the training profiles has to be used for evaluation[5]. If more profiles are added to the training set, this effect becomes less important as the redundancy of this set increases.

   The condensed nearest neighbour has an approximately 6% higher classification error than the normal nearest neighbour for all sizes

[5] We could have considered to use only half of the training set for NN classification. However, we judged it more fair in the comparison to provide to both techniques the same amount of training data.

of the training set. As discussed in Section 5.3.5, profiles were deleted in the condensing procedure that somehow contributed to the decision boundaries.

Figure 5.6 also points out that, of all techniques we considered, the nearest neighbour technique has the lowest classification error. This is because all available data are used for classification, whereas the other techniques, in reducing the available data, also remove non-redundant information.

3. *Classification effort and classifier size*

   If we consider Figures 5.7 and 5.9, the classification effort and the size of the classifier have similar forms, because they are both related to the number of profiles that are present in the resulting classifier. In the nearest neighbour case, all profiles are used in the classifier — the CNN classifiers use the condensed profiles only. Both the classification effort and the classifier size are exactly linearly related to the number of profiles in the training set (NN) or the number of condensed profiles (CNN).

   The selected profiles in the RBF techniques are the centres. The major part of the computation arises from the profile-to-profile distance evaluations. A small number of extra flops is necessary for the nonlinear transform ϕ and the matrix multiplication. The same applies to the classifier size of the RBF techniques. Additionally to the centres, a small number of connection weights $((KN_C + N_C))$ has to be saved.

   The two plots show that in the CNN-, RR- and the RGS classifiers only a very small number of training profiles is left over from the total number that was initially available for training. It means that redundant or nearly redundant profiles are removed from the classifiers at the cost of a reduced classification performance.

4. *Classification effort vs classification error*

   The key question now arises: is it worth losing classification accuracy for an increase in classification speed? This question is answered by a plot of a_1 vs a_2 (Figure 5.10). It demonstrates,

for each classifier, the trade-offs between classification effort and classification error.

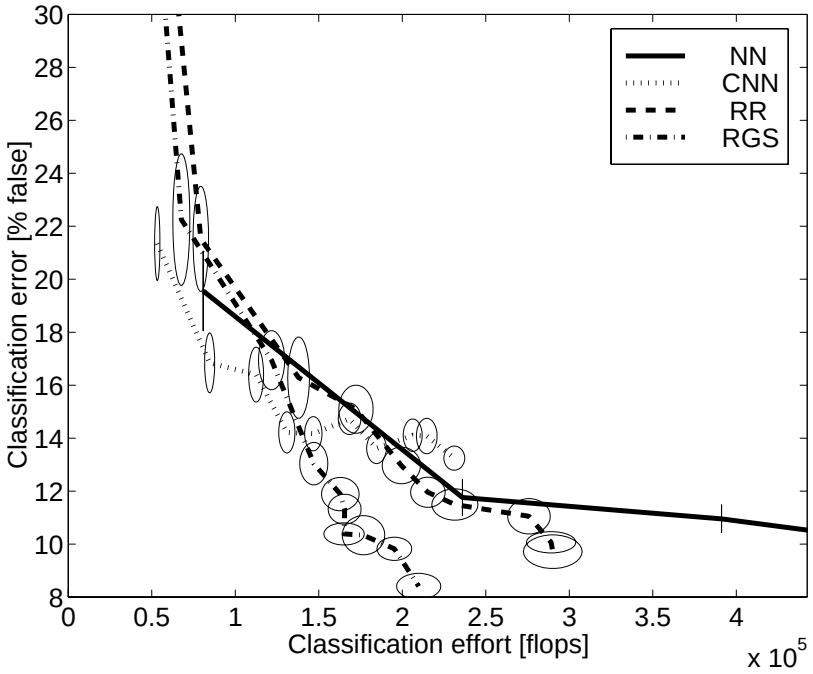

Figure 5.10: *Classification rate (a_1) vs Test effort (a_2) (zoomed). The ellipses denote the 75% confidence level in both directions.*

From this figure one can decide which classifier is most appropriate for a particular classification purpose. The simple approach is to choose a desired classification error on the vertical axis, move horizontally until the first curve in the plot is reached. This classifier should be used as it is the most rapid one. For example, if one desires a classification between approximately 9% and 14% an RGS classifier is the best choice as it gives the quickest answer. The required size of the training set can be found in Figures 5.6 or 5.7.

Conversely, Figure 5.10 can be utilised to find the best classifier given a desired classification speed. E.g. if one is willing to carry out 10^5 flops for one classification, a CNN classifier is the best

	scenario	*crisis*	*wartime*
application	SHORAD	RGS medium train set	CNN small train set
	HIMAD	RGS large train set	CNN small train set
	Fighter aircraft	RGS medium train set	CNN small train set
	Surveillance	RGS large train set	CNN medium train set

Table 5.2: *Best classification technique for given scenario and application*

choice, because it has the minimum error of approximately 16%.

Still, if one desires the minimum classification error possible, a nearest neighbour is appropriate, but it will take a relatively long time to produce an answer.

To return to the previously raised question: unless one desires the minimum possible classification error, using either the CNN or the RGS classification technique has the preference compared to a nearest neighbour rule.

Revisiting Table 5.1 we may insert, using Figure 5.10, the most appropriate classification techniques; see Table 5.2.

We want to stress that filling in this table is merely a demonstration of the method of classifier selection — for a decisive answer on which techniques to use, larger scaled experiments have to be carried out.

5.7 Summary

In this chapter we have described classification tests on range profiles with four different techniques. Although the results were favoured by a small number of classes, it is very encouraging that a reliable classification is found on strictly independent training and test sets and over a large aspect angle range.

A simple nearest neighbour rule achieves the lowest classification error. This result, however, can only be achieved if enough time is available to wait for the answer. Other techniques, such as the Radial Basis Function classifier may have a poorer performance in terms of

classification error, but produce a faster decision. In short, the Nearest Neighbour technique was shown to be most accurate, whereas the neural methods operate most efficiently.

Chapter 6

The Box-Cox metric

6.1 Introduction

[1] [2]When the statistics of a classification problem are fully known, using a Bayesian technique gives the optimal performance. Often, though, such knowledge is not available nor deducible: the options are then to still use a Bayesian classifier or perhaps another pattern recognition technique but now *assume* certain statistical properties. E.g., a *nearest mean classifier* and a *linear discriminant function* both work optimally with *normality*: this special statistical condition requires that samples from all classes are normally distributed with equal variances for each class in each dimension.

Even though it may be known that the problem does not fully agree with the assumptions, the hope is that the classifier is robust enough to still provide adequate results. If the results are *not* considered adequate, the common option is to adapt the technique such that its assumptions are more close to the anticipated statistics of the problem. Another important possibility, however, is to keep using the same, perhaps elementary, technique but now *adapt the statistical properties of the samples*. This can be done by a transformation. Such a transformation is the subject of this chapter. The statistical condition that we desire is normality, the techniques we consider assume normality or, at least, profit from it. The transformation we propose is called the Box-Cox transformation. Several classification techniques will be used, but we shall take the nearest neighbour (NN) rule as the leading technique in three case studies to demonstrate the classification improvement as a result of this transformation.

The reason to use the NN rule is that it is the simplest method that can be applied for range profile classification: we want to demonstrate the potential of the transformation using a straightforward technique without the risk of obscuring its usefulness by choosing a more complex classification method. Also, the literature provides a wealth of reports on this technique, both theoretical and practical, including investigations that consider improved versions — one of them we shall use in a comparison. Finally, an attractive property is that for a large number

[1]This research was initiated by an observation of F. Dommermuth of classification improvement on radar range profiles.

[2]Most of this chapter has been published in [38], a smaller part in [37].

of samples, we know from theoretical grounds that the error does not exceed twice the error found with a Bayes classifier.

The organisation of this chapter is as follows. The next section will provide some theoretical background on the nearest neighbour rule and the influence of the chosen metric. Section 6.3 briefly reviews the metric devised by Short and Fukunaga as it will be used in a comparative study. In the subsequent section the Box-Cox transformation and the associated metric are discussed. Then we will demonstrate the improvement of classification rate using this Box-Cox metric instead of an Euclidean metric or instead of the Short and Fukunaga metric in the first case study: two multivariate exponential PDFs with unequal covariance matrices. In this experiment we also show the results of a Bayesian classifier assuming normal distributions. In Section 6.6 the method is applied to the well-known IRIS set and in 6.7 a considerable improvement in classification is demonstrated if the Box-Cox metric instead of the Euclidean metric is employed. A comparison with the performance of a Radial Basis Function classifier is presented where one might expect that the adaptivity in the RBF approach would make it less dependent on the metric. This chapter is concluded by a summary.

6.2 The Nearest Neighbour classification rule

The previous chapter has already introduced the Nearest Neighbour (NN) classification rule [22]. It assigns an unclassified sample to the class of its nearest neighbour in a set of samples whose classes are known.

The asymptotic convergence of the probability of error ε_{NN} for an infinite number of training samples was first studied by Cover and Hart [19]. They showed that ε_{NN} is bounded by the lowest error possible, the Bayes error ε^*, through: $\varepsilon^* \leq \varepsilon_{NN} \leq 2\varepsilon^*(1 \Leftrightarrow \varepsilon^*)$, independent of the chosen metric.

In the finite sample case, the metric *does* affect the error rate — it determines the convergence to ε_{NN}. Several researchers proposed choices for an improved metric compared to an Euclidean metric for NN

classification. Short and Fukunaga [69] construct, for a two-class problem, the optimum local metric by minimising the difference between the finite sample NN error and the asymptotic NN error. In another approach by Fukunaga and Flick [26] a global metric is proposed. Finally, Myles and Hand [56] devise and compare several generalisations of the optimum local metric for the multi-class problem. The experiments in these papers demonstrate that, by choosing an optimised metric, it is indeed possible to reduce the classification error without increasing the number of samples.

Although it was not used as an assumption in the design of the improved metrics, all simulations in the three papers were based on samples drawn from normal probability density functions (PDFs) with equal covariance matrices for all classes (homoscedasticity) and equal variances in each dimension. As put forward in the introduction, we will denote the occurrence of these three conditions simultaneously by the term *normality*.

It is intuitively clear that, in the finite sample case, also the NN rule performance will improve if the class conditional PDFs tend to normality. After all, in the NN rule with a Euclidean metric the decision boundary is built up from *piecewise* linear discriminant functions.

However, in remote sensing applications, the normality conditions are often violated as the samples (or the elements of multi-dimensional samples) to be processed usually have the dimension of energy. As they are exclusively positive-valued, a one-sided PDF results. This PDF, e.g. an exponential distribution (typical for SAR imaging) or a Poisson distribution (typical for CCD imaging) can be highly asymmetrical and thus far from a normal distribution.

Additionally, it is often found that the variance of the samples (or sample elements) increases with the mean value. E.g. in case of an exponential distribution, the variance *equals* the mean. For multi-dimensional samples this may cause large dissimilarities of the variances in the sample elements.

In the remainder of this chapter we are concerned with the choice for the metric in the case where the PDFs are far from normality which is a situation that often occurs in real-world applications.

6.3 The Short and Fukunaga metric

Short and Fukunaga [69] construct for the two-class case the optimum local metric for NN classification. It is derived from the minimisation of the difference between the asymptotic and finite NN error. They arrive at the following metric (boldface symbols denote vectors):

$$d(\boldsymbol{x}, \boldsymbol{y}) = |\boldsymbol{V}^T(\boldsymbol{x} \Leftrightarrow \boldsymbol{y})| \tag{6.1}$$

where $\boldsymbol{y}$ is a training sample and $\boldsymbol{V}$ is the gradient of the a posteriori probability that a test sample $\boldsymbol{x}$ belongs to class 1:

$$\boldsymbol{V} = \nabla P(1|\boldsymbol{x}) \tag{6.2}$$

$\boldsymbol{V}$ can be estimated by:

$$\hat{\boldsymbol{V}} = \frac{P(1|\boldsymbol{x})}{c}(\hat{\boldsymbol{M}}_1 \Leftrightarrow \hat{\boldsymbol{M}}_0) \tag{6.3}$$

where c is a positive constant and $\hat{\boldsymbol{M}}_1$ and $\hat{\boldsymbol{M}}_0$ are the local sample class 1 mean and the local sample mixture mean, respectively. "Local" is defined as the space within a hypersphere with radius R_K, where R_K is the distance between $\boldsymbol{x}$ and its K'th nearest neighbour.

For the NN computation the magnitude of $\hat{\boldsymbol{V}}$ is irrelevant because we only want to find the smallest distance, so the factor $P(1|\boldsymbol{x})/c$ may be omitted. For computer implementation, the following algorithm is used to classify the sample $\boldsymbol{x}$ given training samples $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ and a choice for K, the number of neighbours.

1. Find the K nearest neighbours to $\boldsymbol{x}$: $\boldsymbol{y}_{j_1}, \ldots, \boldsymbol{y}_{j_K}$.
2. Find which neighbours belong to class 1 and compute the average over these neighbours to find $\hat{\boldsymbol{M}}_1$. Average all neighbours to find $\hat{\boldsymbol{M}}_0$.
3. Compute $|(\hat{\boldsymbol{M}}_1 \Leftrightarrow \hat{\boldsymbol{M}}_0)^T(\boldsymbol{x} \Leftrightarrow \boldsymbol{y}_{j_1})|, \ldots, |(\hat{\boldsymbol{M}}_1 \Leftrightarrow \hat{\boldsymbol{M}}_0)^T(\boldsymbol{x} \Leftrightarrow \boldsymbol{y}_{j_K})|$ and choose the smallest. The class of this vector is assigned to $\boldsymbol{x}$.

A clear drawback of this is method for our applications is that it is designed for the two-class case only. Also, an averaging is required which is not trivial for range profiles where shift-invariance is desired.

Let us interpret this method geometrically. Consider the plane through $\boldsymbol{x}$ perpendicular on the vector $\hat{\boldsymbol{V}}$. In the Short-Fukunaga metric we search for the training sample that is closest to this plane. The class of this sample is assigned to $\boldsymbol{x}$.

6.4 The Box-Cox transformation

In 1964 Box and Cox [11] proposed a parametric power transformation in order to reduce departures from normality. This technique has been extensively studied — a comprehensive review article has been written by Sakia [68].

In many real-world applications it is highly recommended to apply the Box-Cox transformation for a better approximation to normality if the techniques to be applied assume normality or are known to profit from normality. Loosely stated, in many applications it holds that the success of the Box-Cox transformation is due to "the nature of things": the larger the mean value, the larger the variance.

The Box-Cox transformation of x with parameter η is defined as:

$$x^{(\eta)} = \begin{cases} (x^\eta \Leftrightarrow 1)/\eta & \text{if } 0 < \eta \leq 1 \\ \log x & \text{if } \eta = 0 \end{cases} \tag{6.4}$$

For vectors $\boldsymbol{x}$ this transformation is applied to each element separately with the same η. Clearly, a refinement could be made whereby different values of η are used depending on the statistics of particular elements. For simplicity and for the reason that the elements in the samples we consider are of the same nature so that we do not expect large improvements, we have chosen not to increase the number of free parameters.

The reason for choosing $(x^\eta \Leftrightarrow 1)/\eta$ instead of x^η is to achieve continuity for $\eta = 0$. η is a tuning parameter between a linear ($\eta = 1$) and a log-scale ($\eta = 0$). Many researchers present ways to estimate the parameter η, e.g. Perrichi [59] and Carrol [14]. As these methods do not provide the shift-invariance needed for range profiles, we will not attempt to estimate this parameter using their methods, but study the effect of η experimentally. In the applications on the IRIS set and real radar data the value of η can easily be found in a training stage.

Equation 6.4 requires that all elements of the vector $\boldsymbol{x}$ are positive. In our applications this will always be the case as the components of $\boldsymbol{x}$ will have the dimension of energy or length. For generalisations of the Box-Cox transformation for negative values we refer to Sakia [68].

One paper, by Beauchamp and Robson [8], deals with the connection between classification improvement and the Box-Cox transformation. Their starting point is the notion that a *linear discriminant function* is optimal in case of normality [22]. They choose PDFs for $\boldsymbol{x}$ such that the PDF of $\boldsymbol{x}^{(0)}$ is closest to normality and demonstrate the sharp improvement in classification performance for $\eta \to 0$.

In view of the applications of the Box-Cox transformation in a nearest neighbour rule the transformation may equivalently be viewed as an adaptation of the Euclidean distance metric. If $||.||$ is the Euclidean metric, the *Box-Cox metric* is defined as:

$$d(\boldsymbol{x}, \boldsymbol{y}) = ||\boldsymbol{x}^{(\eta)} \Leftrightarrow \boldsymbol{y}^{(\eta)}|| \tag{6.5}$$

Clearly, the Euclidean metric is found for $\eta = 1$. The attractivity of this metric is evident: it is simple and robust and does not depend on class nor the number of classes.

Another way to achieve one of the requirements for normality is to normalise the samples by the variance within their class. The Box-Cox metric, however, is much simpler as it operates on all samples *without* class information. Additionally, division by variance does not reduce possible asymmetries of the class conditional PDFs.

6.5 Case study 1: Two classes, multivariate exponential PDFs

An exponential distribution often occurs as the PDF of energies. Consider the measurement of a complex voltage

$$V = V_r + jV_i \tag{6.6}$$

and assume that both the real and imaginary voltages are random variables with normal distributions, zero mean and variance σ^2. Then it

can be shown, see e.g. [74], that the energy

$$X = V_r^2 + V_i^2 \tag{6.7}$$

has an exponential PDF given by

$$f(X) = \begin{cases} \frac{1}{2\sigma^2} e^{-\frac{X}{2\sigma^2}} & \text{if } X \geq 0 \\ 0 & \text{elsewhere} \end{cases} \tag{6.8}$$

Now suppose that a vector $\boldsymbol{x}$, member of class i, is d-dimensional and has for each of its elements an exponential PDF such as Equation 6.8. If the elements are independent, the joint PDF is:

$$f_i(\boldsymbol{x}) = \begin{cases} \prod_{j=1}^{d} \frac{1}{2\sigma_{ij}^2} e^{-\frac{X_j}{2\sigma_{ij}^2}} & \text{if } X_1, \ldots, X_d \geq 0 \\ 0 & \text{elsewhere} \end{cases} \tag{6.9}$$

where σ_{ij} is the variance of the class i distribution in the X_j direction. For a first experiment we will consider two distributions in two dimensions. We choose:

$$\sigma_{11} = 2, \ \sigma_{12} = 2 \tag{6.10}$$

$$\sigma_{21} = 1, \ \sigma_{22} = 5 \tag{6.11}$$

Note that the covariance matrices for the two classes are unequal (heteroscedasticity) and that for the second PDF the variances in the two dimensions differ.

Consider a new sample $\boldsymbol{q}$ drawn from either of the two PDFs with equal a priori probabilities. The experiment is to classify this sample with a nearest neighbour rule. The asymptotic nearest neighbour error ε_{NN} (one neighbour) and the optimum Bayes error ε^* are given by [22]:

$$\begin{array}{rcl} \varepsilon_{NN} & = & 1 \Leftrightarrow \frac{1}{2} \int_{X_1>0, X_2>0} \frac{f_1^2(\boldsymbol{x}) + f_2^2(\boldsymbol{x})}{f_1(\boldsymbol{x}) + f_2(\boldsymbol{x})} d\boldsymbol{x} \quad = \quad 0.219 \\ \varepsilon^* & = & \int_{X_1>0, X_2>0} \min(f_1(\boldsymbol{x}), f_2(\boldsymbol{x})) d\boldsymbol{x} \quad = \quad 0.162 \end{array} \tag{6.12}$$

The performance of the Euclidean and Box-Cox metrics on the NN classification in the finite sample case was investigated in a simulation experiment. For both metrics we used a single nearest neighbour ($K = 1$). Even though the core of this chapter is a comparison between

metrics for reference a maximum likelihood classifier is included in the comparison [22]. We will refer to this technique as a *Gaussian* classifier. We shall assume that the samples are normally distributed and estimate the parameters (mean and standard-deviation) from the samples.

Obviously this technique will work best when the assumptions are most closely satisfied: the transformed samples are distributed according to a normal distribution. The Gaussian classifier is trained and tested as follows:

1. Perform a Box-Cox transformation on all training and test samples with parameter η.

2. From the training samples $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n$, estimate the averages and variances in each dimension for each class and assume normal distributions to estimate the multi-dimensional class conditional PDFs.

3. For a new sample $\boldsymbol{q}_i$ compute the values of the class conditional PDFs and find which is maximum. The class associated to this PDF is assigned to $\boldsymbol{q}_i$.

Now, let us turn to the simulation experiment:

1. Generate randomly n training samples, $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n$, $n/2$ from class 1, and $n/2$ from class 2.

2. Generate for each class 50 new samples: $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{100}$.

3. Compute the error rate for the Gaussian classifier and the NN using the Box-Cox metric, both for multiple values of η. ($\eta = 1$ gives the Euclidean metric.)

4. Repeat steps 1-3 two-hundred times and find the averages and the standard deviations of the error rates.

Figure 6.1 shows the results of the Box-Cox metric $K = 1$ for $n = 2, 4, 8, 16, 32$ as a function of η. The vertical bars display the estimated standard deviation in the average error rates. As expected, the NN Box-Cox metric is most profitable for a small number of training samples. For larger sample set sizes the differences become negligible.

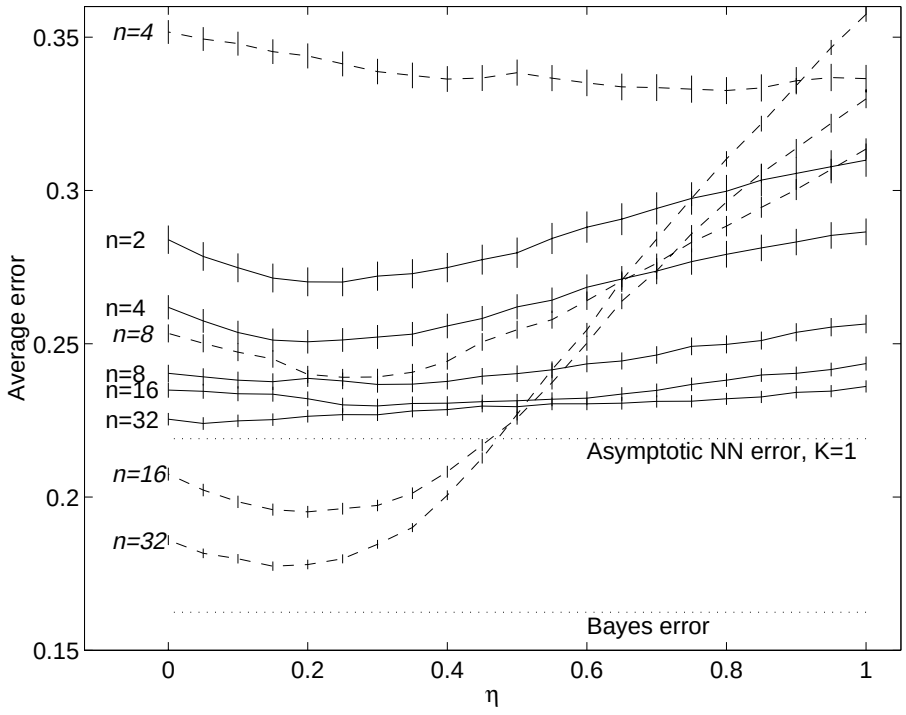

Figure 6.1: *Simulation results of the nearest neighbour classifier using the Box-Cox metric (solid) and the Gaussian classifier on the Box-Cox transformed samples (dashed).*

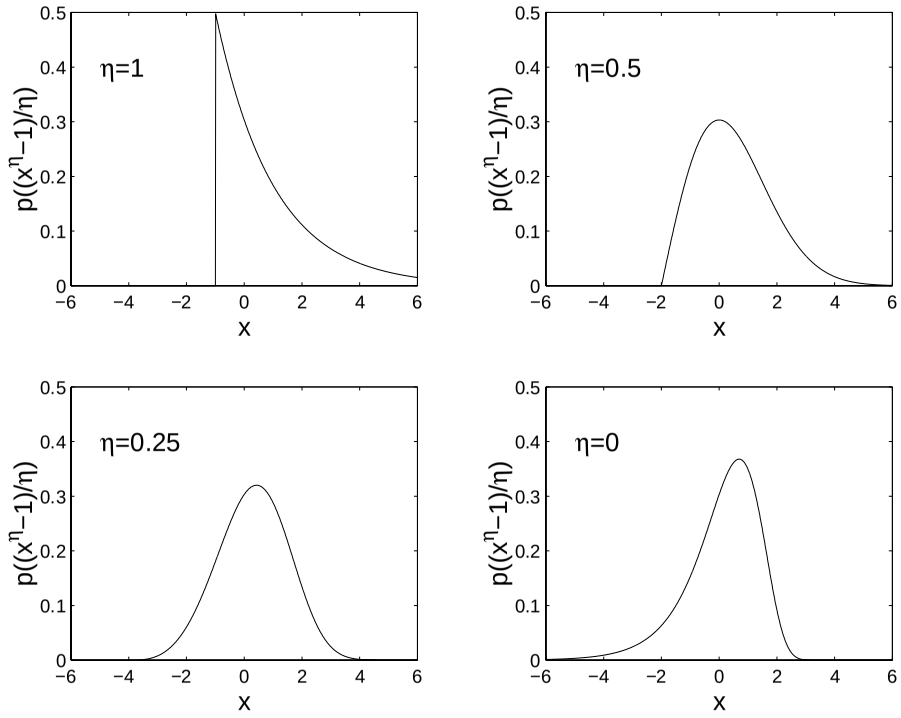

Figure 6.2: *Effect of Box-Cox transformation on an exponential Probability Density Function. As can be seen, for $\eta = 0.25$ the normal distribution is very well approximated. Note that for $\eta = 0.5$ a Rayleigh distribution is found.*

Another observation is that the Gaussian classifier improves dramatically for η approaching $0.2 \Leftrightarrow 0.3$ at larger sizes of the training set. This is not surprising as for these η's the exponential distributions tend to normal distributions as indicated in Figure 6.2.

Observe that the nearest neighbour classifier and the Gaussian classifier show an opposite behaviour: a large effect as function of η is seen for the NN for *small* n and for the Gaussian classifier for *large* n. The reason for this is that the process of estimating the PDFs for a low number of samples is very inaccurate no matter what the value of η is.

The metric performs best for $\eta \approx 0.2 \Leftrightarrow 0.3$. It can be shown that for $\eta = 0.29$ the Box-Cox transformed exponential distribution has the best approximation to a normal distribution in a least-squares sense.

For $\eta = 0$, the logarithm is taken from the samples which makes that both the transformed PDFs have exactly the same variance in the two dimensions (homoscedasticity). As the Box-Cox transformation cannot satisfy all conditions for normality simultaneously (see page 114), an intermediate η is found.

Let us take exponential distributions and investigate the effect of dimensionality of the samples. Using Equation 6.9 we performed the following experiment for $d = 1, 2, 4, 8, 16, 32$ dimensions and for $n = 32$ training samples.

1. Draw each of the σ_{1j}'s and σ_{2j}'s for $j = 1, \ldots, d$ from a uniform distribution between 1 and 5. This defines the PDFs for both classes.

2. Generate randomly n training samples, $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$, $n/2$ from class 1 and $n/2$ from class 2.

3. Generate for each class 50 new samples $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{100}$.

4. Compute the error rate for the Euclidean metric, the Box-Cox metric with $\eta = 0.3$, the Short-Fukunaga metric with $K = 5$ and a Gaussian classifier applied on Box-Cox transformed samples with $\eta = 0.3$

5. Repeat steps 1-4 a hundred times and compute the averages and the standard deviations of the error rates.

The relative reduction of the errors found with the Box-Cox metric, the Short-Fukunaga metric and the Gaussian classifier with respect to the Euclidean metric is shown in Figure 6.3: it shows the average error rates of all techniques divided by the average error rate of the Euclidean metric.

The results show that despite the simplicity of the Box-Cox metric, it is superior to the Euclidean metric and the Short-Fukunaga metric for high-dimensional samples. For $d = 32$ the Box-Cox metric has error rates that are approximately factors three and two better than the Euclidean and Short-Fukunaga metric, respectively. The Gaussian classifier obviously has the lowest classification errors which could be expected from the results of the previous experiment.

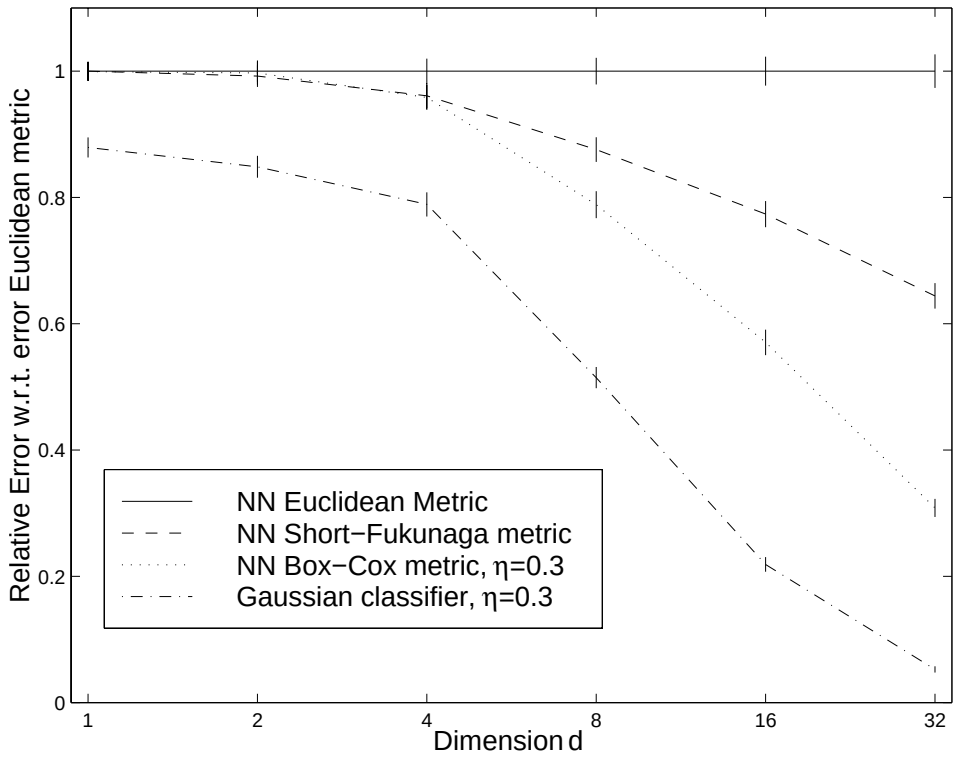

Figure 6.3: *Simulation results for the three metrics as a function of the dimensionality of the 32 samples. For completeness also the results of the Gaussian classifier are shown.*

6.6 Case study 2: The IRIS data

On the IRIS data (three classes of irises, fifty samples per iris, four dimensional samples representing sepal length, sepal width, petal length and petal width, respectively) has first been used by Fisher [25] in his classic paper on linear discriminant functions. As it has been one of the favourite examples for illustrating clustering procedures and for carrying out pattern recognition experiments we will devote this second case study to the IRIS data set. We will show that a modest but computationally cheap classification improvement can be gained using the Box-Cox metric. As the vector components have unit of centimetres, they are positive valued and the Box-Cox metric can be applied without any adaptation.

The hope for finding an improvement using the Box-Cox metric on the IRIS data is again based on "the nature of things": the larger the mean of, e.g. , the sepal length of an iris, the larger its variance.

On the IRIS set we perform two experiments. The first demonstrates the approximation to normality as a function of η. In a second experiment we will show that the NN rule with a Box-Cox metric is indeed a better classifier than the NN rule with an Euclidean metric. The parameter η is found in a training stage.

In the first experiment a portion from the set is chosen randomly to classify the samples that are left over. In detail, $n/3$ training samples are selected from each of the three iris species. The remaining $150 \Leftrightarrow n$ samples are classified using the Box-Cox metric for multiple values of η and the classification error is computed. For each n, this procedure is repeated 200 times and the average error rate and its standard deviation are calculated. Figure 6.4 shows the results.

The optimum power is close to zero which reveals that the PDFs of the IRIS data have their best approximation to normality if a logarithm is applied to the samples. Also it is clearly shown that the improvement is best when the number of training samples is low.

For this reason the second experiment uses an NN rule with only three training samples. Admittedly, it is not very realistic to classify samples with only one training sample per class. The purpose, however, is to show the usefulness of the Box-Cox transformation for real data; this can be done best with a low number of samples. The experiment

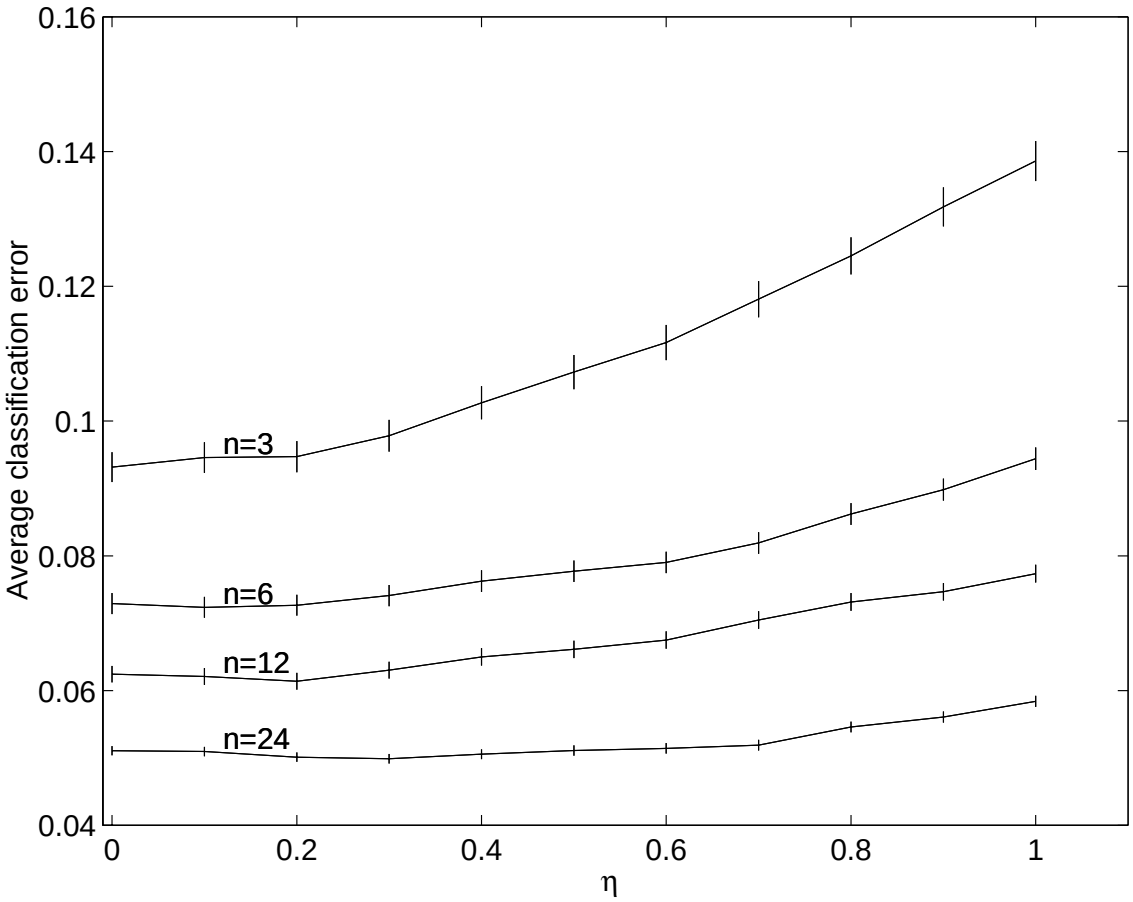

Figure 6.4: *Results of the Box-Cox metric on the IRIS data.*

proceeds as follows:

> Let us separate the IRIS set randomly into two equally sized sets, the *training* set and the *test* set. Both sets contain an equal amount of samples from each class. From the training set we find an optimum η in the following way:
>
> 1. Select randomly three samples from the training set, one sample from each class.
> 2. Classify the remaining samples in the training set using these three samples, a NN rule and the Box-Cox metric for $\eta = 0, 0.05, 0.1, \ldots, 1$.
> 3. Repeat steps 1 and 2 twenty times and average, for each η, the error rates. Find the $\eta_{\min}$ where the lowest average classification error is achieved.
>
> Let us now find the average test error:
>
> 1. Select randomly three samples from the training set, one sample from each class.
> 2. Classify the test samples with this three samples using
>   - the Box-Cox metric for $\eta = \eta_{\min}$, and
>   - the Euclidean metric.
> 3. Repeat the previous two steps twenty times and average the error rates.

We call this complete procedure a *trial*. Carrying out 20 trials we found the following results:

- $\eta_{\min}$ varied between 0 and 0.6, average $\eta = 0.13$.
- The average test errors over all trials were 0.097 and 0.134 for the Box-Cox metric and the Euclidean metric, respectively.
- For each of the twenty trials, using the Box-Cox metric gave a lower error rate than the Euclidean metric. On average the error rates found with the Box-Cox metric were 27% lower than the rates found with the Euclidean metric.

6.7 Case study 3: Radar range profiles

6.7.1 Choice of the metric

Analogously to the metric used in the previous chapter, we apply a metric that is independent of the position of the scatterers in the profile. Additionally, a Box-Cox transformation is applied. We define the *sliding Box-Cox metric* by:

$$d(\boldsymbol{x}_1, \boldsymbol{x}_2) = \min_l ||\boldsymbol{x}_1^{(\eta)} \Leftrightarrow \boldsymbol{x}_2^{(\eta,l)}|| \tag{6.13}$$

where $\boldsymbol{x}_2^{(\eta,l)}$ is the original normalised profile transformed with a Box-Cox parameter η and cyclically shifted over l positions to the right. If $\boldsymbol{x}_2^{(\eta)}$ and $\boldsymbol{x}_2^{(\eta,l)}$ are regarded identical for all l, it can be proved (see Appendix B) that this distance measure is indeed a metric.

6.7.2 Results using the nearest neighbour rule

We use a subset from Data set I, the S-Band data of four different aircraft introduced in Section 4.3. From the input set we selected, randomly, 52 profiles per class for training. The same amount of profiles per class was randomly selected from the test set. Recall that input and test set are measured in different legs and are thus independent. These two sets will be used in this and the next section for classification experiments. As a large number of experiments is performed in these sections, this reduction limits the computational burden.

With a straightforward method, we estimate a proper value for the Box-Cox parameter η using the training set. For each class, we select randomly half of the profiles. This set (of $4 \times 52/2$ profiles) is used to classify the remaining profiles in the training set with a nearest neighbour rule and the Box-Cox metric for several values of η: $\eta = 0, 0.025, 0.05, \ldots, 1$. This procedure is repeated twenty times. We found that for each of the twenty experiments the optimum η (i.e. the η where the lowest classification error is achieved) was between 0.025 and 0.35, the average was 0.125 with standard deviation 0.076.

Let us now classifiy the test set using the full training set and an NN classification with the Box-Cox metric for several values of η. Figure 6.5 shows the results.

Using $\eta = 0.125$, found from the estimation procedure which involved the training set only, we observe that the classification error reduces with a factor 4.5 compared to the Euclidean metric ($\eta = 1$). If, from all twenty estimates of the optimum power η using the training set, the worst η is chosen ($\eta = 0.025$) still a classification improvement of more than a factor three is found.

It demonstrates convincingly that the classification error reduces considerably if the Box-Cox metric is chosen instead of the Euclidean metric. A good value for η can very well be found from the training set.

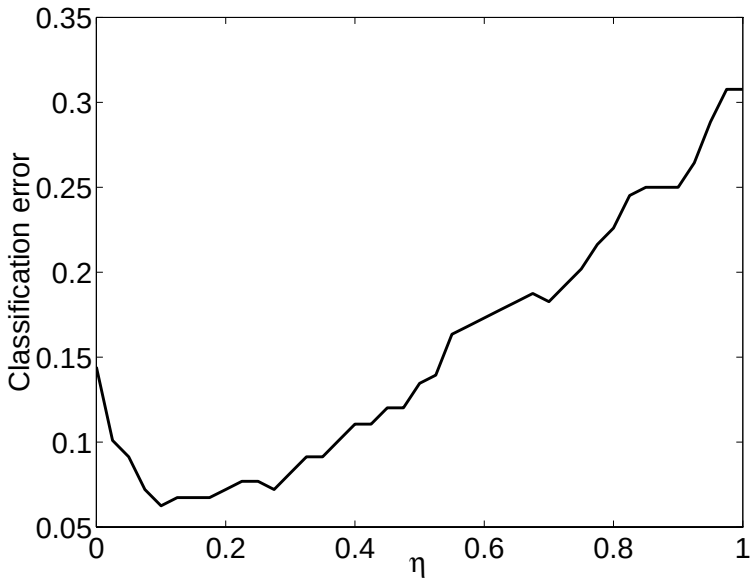

Figure 6.5: *Classification error as function of the parameter η of the Box-Cox metric.*

6.7.3 Results using Radial Basis Functions

[3]The NN rule works on distances between samples only. It is interesting to submit the other technique that works on this information, the RBF

[3]This section constituted the core of the paper [37].

network, to the same data.

In the application of the latter technique, we would be inclined to rely on the training stage to 'repair' a less optimal metric. The expectation is therefore that η has less influence on the classification rate than the nearest neighbour classifier.

We have shown that an evaluation set (Section 5.3.7) can be used to find the optimum number of centers. We can simultaneously use it to estimate a good value for η. For this we simply train RBF-networks for different values of η ($\eta = 0, 0.1, 0.2, \ldots, 1$), and take that network where the lowest evaluation error (Equation 5.15)at the optimum number of centers is achieved.

This is repeated five times for random separations of the training set in a design- and evaluation set. We found for each of the five trials that the value of η where the lowest evaluation error was achieved, was at $\eta = 0.1$. This is in agreement with the result for the nearest neighbour rule. The evaluation errors are shown in the upper plot in Figure 6.6.

For this $\eta = 0.1$, the classification error on the independent test set was on average 0.06 (Figure 6.6, lower figure). The Euclidean metric ($\eta = 1$) performed poorly: we found an error rate of 0.7, only slightly better than a random-class assignment of 0.75. The results show that the RBF method performs even worse than the nearest neighbour technique if η does not have the optimal value.

We can understand this by realising that more training samples are required when the statistics are far from normality, to achieve the same performance compared to the case where (near-) normality is obtained. For RBF networks, half of the training profiles are used for evaluation to avoid overfitting. This does not significantly influence the performance when the value of η is optimal. However, the performance of the RBF classifier deteriorates more seriously than the NN classifier for non-optimal values of η because it is designed using half of the training set.

Nevertheless, the expectation that the settings of the connection weights in the training phase compensate for a sub-optimal metric is refuted. The results suggest that an RBF technique based on the Euclidean metric works best if the samples satisfy normality.

We must conclude therefore that the information for the profile classification is extracted in the very first step: the computation of the

distance between range profiles. If this is not done properly, essential discriminatory information is lost.

6.8 Conclusions

Departures from normality are encountered in many real world discrimination problems. If measured samples need to be classified using a nearest neighbour rule a Box-Cox metric may give a considerable improvement of classification accuracy. This simple and robust metric does not depend on class nor on the number of classes. A single free parameter can easily be found from a training procedure. In experiments using the nearest neighbour method, the decrease of error rate compared to other metrics has been demonstrated convincingly on synthetic data, the IRIS set and range profiles.

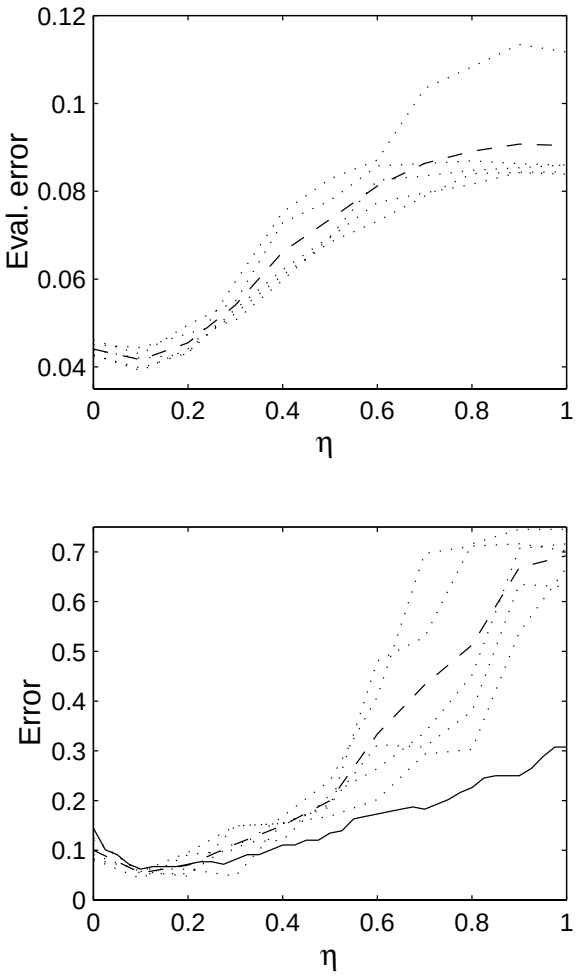

Figure 6.6: *Results of the Radial Basis Function classifier applied to radar range profiles using multiple values of the Box-Cox parameter η. The evaluation error (upper figure) from each of the five trials (dotted) and their mean (dashed) is shown. The lower figure displays the classification errors (dotted), and their mean (dashed). For comparison it also shows, solid, the classification errors for the nearest neighbour method from Figure 6.5.*

Chapter 7

Comparison between measured and predicted range profiles

7.1 Introduction

[1]For a few decades, Radar Cross Section (RCS) prediction techniques have been under development for computing the total amount of radar reflection of an object at a certain frequency and seen at a particular aspect angle.

Obviously, the assessment of the accuracy of RCS-prediction techniques against real radar measurements is crucial to acquire confidence in predictions when measurements are *not* available. It is therefore remarkable that, apart from a few comparison reports on simple objects (e.g. flat plates [66], dihedrals [6] and so-called *Bruderhedrals* [70]), no open-literature accounts on the assessment of the resemblance between predictions and measurements can be found, let alone a comparison on *complex* targets.

In this chapter a start is made to fill this gap as we will perform an investigation on full sized, in-flight aircraft instead of elementary objects in a controlled environment. In the literature, comparison is commonly done on the total strength of the radar reflection, the radar cross section (RCS). We, however, perform the comparison on the distribution of RCS along the line of sight, that is, on the shape of range profiles. This is done to gauge whether they can be used for reliable target recognition. Additionally, the range profile is taken instead of absolute RCS to become independent of circumstances that heavily influence the accurate estimate of absolute RCS of in-flight targets — we mention the calibration (i.e. , the conversion-factor between measured voltages and absolute RCS) of the radar, the estimate of target distance and the aspect angle. For the same reason we have normalised the range profiles in previous chapters. A clear additional advantage using profiles is that the locations in range of particular scatterers and their simulated counterparts are known. We can hence assign contributions in the range profiles to physical scattering processes.

The eventual goal we want to achieve with the predictions of radar range profiles is to be able to classify measured profiles using predicted profiles. It is thinkable that an adequate classification rate can be found even if measured and predicted range profiles do not resemble closely. A

[1]Parts of this chapter have been presented at [35, 34].

situation can be thought of that in the high-dimensional "range profile space" measured and predicted profiles from the same aircraft indeed represent separated subspaces but are still closer to each other than to profiles from other classes. An example is where range profiles from a very large aircraft can always be distinguished from those of small aircraft, even though the features within the profiles do not match between measurements and predictions.

Admittedly, for the main purpose of this thesis only the classification rate matters. We, however, judge it essential to assess the resemblance between measurements and computations in much more detail because it provides understanding when the recognition accuracy is inadequate: in the mentioned example, this accuracy will drop when more very large aircraft types are added such that classification on size alone is no longer sufficient. Hence the comparison provides similarity information that would be obscured in a classification experiment as the outcome of such a test, the recognition error, does not reveal the deficiencies in the modelling process.

The organisation of this chapter is as follows: in the next section we graphically display the correspondence between predictions and measurements and introduce and compute a measure of similarity for the comparison. One source of differences is the error in the aspect angle estimate: in Section 7.3 we improve the aspect angle estimate utilising the range profile data. Subsequently, we will go into the reasons for the dissimilarities between the simulations and the measurements. In the final section we draw some conclusions.

For the remainder of this thesis we will use the abbreviations PRP and MRP for 'predicted range profile and 'measured range profile', respectively.

7.2 Comparison on exact aspect angles

For the comparison we regard Data set II (MRPs) and Data set IV (PRPs). From Chapter 4, recall that all simulated data are from RAPPORT and a model of a Boeing 737-500, and the measurements are done on a Boeing 737 from either the 300 or 500 series. Also bring to mind that **1)** the 737-500 has a 2.4 m shorter fuselage than the 737-300,

2) the engines of the 737-500 model are closed near the front entrance and **3)** the landing gear bay of the model is inaccurately modelled (it is too 'boxy').

The quantification of the similarity between a predicted and a measured range profile is assessed with a straightforward measure: the Maximum Correlation Coefficient μ. This number is the peak value of the normalised correlation function. If $\boldsymbol{x}$ is a vector representing the MRP and if $\boldsymbol{y}$ is a vector representing the PRP, this similarity measure is defined as

$$\mu = \max_i(\boldsymbol{x}^{(i)} \cdot \boldsymbol{y}) \tag{7.1}$$

Here "$\cdot$" denotes the inner product of the two vectors and $\boldsymbol{x}^{(i)}$ is the original vector $\boldsymbol{x}$, but circularly shifted over i positions to the right. For example, if $\boldsymbol{x}^{(0)} \equiv \boldsymbol{x} = \frac{1}{\sqrt{30}}[1\ 2\ 3\ 4]$ then $\boldsymbol{x}^{(2)} = \frac{1}{\sqrt{30}}[3\ 4\ 1\ 2]$. Both $\boldsymbol{x}$ and $\boldsymbol{y}$ are normalised: it means that the sum of squares of the elements (=total energy) equals one. Therefore, if the PRP and the MRP are identical apart from a discrete shift, μ equals unity.

The resulting MRPs and PRPs at the same aspect angles are shown in Figure 7.1 in the two upper diagrams and the lower left diagram. They show ten measured range profiles (blue), each of them aligned with the predicted profile at the same aspect angle (red). The aircraft contour is aligned with the PRP's. From the fifty profiles we show in this figure only the five that have the poorest correlation (the downmost five profiles) and the five that show the best correlation (the topmost five profiles). For all profiles the *magnitudes* are shown. The average Maximum Correlation Coefficients are 0.72, 0.80 and 0.69 for leg 1, 2 and 3, respectively. Figure 7.2 shows another presentation of the data. The positions of the four figures correspond to those of Figure 7.1 but here all fifty measured and computed range profiles are shown in grey-scales. Before we will discuss these results, let us first proceed with the approach to estimate an improved aspect angle.

7.3 Improvement of aspect angle estimate

Comparing an MRP to a PRP at exactly the same aspect aspect angles disregards the possibility that there are likely to be errors in the aspect

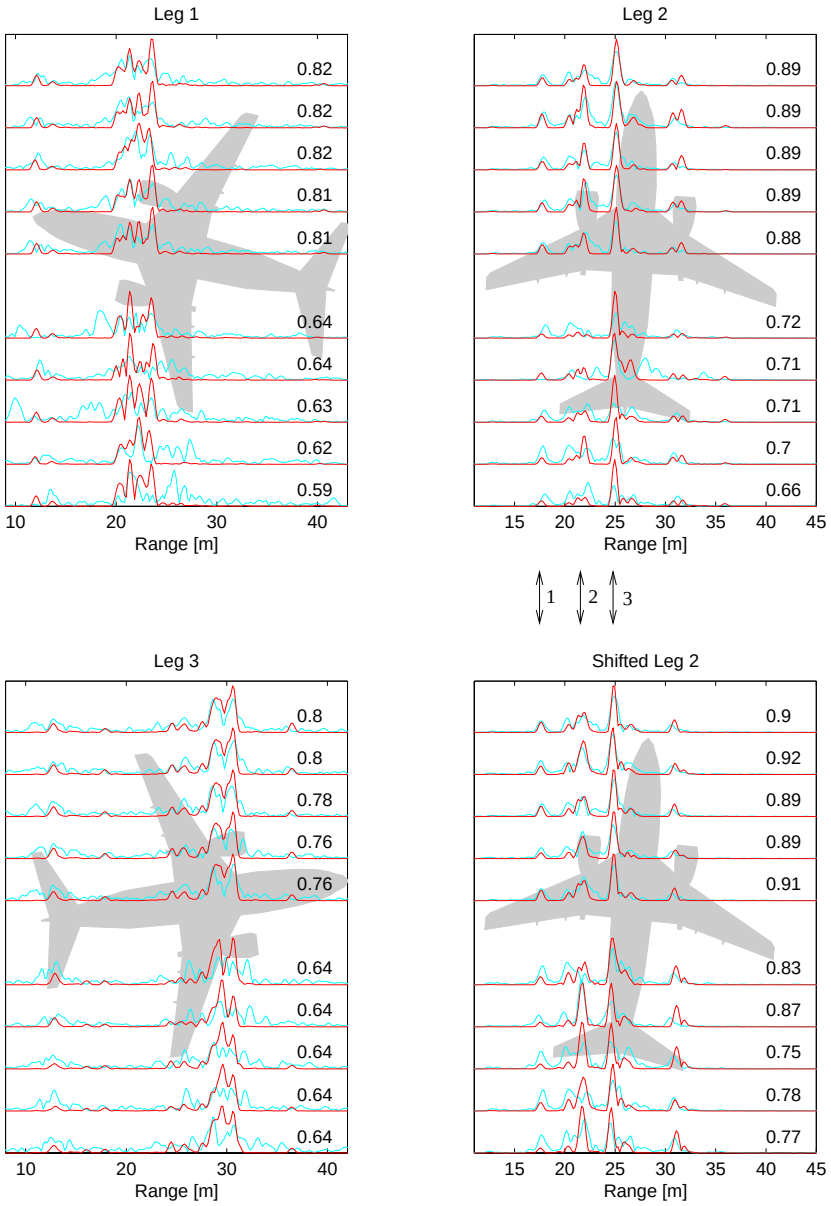

Figure 7.1: *For each of the three legs the best five (top) and the poorest five (bottom) matching measured (blue) and predicted (red) range profiles are shown. The radar is situated at the left hand side. The numbers in the figures display μ. For the two topmost figures and the lower left figure, the PRPs are computed at the estimated aspect angles. The profiles in the lower right show the results of optimal shift from leg 2 (section 7.3). The arrows show where clear physical features of the aircraft can be observed (Section 7.4).*

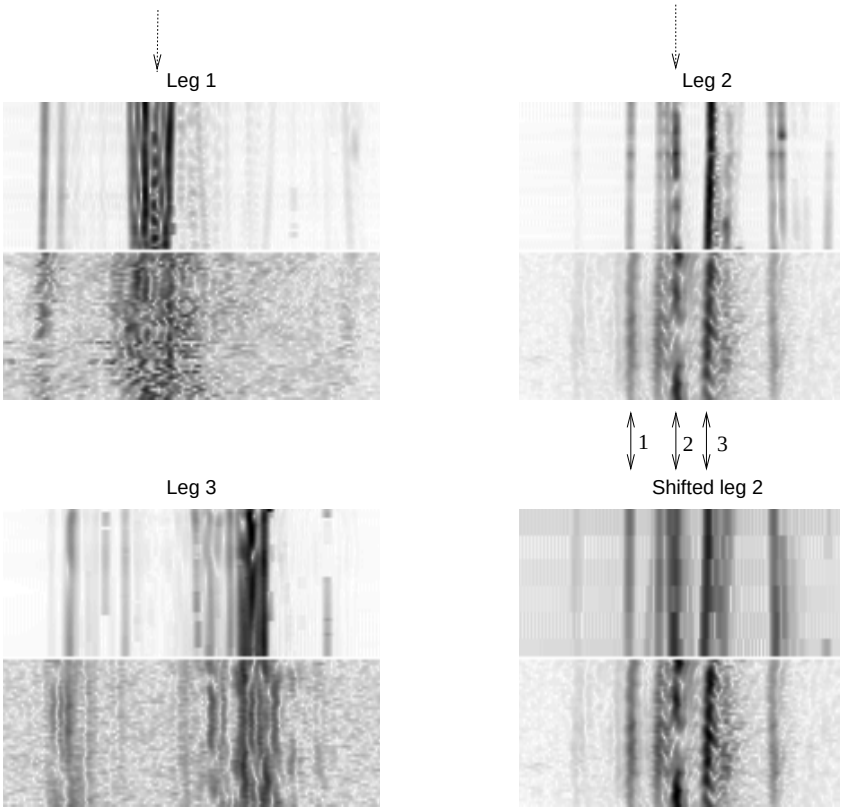

Figure 7.2: *For each of the three legs fifty measured and fifty predicted range profiles are shown as a grey scale image. (white: lowest, black: highest amplitude). The fifty predictions are shown at the top and the fifty measurements at the bottom of each sub-figure. The profiles in the lower-right show the results of optimal shift from leg 2. The horizontal extent is 35 m for all images. The dotted arrows point at range cells where the speckle phenomenon is prominent, the solid arrows show where clear physical features of the aircraft can be observed (Section 7.4).*

angle estimates of the MRPs. For leg 2 we therefore carried out the following procedure using the PRPs computed at the grid of aspect angles.

1. Shift the entire leg over a chosen angle in both aspect azimuth and aspect elevation.

2. Consider the aspect angle of a shifted MRP. Find the PRP in the grid of aspect angles that is closest in aspect angle.

3. Perform step 2 for all MRPs, such that for each MRP a PRP is found. Note that for several MRPs the same PRP can be found, as neighbouring MRPs differ less in aspect angle than the neighbouring PRPs in the grid.

4. Compute μ for each MRP-PRP pair and average.

If the procedure is repeated for several shifts in aspect azimuth and aspect elevation, Figure 7.3 is the result. It shows that the average correlation coefficient increases from 0.80 to 0.85 if a proper shift is chosen. (The worst possible shift gives $\mu = 0.63$.) The profiles in the

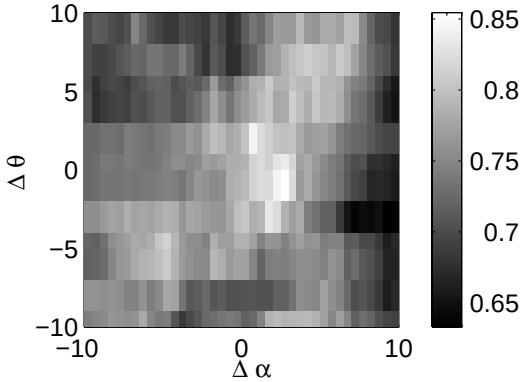

Figure 7.3: *Average Maximum Correlation Coefficients* $< \mu >$ *as function of the shift over aspect azimuth and aspect elevation for leg 2. The maximum* $< \mu >$ *is found at a shift of* $\Delta\alpha = 2.5$ *and* $\Delta\theta = \Leftrightarrow 2$.

lower right diagram of Figures 7.1 and 7.2 show the same MRPs of leg 2,

but now aligned with profiles that were found through this optimisation procedure. A slightly better match is observed corresponding to an increase of μ. We have thus found a better estimate of the target attitude, i.e. $(\hat{\alpha}, \hat{\theta})$, with respect to the radar.

7.4 Observations and discussion

1. *Overall results*
   Viewing Figures 7.1 and 7.2 we observe for leg 1 (near nose-on) some correlation between the two main collections of scatterers (nose and wings/engines). For leg 3 (near tail-on) a fairly good agreement between MRPs and PRPs is seen. The correspondence for leg 2 (near-broadside) is much better. This is supported by the average values of μ for the legs: 0.72 and 0.69 *vs* 0.80. Even for the MRP-PRP pairs with lowest μ, the correspondence is quite good for leg 2 ($\mu = 0.66$) and still present for leg 3 (μ=0.64).

2. *Broadside leg vs off-broadside legs*
   The convincing correspondence for the broadside case, leg 2, is clearly favoured by the aspect angle under which we see the aircraft: we do not have reflections from cavities nor turbines we have at head-on and tail-on aspect angles. Also, at these aspect angles not many range cells contain multiple scatterers that give rise to inaccurate amplitudes.

   When we observe an aircraft the noise level on the measured range profiles increases with the distance to the aircraft. Also, the engine perturbations lead to contributions that appear as a higher noise level. (On page 142 we will further investigate this.)

   The noisier appearance of the first leg can be fully attributed to the engine contributions as the aircraft is measured at a smaller distance than leg 2 (17 km and 26 km, respectively).

   The third leg also has a higher noise level than leg 2. Apart from the perturbations on the turbines, this is caused by the larger distance to the aircraft (38 km). Note that the disturbances due to the turbines (tail-on) are of less influence than those due to the compressor blades (nose-on) as the latter have a larger surface.

3. *Boeing 737-300/500 ambiguity*
   For leg 3, we may have actually measured a Boeing 737-300 instead of the somewhat smaller Boeing 737-500 as we do see extra signal in the MRP left from the leftmost scatterer in the PRP.

4. *Peak amplitudes*
   Considering Figure 7.1 again, it is seen that, although the peak positions are quite well predicted, the amplitudes match less well. One of the probable causes is speckle: to predict the amplitudes of range cells that contain multiple scatterers, the model and the real target should have the same aspect angle within a few hundredths of a degree.

   This is convincingly shown by Figure 7.2 where certain vertical lines (i.e. , the peak amplitude behaviour in time) display this feature, e.g. in leg two in both the predictions and measurements. The ranges where this occurs are pointed at with a dotted arrow. Another example is found in the predictions of leg one.

   Also, the target approximation by small flat patches (instead of round surfaces) and a perfectly conducting surface (instead of dielectric surfaces) is only a first approximation to the actual scattering mechanisms, and is therefore likely to produce inaccuracies in amplitude.

   One of the obvious causes of amplitude mismatch is the normalisation. As no noise-power is present in the PRPs its normalisation pushes the signal components to higher values compared to the MRPs. This is mainly seen in the first and the last leg, where the signal-to-noise ratio of the MRPs is significantly poorer compared to leg 2.

   Another observation is that the spaces between the profile peaks are 'filled' for the MRPs and are much less filled for the PRPs. This influences the amplitudes as the profiles are normalised.

5. *Unmodelled reflective processes*
   Several reflective processes that occur in reality are presently not accounted for in the RCS-prediction code. We mention the modelling of edge-diffraction, creeping waves, cavities and rotating

engine parts.

Possible scatter contributions not modelled on the geometrical model are: antennas, dielectric materials, transitions between dielectric materials, surface roughness, rotating engines and features that differ from aircraft to aircraft from the same type.

As the results show, for broadside views the aircraft range profiles can very well be modelled by Physical Optics and Ray tracing only. The main difference is that the MRPs have a clear signal component between the main peaks in the profile. For nose-on and tail-on aspect angles, the Physical Optics and Ray tracing approximation predicts the most prominent scatterers, but more reflective processes and better models need to be utilised to account for the other contributions in the range profile. Apparently, most of the extra non-explained signal in these profiles is due to the engines: cavities and rotating fans and turbines.

6. *Engine perturbations*
We envisage that the engine rotation does not produce discrete peaks in the range profile at other places than the real positions of the fan. The constant rotation of the engine shaft will give rise to features equidistant in time. The pulse repetition interval, however, decreases smoothly as the emission of the waveform progresses; thus the features are reckoned to be decorrelated and give rise only to an overall increase of the noise-level (treated earlier on page 39).

   A simple model of the engine [2] supports this view. The Boeing 737-500 is powered by two CFM56-3 engines [49]. This "high bypass ratio" turbofan has an area of the first fan stage that is much larger than the second one — therefore most of the modulated signal is reckoned to come from the first fan rather than from the other rotating parts within the engine. To examine whether any discrete responses in the nose-on range profiles are to be expected due to the rotation of this fan, a basic model can be designed where a point scatterer at the location of the fan is sinusoidally moving back and forth with an amplitude of the fan blade thickness. This thickness, which depends on the distance

from the shaft, is typically two wavelengths on S-band. The frequency at which the point scatterer moves is chosen as the BCF, the *Blade Chopping Frequency*. This quantity corresponds to the displacement of the fan over one blade. For the fan the BCF is approximately 4 KHz. (Note that the pulse repetition frequency of the waveform changes gradually from 1.8 to 2.4 KHz.)

Applying this model to our waveform, looking at the aircraft at an aspect azimuth of 13° (the same as for leg 1) and giving each of the point scatterers an amplitude of $1/(L+1)$ ($L+1$ is the number of pulses in the waveform) we found Figure 7.4 for a BCF of 4 kHz and several fan thicknesses. It shows that the contribution from the rotating fan consistently gives an increase of noise level throughout the range profile rather than producing discrete peaks other than the physical locations of the two fans. The two peaks are most apparent at the smallest fan thickness because there the fan nearly appears flat at the used wavelength ($\lambda \approx 0.1$ m). As the fan is chosen thicker, more of the energy is spread over full range profile.

The contributions in the measurements that occur behind the main peaks in the profiles of leg 1 and 3 are therefore due to multiple bounces in the engine cavity. These contributions seem to come from behind the physical engine position as they bounce a few times in the interior before returning to the radar.

7. *Physical scatter-locations*
   We also observe that for leg 2 the main features on the aircraft, like the *flap tracks* (the two dihedral-like structures on each of the wings), the engines and the fuselage, can well be seen in the range profiles. These features, respectively denoted by the solid arrows 1, 2 and 3 in Figures 7.1 and 7.2 are shown to be very stable throughout the leg.

8. *Similarity measure*
   We have regarded the Maximum Correlation Coefficient μ as a measure of similarity between MRPs and PRPs. This parameter can however be quite low, even if a reasonable correspondence between the peaks is observed. As an example, see Figure 7.1,

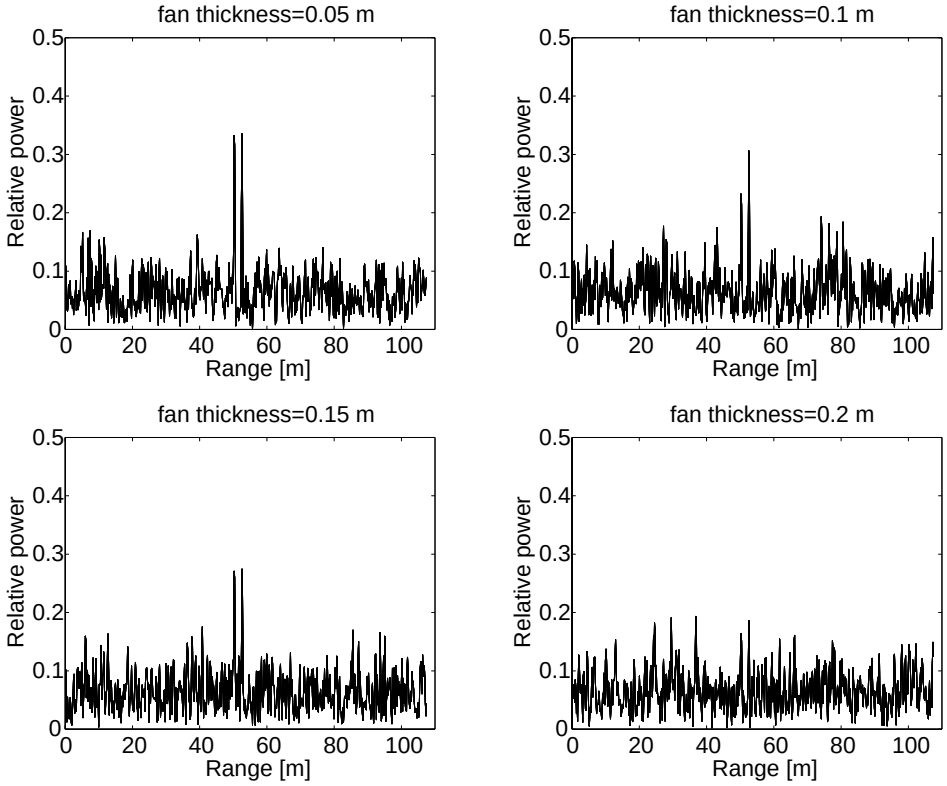

Figure 7.4: *Simulated range profiles (amplitudes) using the ORFEO waveform for Boeing 737-500 engines where the aircraft is seen at an aspect azimuth of* 13°. *We have taken four different fan thicknesses and a BCF of 4 KHz. The resulting two peaks are positioned at the physical locations of the two fans.*

bottom-most profile in the lower right figure. For this MRP-PRP pair, a low μ is found even though most peaks in the PRP are also present in the MRP. The reason is that μ is sensitive to differences in the relative amplitudes of scatterers — unfortunately these are the features that are difficult to predict accurately (see point 4).

7.5 Conclusion

This chapter has shown that RCS-prediction techniques together with an accurate model of an aircraft can produce synthetic range profiles that are accurate for broadside aspect angles and fairly accurate at tail-on and nose-on aspect angles. In the latter two cases, the differences are due mainly to unmodelled reflective processes concerning the engines. The results are encouraging for the recognition of measured profiles with predictions, in particular at broadside views.

Chapter 8

Classification of measured range profiles with predictions

8.1 Using synthetic data: philosophy and objectives

Why do we want to use predicted range profiles for classification?

The answer is that it is impossible to fill a database with measured range profiles only. They are very sensitive to aspect angle, even though we have shown in Chapter 4 that it is not necessary to sample the range profiles so densely that the speckle variations are followed. It has shown to be sufficient to sample in the order of the angle associated with the movement of a scatterer outside the limits of a resolution cell, Rotational Range Migration.

Still, we need a large number of range profiles for a single aircraft to characterise it over the desired range in aspect azimuth and aspect elevation. Additionally, we do not want to perform a supplementary measurement each time a new type of aircraft is made.

For military aircraft the need for a solution to the database problem is even more urgent. New aircraft from hostile nations will never participate in a measurement campaign. Also, fighters have different stores configurations depending on their mission: center tanks, pods and missiles. To measure the radar signatures of all occurring stores configurations is a formidable task — if possible at all.

A way to avoid lengthy and expensive campaigns is the use of RCS-prediction techniques in conjunction with a high-quality computer model of the object. The model can be viewed at any aspect angle and range profiles can be computed through the numerical simulation of electromagnetic scattering processes. Once we have confidence that range profiles can reliably be simulated for complex targets like aircraft, profiles from new aircraft can be produced using a computer model only and measurements can be avoided.

Evidently, the gain for military aircraft is even larger: models of stores can be fitted onto the computer model and profiles can be recomputed. Additionally: RCS-predictions may reveal the variation of range profiles as a function of configuration. It may turn out that changing the configuration slightly does not significantly alter the range profile.

Finally, if we use a different radar or apply other waveform parameters, well-validated prediction techniques may avoid the necessity to

re-measure the profiles for the database.

This approach (i.e. , using predictions for the classification of measurements) is new: to the best of our knowledge no reports on this field have appeared in the open literature.

In the remainder of this chapter we will show that the approach is feasible, albeit on a relatively small data set. Clearly, it shows a "proof of principle" only — the lack of data did not allow us to perform a decisive test.

In the next section the background will be given for the experiments in the remainder of the chapter. We shall deal with the preprocessing of range profiles which includes more steps than we have used sofar. One of the interesting additional procedures is to low-pass filter the profiles in order to reduce the uncertainties in predictions and measurements. Subsequently, we introduce the declaration *ambiguous*: here the classfier is not forced to produce a class declaration, but is also allowed to leave the profile unclassified. Then, we will present two distinct ways to classify the leg as a whole. After this preparatory section, the methods and procedures will be applied to the range profiles from the ORFEO measurements and RAPPORT predictions in the third part of this chapter. Finally, we shall draw the conclusions.

8.2 Theoretical background for experiments

8.2.1 Range profile preprocessing

Sofar we have classified measured range profiles (MRPs) with profiles that were also measured. In this chapter we will classify measured range profiles with *predicted* range profiles (PRPs). It is necessary to understand that there is a fundamental difference between the experiments in the first case and those in the latter. Let us elaborate on this now.

Consider a single type of aircraft. If we use the MRPs for both training and testing, we assume that for a very large number of samples in both the input and test set the derived statistical properties of both sets are equal because, eventually, in both the input and test set most occurring configurations of the aircraft are seen under a sufficient

number of aspect angles.

In the hybrid case the input set contains PRPs, the test set MRPs. Their nature is very different due to the large number of simplifications in the simulation process and the absence of noise in the predicted range profiles. Therefore, no matter how many profiles are computed or measured, the statistical properties of the MRPs and the PRPs will always be different. Loosely (and certainly too pessimistically) stated, we try to "classify apples with oranges".

Supplementary to the preprocessing steps we have applied in previous chapters on measured range profiles alone, we should therefore put an extra effort into removing as many as possible properties from both the MRPs and PRPs that do not contain information on the target type but are a result of the generation process of the profiles only.

We have identified five operations to accomplish this task; those mentioned under numbers 2,3 and 4 have been treated before in the case when both the input and test set contained MRPs.

1. *Vertical offset removal*
   Visual comparison between predictions and measurements show that the amplitudes in the parts of the range profile where no signal is present, are much larger for the MRPs than for the PRPs. Clearly, this is most apparent when the signal-to-noise ratio is low. A typical example is shown in Figure 8.2, topmost figure.

   In this way, mismatch is possible as noise contributions are confused with real radar reflections. Wishing to avoid such an unwanted correlation, it is advisable to subtract an offset from each of the elements of the range profile.

2. *Vertical linear scaling*
   We do not want the classification to be dependent on the total energy that is contained in the profiles. We therefore should scale the profiles with certain factors to give the MRPs and PRPs comparable amplitudes.

3. *Horizontal offset removal*
   The positions of the scatterers in the range profile are undefined and an optimal horizontal shift has to be found for adequate classification.

4. *Vertical nonlinear scaling*
   To emphasise features that are common between predictions and measurements, a nonlinear scaling is desired, known from Chapter 6 as the Box-Cox transformation.

5. *Adaptation of resolution*
   Due to the simplifications in the simulation, in several ways MRPs differ from PRPs. It is thinkable, therefore, that the optimal range resolution is not the highest resolution available. The influence of details that differ between MRPs and PRPs can be diminished when, at the same time, sufficient low-frequency features are left for proper discrimination. Hence, it is worthwhile to search for an optimum resolution.

Ideally, we would like to find the parameters involved in these five operations in an optimisation procedure. Such an approach could be followed when at least two independent test sets were available. A search in parameter space to minimise the classification error on one of the test sets hence provides an estimate of the optimum values of the parameters. To find the true classification error, the estimated parameters are used to classify the second set.

We, however, do not have enough data for performing such an optimisation. Therefore, we shall resort to **1)** elimination of the dependencies from each profile individually (the vertical offset will be estimated and subtracted from each profile; dependency on linear scaling shall be eliminated by normalisation), **2)** removing dependency on horizontal offset by choosing a proper metric (we shall, as before, use the Sliding Euclidean Metric), and **3)** taking a fixed value for the nonlinear scaling parameter (based on prior experience we will choose a Box-Cox parameter of $\eta = 0.2$).

We have hence performed four of the five operations. Finding a parameter for the fifth operation is more problematic. We do not have a proper choice in advance for the parameter that will be used to control the resolution, the smoothing parameter σ, nor can we think of a way for elimination. We shall therefore simply study the effect of the resolution parameter on the error in the test set. As will be shown, the important result is that a clear minimum is achieved which shows that the effect is present. The exact values of the optimum σ_{opt} and the corresponding

minimum test error are of less significance — the same experiment should be repeated on an *independent* test set to find whether the values of σ_{opt} and the test error can be reproduced.

8.2.2 Ambiguous declarations

In this chapter we will use the nearest neighbour rule as the baseline technique for the classification of MRPs with PRPs. This rule, in its simplest form, assigns to an MRP the class of the nearest PRP according to some distance metric.

Hence, the classifier is forced to make a decision about a particular MRP, even when two PRPs of different classes both have approximately the same distance to this MRP.

A way to relax this strictness is the introduction of a 'zeroth' class that contains ambiguous classifications. Compared to the true class of a profile, a classification can then have three possible outcomes: error, correct and ambiguous. In other words, if $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_{N_{\text{train}}}$ are the profiles in the input set (PRPs), then $C_{\text{NNA},\boldsymbol{p}_1,\ldots,\boldsymbol{p}_{N_{\text{train}}}}(\boldsymbol{q})$ denotes the nearest neighbour classification with an ambiguous option from the measured profile $\boldsymbol{q}$; the possible outcomes are $0, 1, \ldots, N_C$.

It is recommended in future research to pay attention to ambiguous declarations concerning only a *subset* of classes. For example, if in a military scenario aircraft 1 and 2 are very similar and they are both friendly, an ambiguous classification like "this profile is from either aircraft 1 or 2", is highly desirable. A further step, knowing that we are measuring one out of these two aircraft, could be to use other features or more measurements to finalise the discrimination. In that case we have effectively devised a *hierarchical* classifier. For this research we have limited ourselves to only one ambiguous class.

How can we adapt the nearest neighbour to make the label "ambiguous" one of the possible outcomes? Sofar, we have used only one neighbour in our classification experiments. An extension is to use multiple neighbours and analyse the classes of these neighbours. In short, we use the $(k, l) \Leftrightarrow$ nearest neighbour (extension of the two-class case provided by, e.g, DeVijver and Kittler [20]):

Compute the $k \geq 1$ nearest neighbours to a test profile. If

> the most abundant class among the k nearest neighbours has at least l more votes than the second-most abundant class, then this class is assigned to the test profile. If not, the test profile is assigned to the ambiguous class. Here $k \geq l$, $k \geq 1$ and $l \geq 0$.

Example: suppose that from a range profile five neighbours are computed, and class A has three votes, class B one, class C one and all other classes none. Then both the $(5,1)$ and the $(5,2)$ nearest neighbour rules assign the range profile to class A and the $(5,5)$, $(5,4)$ and the $(5,3)$ nearest neighbour rules put the profile into the ambiguous class.

Note that the (k,k) and the $(k, k \Leftrightarrow 1)$ nearest neighbour rules are identical. If the k neighbours belong to class A, both rules assign the profile to class A; if one neighbour belongs to another class, the difference $l = (k \Leftrightarrow 1) \Leftrightarrow 1 = k \Leftrightarrow 2$ and again both rules provide the same answer: the profile is assigned to the ambiguous class.

Also remind that both the $(1,1)$ and the $(1,0)$-nearest neighbour rules are the same as the original one-nearest neighbour rule from the previous sections.

8.2.3 Cost assignments

To assess the usefulness of the ambiguous class, we have to assign costs to either of the three possible outcomes. If c_c, c_e and c_a are the costs for a correct, erroneous and ambiguous classification, respectively, and $C(\boldsymbol{q})$ the true class of $\boldsymbol{q}$, we assign linear costs:

$$c(\boldsymbol{q}) = \begin{cases} c_c & \text{if } C_{\mathrm{NNA},\boldsymbol{p}_1,\ldots,\boldsymbol{p}_{N_{\mathrm{train}}}}(\boldsymbol{q}) = C(\boldsymbol{q}) \\ c_e & \text{if } C_{\mathrm{NNA},\boldsymbol{p}_1,\ldots,\boldsymbol{p}_{N_{\mathrm{train}}}}(\boldsymbol{q}) \neq C(\boldsymbol{q}) \\ & \text{and } C_{\mathrm{NNA},\boldsymbol{p}_1,\ldots,\boldsymbol{p}_{N_{\mathrm{train}}}}(\boldsymbol{q}) \neq 0 \\ c_a & \text{if } C_{\mathrm{NNA},\boldsymbol{p}_1,\ldots,\boldsymbol{p}_{N_{\mathrm{train}}}}(\boldsymbol{q}) = 0 \end{cases} \tag{8.1}$$

Without loss of generality we may take $c_c = 0$ and $c_e \geq c_a > 0$.

Operationally, the label "ambiguous" can be viewed as a postponement of the classification. As long as the classifier is not sure, no declaration on the target type is made. The cost for an ambiguous

classification therefore depends on the time that is available for a classification. Again the scenario is decisive here: in crisis, the cost for an ambiguous classification could be very low compared to an error. In time of war and close by aircraft, the cost will approach the cost of an error since a rapid answer is required.

To give insight in the use of an ambiguous class, let us come up with a simple Equation for c_a. Imagine a military environment and an observer that desires a classification of an observed target. Initially assuming that the aircraft is hostile, a certain maximum weapon range is presupposed. Now, let N be the number of profiles that can be measured before the observer is within the weapon reach, given an assumed maximum velocity of the aircraft. Regardless of scenario or application for large N the cost of an ambiguous classification should be low, as a long time is available. When N approaches 1, the cost for an ambiguous declaration should move towards the cost for an error. This is desirable as an operator is not interested in the assignment "ambiguous" when the object is too close. Given these boundary conditions, a viable choice for the cost of an ambiguous classification could then be

$$c_a = c_e / N^{\frac{1}{\beta}}. \tag{8.2}$$

The "stress parameter" β is a positive real number depending on the scenario and application: in times of war, β is large (that is a slow decay of c_a as a function of N), and in crisis, β is low to account for a fast drop-off in c_a. After all, in the latter case one wants to be very sure about the classification to avoid escalation. Figure 8.1 shows the ratio c_a/c_e for several values of β.

8.2.4 Leg classification

The measured data we consider in this chapter are grouped in *legs* of fifty profiles each. It means that we know in advance that these profiles originate from the same aircraft type as the radar has been tracking the aircraft during the acquisition. Additionally we know that from a leg, the estimate of the *change* in aspect angle is accurate, but a large overall shift could be present. These two sources of *a priori* information can be used for extra classification accuracy. We shall use two ways in

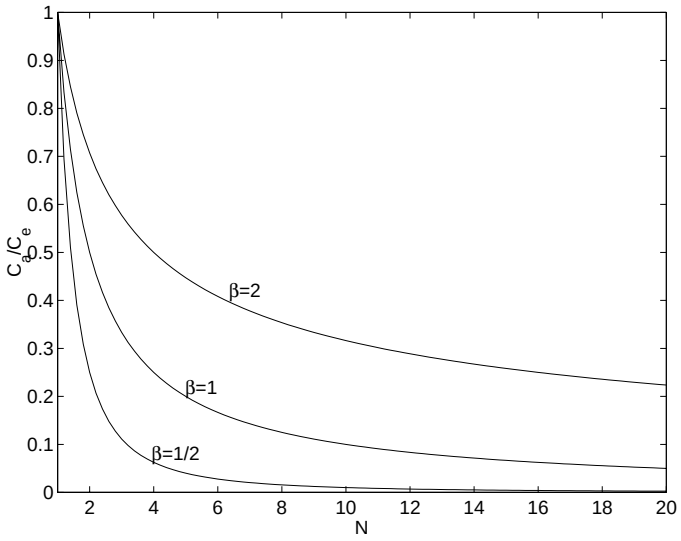

Figure 8.1: *Cost for an ambiguous classification for three values of the "stress-parameter" β.*

this chapter to exploit this information, the *Majority voting* uses only the first kind of *a priori* knowledge, the *Full leg classification* employs both sources.

- *Majority voting*

  An extension to the classification of single range profiles is to count the number of classifications in favour of each class, and assign the entire leg to the class that received most votes.

- *Full leg classification*

  When a single MRP is classified, the closest PRP within (a subset of) the input set is searched for. It means that we allow two subsequent MRPs that have nearly the same aspect angle, to match with two PRPs that have very different aspect angles. In the full leg classification, we will not permit this, but shift the leg as a whole in aspect angle. We shall make use of the following procedure.

  1. Choose a grid of $i = 1, \ldots, N_s$ shifts in α and θ.

For each class C_k ($k = 1, \ldots, N_C$), do for all i the following:

(a) Shift the leg over $\Delta\alpha_i$ and $\Delta\theta_i$.

(b) Find for each of the fifty MRPs in the leg the PRPs that are closest in aspect angle and compute the Sliding Euclidean Distances.

(c) Average these fifty distances. This gives an average distance for each shift and each class: $\overline{d}_{ik}$.

2. Find for each indiviual class C_k the lowest average distance over all shifts: $\overline{d}_k = \min_i \overline{d}_{ik}$. (The shift where this occurs gives an estimate of the leg-shift for this class.)

3. Find for which class the lowest average distance is found — this class, $C_{\arg(\min_k \overline{d}_k)}$, is assigned to the leg.

8.3 Experiments

8.3.1 Available data and preprocessing

Two collections of range profiles are used in this chapter. The first set consists of measured range profiles, the second profile collection is simulated using the RCS-prediction code RAPPORT. These sets are Data set III and IV, respectively, specified in Chapter 4, see Figure 4.3, page 64.

Recall that both collections contain profiles from five civil aircraft, seen at approximately broadside aspect angles: the Boeings 737-500 and the 747-400, the Fokker 100, the Airbus A310 and the McDonnell Douglas of the 80-88 series. The measured data comprise of eight legs of fifty profiles each. For each aircraft profiles are predicted at 215 grid-points in aspect azimuth and elevation, see Figure 4.10, page 75.

As in all previous classification experiments we have normalised the profile (both measured and predicted) such that the total energy in each profile equals unity. As announced earlier, we use the Box-Cox transformation with parameter $p = 0.2$.

We estimate the baseline by searching for that consecutive part in the range profile of 30 elements that contains the lowest average energy (from a total of 648 elements). This average is subtracted from the

range profile. The procedure is carried out for both the PRPs and the MRPs. No new normalisation is done. As an example, Figure 8.2, topmost picture, displays an MRP with the best matching PRP with aspect angle near the MRP aspect angle. The lower diagram shows the same MRP and the best matching PRP after baseline removal.

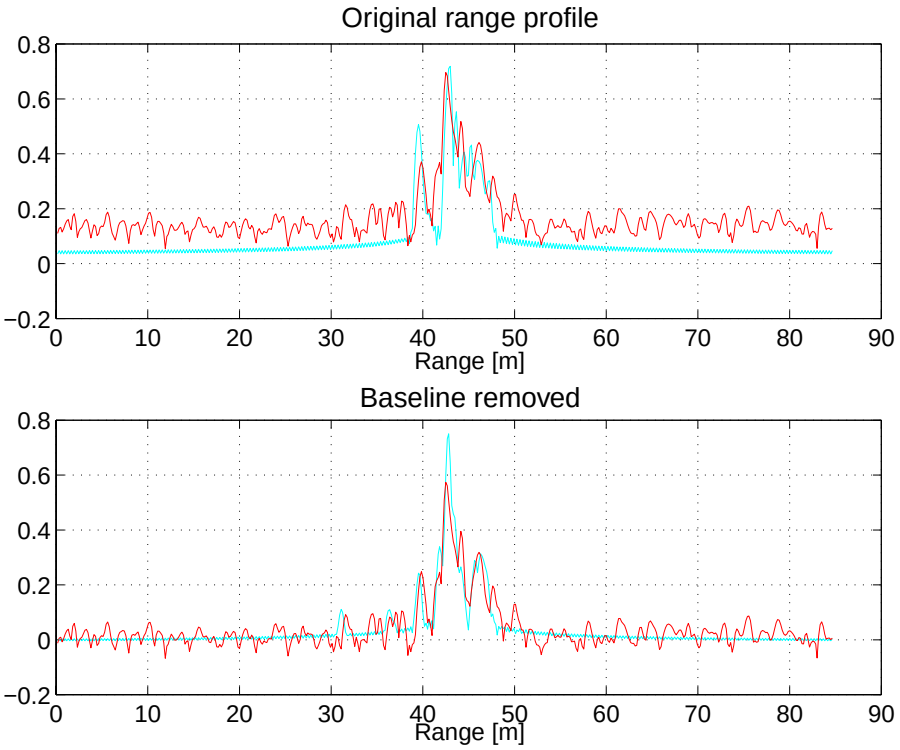

Figure 8.2: *MRP of a Fokker 100 and the best matching Fokker 100 PRP, with (topmost figure) and without (downmost figure) baseline removal.*

8.3.2 One-Nearest neighbour rule

A simple one-nearest neighbour rule is applied to classify the MRPs with the PRPs. It is not necessary to find the nearest PRP searching in the full input set. To limit the computation time (and possibly also to reduce the classification error) only PRPs need to be considered that

are close enough in aspect angle to the MRP. The set of PRPs from which the nearest neighbour will be selected is found by simply taking a circle of five degrees around the estimated aspect angle of the MRP (according to the uniform distribution of Equation 4.1) and choosing all PRPs from all classes with aspect angles within this circle[1]. Then, the distances to all PRPs are computed using the Sliding Euclidean Distance, and the class of the PRP that has the smallest distance is assigned to the MRP.

We found that the classification error was 24.3%. Figure 8.3 displays the results, including the confusion matrix.

We observe that the three Boeing 737-500 legs are classified nearly perfect. Apparently, and as was seen in the previous chapter, the correspondence between measured and predicted profiles at these aspect angles is very good. Also, a nearly perfect classification is seen for the Boeing 747-400. This may not come as a surprise as this aircraft is by far the largest object in the database.

The main source of confusion is between the `MD80` and the `FK10`. The `MD80`-profiles in leg 7 for example, are nearly all classified as the Fokker 100. Looking at Figure 4.6, page 69, we can understand why the classifier has difficulties to arrive at the correct decisions regarding these aircraft. (Call to mind that we look at the aircraft at the approximate angles displayed by the figure: near broadside.) This is because the two aircraft that resemble each other most closely are the `MD80` and the `FK10`: for both aircraft the engines are placed on the fuselage and the horizontal stabilisers are on top of the tail and the aircraft have approximately the same size.

Another origin of errors is the similarity between the Airbus and the Boeing 737-500; they also appear similar if looked upon from broadside even though the Airbus is considerably larger. The engines are placed on the wings and the horizontal stabilisers are mounted on the fuselage.

Despite the confusions, we believe that this experiment has shown that it is possible to identify aircraft with predictions. The error rate is still high, but the main sources of confusion are understood.

With a majority voting for the classification of the full leg, one leg of

[1]Strictly speaking, as we deal with a confinement in two *angles*, we should talk about a *cone* instead of a circle.

the MD80 is classified erroneously; also the leg of the Airbus is classified incorrectly, albeit with a very small margin (26 votes for the B73S and 24 for the EA31).

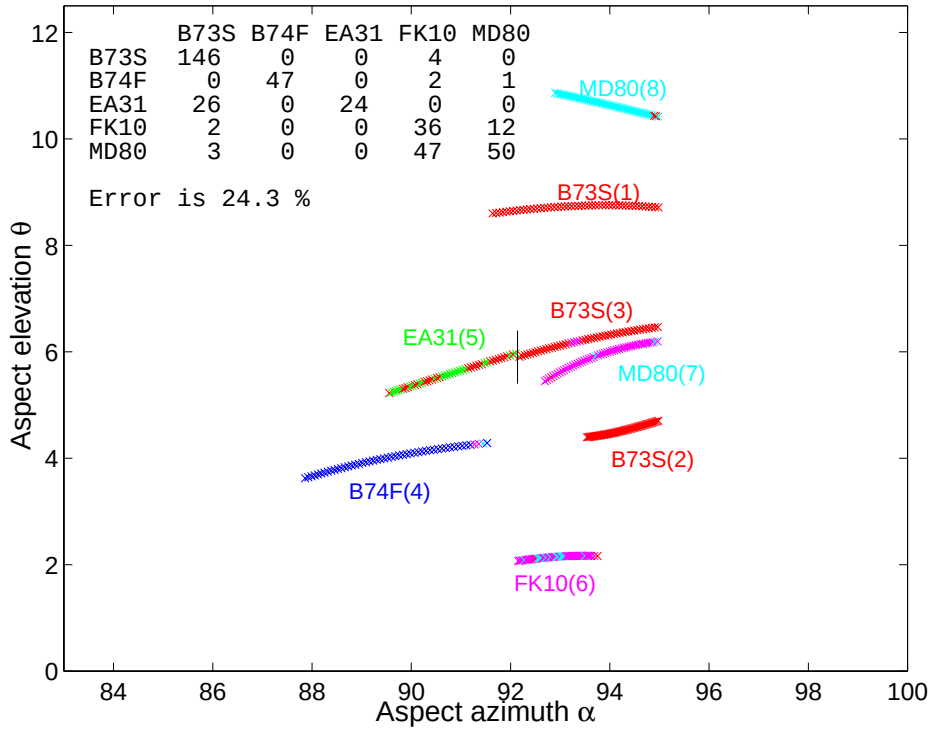

Figure 8.3: *Results of one nearest neighbour classification. Each cross represents a range profile; its colour the classification. Red, blue, green, magenta and cyan represent the B73S, B74F, EA31, FK10 and the MD80, respectively. In the upper left corner the confusion matrix is shown. Each line in this matrix shows the declarations of the profiles for a particular aircraft. E.g., the B73S is classified correctly 146 times, 4 times it was recognised as an FK10.*

8.3.3 Classification with an ambiguous class

We have seen in the previous section that the confusion between similar types of aircraft leads to a rather high classification error. Regarding legs 5 and 6, it seems that the nearest neighbour rule is forced to make a decision, even though it is not sure about it. This is demonstrated by the alternating classification outcomes in, most visibly, leg 5: the Airbus is classified correctly at the start, but a few profiles later it is recognised as a Boeing 737-500. As the profiles are quite similar over such a time-scale, the conclusion is that they are placed near the decision boundary between the two classes.

We apply the (k, l)-nearest neighbour method now with the combinations of k and l, for $k = 1, \ldots, 10$ and $l = 0, k \Leftrightarrow 1$. Figure 8.4 shows the results.

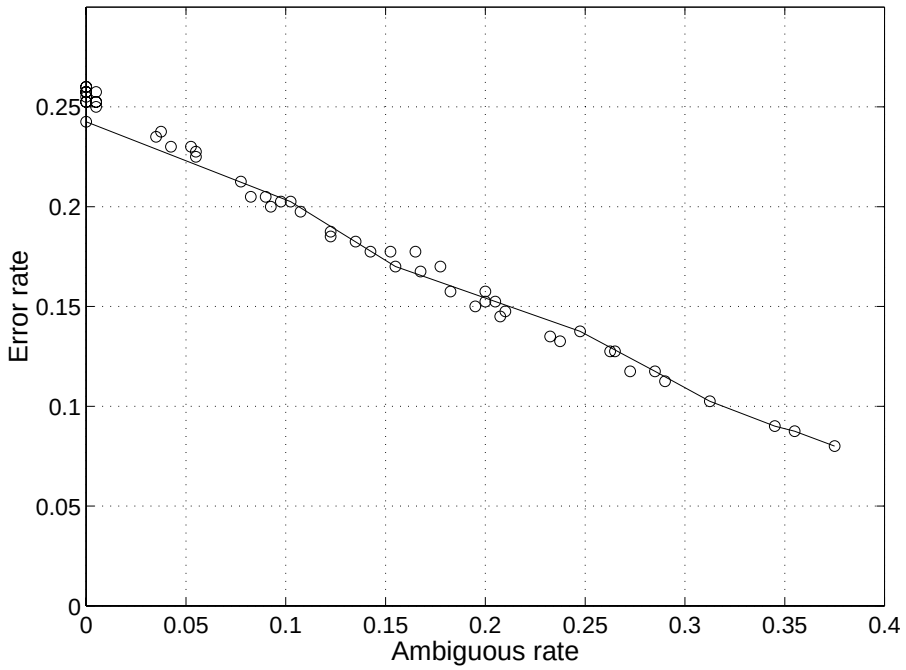

Figure 8.4: *Classification of profiles with an ambiguous class. The combinations for which $k = l$ are connected. The leftmost point of this line is the original 1-nearest neighbour.*

The points in the plot of the ambiguous rate versus the error rate represent in good approximation a straight line with a slope of -0.47. It means, roughly, that for every two ambiguous classifications, one belongs to the originally correctly classified profiles and the other to the originally erroneous classifications.

Then for $c_a \lessapprox 0.47 c_e$ it is profitable to include the ambiguous option — one would choose the rightmost point of the curve. For $c_a \gtrapprox 0.47 c_e$ the one-nearest neighbour gives the lowest total cost. As a tentative example, let us revisit Equation 8.2 and take a stressful environment ($\beta = 4$). Then we find that if the number of profiles we can measure of an unknown object (given distance and assumed maximum velocity/weapon range) is less than twenty the one-nearest neighbour would be chosen. If N is larger, the ambiguous option should be included. (The value of $N = 20$ corresponds to 3.1 s observation time for our waveform.)

As an example we show for one combination, the (2,1)-nearest neighbour, the results on the individual profiles in Figure 8.5. In this case using the majority voting gives an an incorrect classification of one leg (number 7) and two ambiguous legs (numbers 5 and 6).

8.3.4 Optimisation of range resolution

The classification error depends on the resolution of the range profiles. Let us first present the method to investigate this dependency and show some results. After that, we will explain the outcomes. We compare the classification errors using the (k, l)-nearest neigbour. Figure 8.4 has shown that the combinations of k and l when $k = l + 1$ (or, equivalently, $k = l$) give approximately the same results as for the other combinations, so we use only this curve for our comparison.

A lower profile resolution is accomplished by smoothing the range profiles with a moving-average low-pass filter: a Gaussian window, width σ is used. We have performed this for $\sigma = 0, \ldots, 14$ ($\sigma = 1$ corresponds to 0.17 m). Figure 8.6 displays the outcomes. We see that for $\sigma = 4$ the best performance is achieved.

As an example we again take the $(2, 1)$-nearest neighbour and choose the following cost assignments: $c_c = 0$, $c_a = 0.3$ and $c_e = 1$. Figure

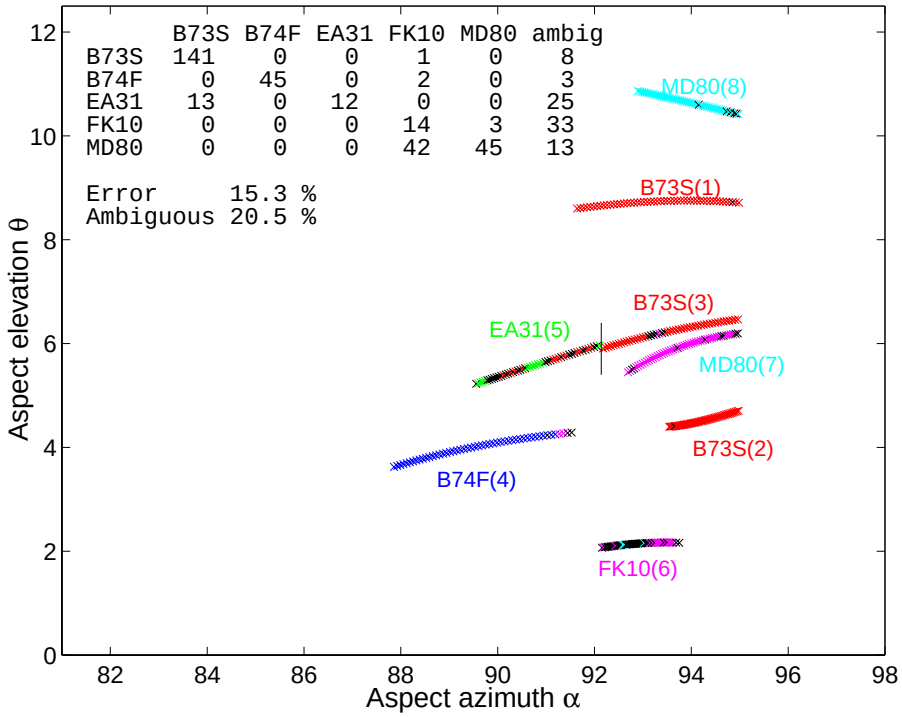

Figure 8.5: *Results of one (2,1)-nearest neighbour classification. Again, each cross represents a range profile; its color the classification. Red, blue, green, magenta and cyan represent the B73S, B74F, EA31, FK10 and the MD80, respectively. The black crosses indicate profiles that received an ambiguous classification. In the upper left corner the confusion matrix is shown.*

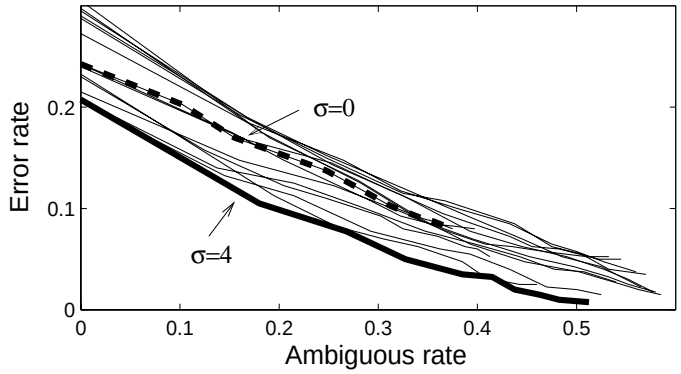

Figure 8.6: *Results of different resolutions induced by the smoothing of range profiles. The curves show for $\sigma = 0, \ldots, 14$ for σ the ambiguous rate versus error rate for the $(k, k \Leftrightarrow 1)$-nearest neighbour. For $\sigma = 4$ (bold, solid line) the range resolution is found to be optimal. The curve for the original profiles is also highlighted (dashed, bold line).*

8.7 illustrates the total cost as a function of σ. As can be seen, again a clear minimum at $\sigma = 4$ is present. These results do not depend critically on the choice for the cost assignments: the same minimum is found for $0 \leq c_a \leq 0.77 c_e$. For larger c_a higher values of σ (up to 6) are found.

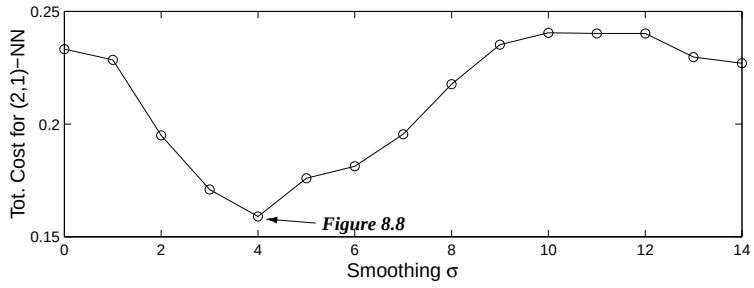

Figure 8.7: *Cost as function of smoothing parameter σ for the $(2, 1)$-nearest neighbour.*

Figure 8.8 shows the classification of the individual legs for a particular point on the curve for $\sigma = 4$. Note now that a majority voting leads

to only one erroneously classified leg (number 8) and one is assigned to the ambiguous class (leg 6).

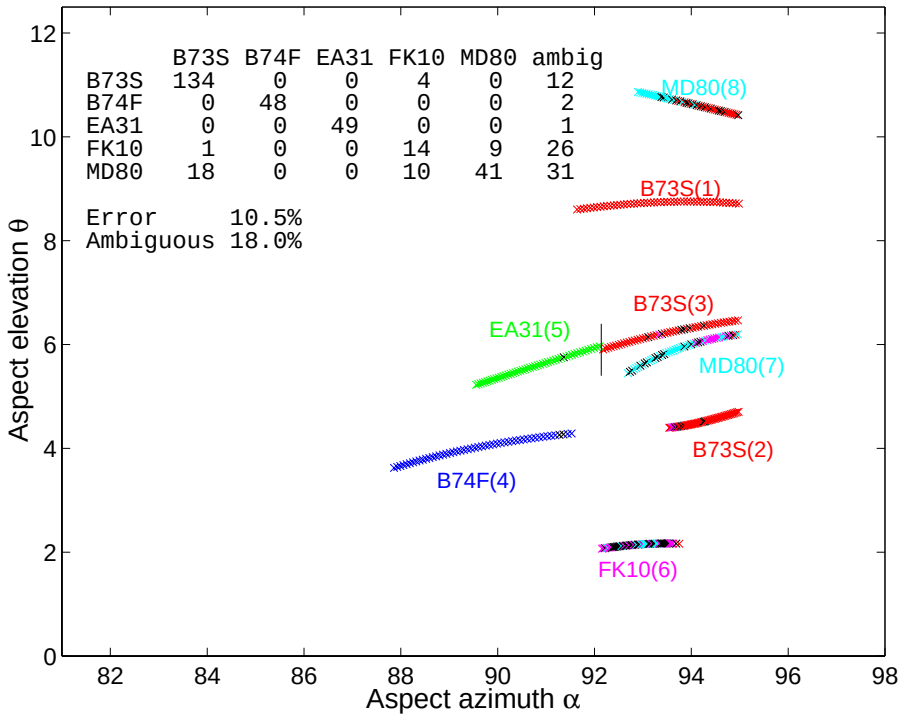

Figure 8.8: *Results of nearest neighbour classification with* $(k, l) = (2, 1)$ *and a smoothing parameter* $\sigma = 4$*, each cross represents a range profile; its colour the classification. Red, blue, green, magenta and cyan represent the B73S, B74F, EA31, FK10 and the MD80, respectively. Black crosses denote profiles that were assigned to the ambiguous class. In the upper left corner the confusion matrix is shown.*

Let us discuss the results. It seems obvious that the classification error increases when the resolution becomes very coarse: too little discriminative features are left in the profiles to allow accurate classification. Some more explanation is needed to understand why the error drops if moderate reductions in resolution are effectuated; after all, intuitively we might expect that the best classification would be achieved at the highest resolution. We will summarise five possible reasons. Apart from the fifth one, they boil down to the hypothesis that

the smoothing makes the errors in the numerous approximations and assumptions less important.

1. *Radar impulse response*

   The radar system has a certain impulse response that is not infinitely short. The latter response is however assumed for the predictions. The computed range profile should therefore be convolved with this response such that it looks more closely to the measured range profile. The true response is unknown, but could be comparable with the chosen smoothing.

2. *Geometrical model imperfections*

   Small errors and omissions could be present in the geometric model, such that the resemblance with the real range profile improves if looked only at the low-frequency features in the profile.

   It does not seem coincidental that, looking at the example of figure 8.8, the overall enhancement is mainly due to the improvement on the roughest model, the Airbus.

   Also, peaks could just 'miss' each other due to imperfections in the model. E.g., suppose that a wingtip gives rise to a very sharp peak in both the measured and predicted range profile. If the model is not particularly accurate, these sharp peaks could fail to coincide in the correlation process and therefore not give a contribution. A smoothing enables the peaks to overlap partially, and eventually a better correlation can be found.

3. *Approximations in RCS-prediction code*

   Reflections from details smaller than 0.5 meters on the geometrical model (if present at all) should be treated suspiciously as the code is designed to operate in the high-frequency domain, and therefore on larger parts. This does not take away that the MRPs *do* contain reflections from such details. With this in mind, note that the optimum smoothing ($\sigma = 4$) corresponds to 0.68 m, which is in the same order of magnitude of the smallest detail that is reliably modelled.

Additionally, several reflective processes are not included in the RCS-prediction code (see Chapter 4). Again, a smoothing may reduce the influence of the approximations that are made.

4. *Undersampling*

   The reduced classification error may be due to an undersampling of aspect azimuth and aspect elevation: the numbers in the Table 4.2, page 64 all depend on the range resolution. As we increase the range resolution cell, we also increase these numbers.

5. *Signal-to-Noise ratio*

   Due to the smoothing, the signal-to-noise ratio of the range profiles increases.

Apart from a decrease in classification error, another advantage of the smoothing can be mentioned. If the profiles are smoothed with a $\sigma = 4$, we can safely reduce the number of data points in a range profile by a factor four and hence profit from a large data reduction, and thus from shorter classification times.

8.3.5 Full leg classification

In the previous experiments we have, for each profile in the leg separately, estimated the aspect angle in the leg by searching for the nearest range profile within a circle of five degrees around the aspect angle found from the tracking data.

Now we want to investigate whether we can do better by shifting the leg as a whole in aspect angle. We employed the procedure from Subsection 8.2.4, taking the smoothed ($\sigma = 4$) range profiles.

Using this procedure with a grid in both aspect azimuth and elevation of 1°, we have found that all legs were classified correctly.

The technique we have used sofar for classifying a leg as a whole, the Majority voting, did not give an improvement of the total classification rate compared to be averaged results for the individual classifications. The current method, where we permit a bias only in aspect angle for the full leg instead of allowing to estimate the angles for each MRP

individually, *did* give an enhancement. It demonstrates that an accurate aspect angle estimate is very beneficial for the recognition accuracy.

Two remarks need to be made here. Firstly, the opportunity to classify fifty profiles in one batch is not likely to be given in operational situations as the time on target in our experiment, nearly eight seconds, is very long. Secondly, we have looked here at slowly manoeuvring aircraft — MRPs from a more violently steered aircraft could not be submitted to a full leg classification using a single bias only.

8.4 Conclusions

We have shown a proof of principle that measured range profiles can be classified with predicted profiles.

The smoothing of the range profiles and the leg-classification were successful improvements compared to the nearest neighbour method on the data set we have considered. Classification improvement due to the applied techniques was evident and the reasons for the improvements were understood. If one has enough time to measure multiple range profiles in a sequence, then the introduction of an ambiguous class in a nearest neighbour classifier is fruitful as the total cost can be brought down.

What the classification enhancement due to the smoothing is concerned, we have to be very cautious to claim that the same will happen in general. The experiment could not be repeated on another test set to check if similar improvements at the same smoothing parameters are observed, simply due to the lack of independent test data.

Chapter 9

Aircraft recognition with radar range profiles: the way ahead

In this final chapter, we wish to conclude this thesis by looking back at the objectives that were presented in the first chapter and by giving directions for future research. This will be done by revisiting the main results and use them to assess what research and developments are desired in the future. Additionally, we will mention relevant novel developments in both algorithms and hardware.

- *Range profile measurement*

  All measured range profiles used this thesis are acquired employing a stepped frequency waveform to synthesise a large bandwidth from individual pulses. Much attention was paid to the timing of the pulses — to achieve robustness for any velocity of the aircraft as well as moderate accelerations we have successfully employed the Velocity Tolerant Waveform.

  Some of the currently available radars (and expectedly many more in the future) have the possibility, however, to emit a large bandwidth in a single pulse. Even though the distances at which a sufficient signal-to-noise ratio can be achieved are smaller (recall that in the stepped frequency waveform all pulses in the sweep are integrated) using such an *instantaneous high bandwidth* pulse offers substantial advantages. Firstly, the time-on-target is very short as a full range profile is measured in a single pulse. Secondly, the highly variable reflections from the compressors and turbines of the aircraft engine(s) (Jet Engine Modulation) do not contaminate the full range profile, but will be located at their physical locations. Measuring a sequence of range profiles, a JEM-spectrum can be produced by monitoring the variability of the range cells that contain the turbines/compressor parts. Hence a simultaneous JEM and range profile classification is possible.

  It is expected that a considerable research effort will be devoted to profit from these technological advances.

- *Classification of radar range profiles*

  Range profiles are discriminative for target type. The robustness of featureless classification made us decide to use techniques based on profile-to-profile distances, rather than to rely on feature

extraction. We have presented a methodology for comparison of different techniques. Furthermore, a dramatic drop in classification error was achieved applying the Box-Cox transformation. Apart from the benefit for radar range profile classification, this nonlinear transformation deserves to be investigated more closely for the use in discrimination problems that concern real world data in a much broader area than radar range profiles only.

Despite the choice we have made in our approach, it is worth paying future attention to the selection of features from the range profiles; the challenge is to make them truly shift-invariant. It is probably too optimistic to expect large benefits in terms of classification accuracy — the main advantage would lie in the increase of classification speed when the classification is done in a lower dimension. For the recognition of measured data with predicted data, an important path to follow is to search for features that are common between the two types of range profiles, rather than to force featureless classification even though it is known that so many simplifications are present in the predictions.

A significant improvement was gained in the classification of sequences of range profiles. We have only considered two different techniques, but we judge more approaches worth investigating are available. One could, e.g. , examine the *dynamics* of a sequence of profiles or find ways to *update* the knowledge on target type each time a new profile from the same aircraft is added.

In the previous chapter we have also touched upon the problem of ambiguous declarations. Many more different but very useful answers could be produced such as an ambiguous classification between a *subset* of types: e.g. if the classifier cannot decide between two different classes, but both types are friendly, then such a declaration is highly valuable in an operational environment. If one still desires classification between these two types, we can add another classifier tailored to discriminate between the two types. (hierarchical classification). It can be trained on specific features in the signatures that were not used before or, possibly, on new measurements.

Also, the classifier should be able to produce the declaration "unknown" when example profiles of the observed object are not in the database. For operational applications such outcomes of classification systems are very important and deserve a considerable amount of attention in future research.

- *Database problem*

  In this dissertation a first step has been made to solve the problem of availability of range profiles. The database should contain signatures seen at a large number of aspect angles from all targets of interest. We have shown that the computations of range profiles using software for electromagnetic computations on computer models of aircraft can be used for mimicking and recognition of measured range profiles. A substantial amount of additional research and development is required, however, to arrive at the anticipated goal.

  One path is to gradually improve the simulation software and the target models. This is a long term effort from which other fields of radar research will profit too, such as RCS-*reduction* programmes. Recently, the RCS-prediction code RAPPORT has been upgraded to include edge-diffraction. Along with these improvements better, more detailed aircraft models are required. This is easily said, but we are aware that these are very difficult to acquire or build. Additionally, an effort should be made to model engine inlets and compressor and turbine blades as they could be important contributors to the total radar reflection when the target is looked upon from near nose-on or tail-on.

  In parallel, we should follow the path of adapting our classifiers by teaching them to look at only those features that are *common* between predictions and measurements and *distinctive* between different target types. By doing this, we become less dependent on the amount of detail in the modelling process. This could be achieved using physical insight; for example, we may attempt to elicit only those features from the measured data that can be reliably predicted with the available RCS-prediction code. Alternatively, the approach could be data-driven: applying automatic

feature recognition techniques, characteristics that are the same for both predictions and measurements may naturally appear.

The work in this thesis has shown that radar range profiles can be exploited for target recognition. For the actual application in real radar systems, though, we still have a large but very challenging and intriguing research effort ahead of us. We predict, however, that aircraft recognition with radar range profiles will become a vital function of many operational radars in the future.

Appendix A

Derivation of radar range profile of point scatterer

A stepped frequency waveform consisting of L pulses at frequencies $f_l = f_0 + l\Delta f$ is emited towards a scatterer at distance R that has a reflectivity ρ. On receival of the reflected pulse, the modulation of amplitude and the change in phase is measured. These two quantities can, for each pulse l, be expressed in a complex number G_l (Equation 3.3 on page 27):

$$G_l = \rho e^{-4\pi j f_l R/c} \tag{A.1}$$

Taking the IDFT from this sequence we can write:

$$g_m = \sum_{l=0}^{L} G_l e^{2\pi j l m/(L+1)} \tag{A.2}$$

$$= \rho e^{-4\pi j f_0 R/c} \sum_{l=0}^{L} e^{2\pi l(m/(L+1)-2\Delta f R/c)} \tag{A.3}$$

Now we shall use the following equality (see reference [78], Equation 5.16)

$$\sum_{l=0}^{L} e^{j\xi l} = \frac{\sin((L+1)\xi/2)}{\sin(\xi/2)} e^{jL\xi/2} \tag{A.4}$$

substituting $\xi = \frac{2\pi y}{L+1}$ with $y = m \Leftrightarrow \frac{2BR}{c}$ where $B \equiv (L+1)\Delta f$ is the bandwidth, we find:

$$g_m = \rho e^{4\pi j f_0 R/c} e^{jL\pi y/(L+1)} \frac{\sin(\pi y)}{\sin(\pi y/(L+1))}. \tag{A.5}$$

taking the absolute value and squaring the result we find Equation 3.4:

$$|g_m|^2 = \rho^2 \left| \frac{\sin \pi y_m}{\sin \frac{\pi y_m}{L+1}} \right|^2. \tag{A.6}$$

Appendix B

The sliding Box-Cox metric

The sliding Box-Cox metric is defined as:

$$d(\boldsymbol{x}, \boldsymbol{y}) = \min_{l} ||\boldsymbol{x}^{(\eta)} \Leftrightarrow \boldsymbol{y}^{(\eta,l)}|| \tag{B.1}$$

where $\boldsymbol{y}^{(\eta,l)}$ is the original vector cyclically shifted over l positions to the right and Box-Cox transformed with parameter η. The equality symbol "$\stackrel{s}{=}$" is used to express that the vector on the left hand side is a shifted version of the vector on the right hand side over an arbitrary number of positions, i.e. $\boldsymbol{y}^{(\eta,l)} \stackrel{s}{=} \boldsymbol{y}^{(\eta)}$ for all l.

If a vector and a shifted version of this vector are regarded identical, the following three requirements have to be fulfilled to prove that $d(\boldsymbol{x}, \boldsymbol{y})$ is a metric:

1. $d(\boldsymbol{x}, \boldsymbol{y}) = 0 \Leftrightarrow \boldsymbol{x} \stackrel{s}{=} \boldsymbol{y}$
2. $d(\boldsymbol{x}, \boldsymbol{y}) = d(\boldsymbol{y}, \boldsymbol{x})$
3. $d(\boldsymbol{x}, \boldsymbol{y}) \leq d(\boldsymbol{x}, \boldsymbol{z}) + d(\boldsymbol{y}, \boldsymbol{z})$

Proof:

1. (a) $d(\boldsymbol{x}, \boldsymbol{y}) = 0 \Rightarrow \boldsymbol{x} \stackrel{s}{=} \boldsymbol{y}$
   $d(\boldsymbol{x}, \boldsymbol{y}) = \min_l ||\boldsymbol{x}^{(\eta)} \Leftrightarrow \boldsymbol{y}^{(\eta,l)}|| = 0 \Rightarrow \boldsymbol{x}^{(\eta)} = \boldsymbol{y}^{(\eta,l_{\min})} \Rightarrow \boldsymbol{x}^{(1)} = \boldsymbol{y}^{(1,l_{\min})} \Rightarrow \boldsymbol{x} \stackrel{s}{=} \boldsymbol{y}$
   Here $l_{\min}$ is the value for l for which $\min_l ||\boldsymbol{x}^{(\eta)} \Leftrightarrow \boldsymbol{y}^{(\eta,l)}||$ is minimised.

   (b) $\boldsymbol{x} \stackrel{s}{=} \boldsymbol{y} \Rightarrow d(\boldsymbol{x}, \boldsymbol{y}) = 0$
   $\boldsymbol{x} \stackrel{s}{=} \boldsymbol{y} \Rightarrow \boldsymbol{x}^{(\eta)} \stackrel{s}{=} \boldsymbol{y}^{(\eta)} \stackrel{s}{=} \boldsymbol{y}^{(\eta,l)} \Rightarrow \min_l ||x^{(\eta)} \Leftrightarrow y^{(\eta,l)}|| = 0$

2. $d(\boldsymbol{x}, \boldsymbol{y}) = \min_l ||x^{(\eta)} \Leftrightarrow y^{(\eta,l)}|| = \min_l ||x^{(\eta,l)} \Leftrightarrow y^{(\eta)}|| = \min_l ||y^{(\eta)} \Leftrightarrow x^{(\eta,l)}|| = d(\boldsymbol{y}, \boldsymbol{x})$

3. Given $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$, pick i such that $||\boldsymbol{x}^{(\eta)} \Leftrightarrow \boldsymbol{z}^{(\eta,i)}||$ is minimised. Then, pick j such that $||\boldsymbol{y}^{(\eta,j)} \Leftrightarrow \boldsymbol{z}^{(\eta,i)}||$ is minimised. Then $d(\boldsymbol{x}, \boldsymbol{y}) \leq ||\boldsymbol{x}^{(\eta)} \Leftrightarrow \boldsymbol{y}^{(\eta,j)}|| \leq ||\boldsymbol{x}^{(\eta)} \Leftrightarrow \boldsymbol{z}^{(\eta,i)}|| + ||\boldsymbol{y}^{(\eta,j)} \Leftrightarrow \boldsymbol{z}^{(\eta,i)}|| = d(\boldsymbol{x}, \boldsymbol{z}) + d(\boldsymbol{y}, \boldsymbol{z})$

Bibliography

[1] S. Abrahamsson, B. Brusmark, G.C. Gaunaurd, and H.C. Strifors. Target identification by means of impulse radar. In *SPIE Vol. 1471 Automatic Object Recognition*, pages 130–131, 1991.

[2] R.G. Aldous. Private communication.

[3] D.J. Andersh, M. Hazlett, S.W. Lee, D.D. Reeves, D.P. Sullivan, and Y. Chu. XPATCH: A high-frequency electromagnetic-scattering prediction code and environment for complex three-dimensional objects. *IEEE Antennas and Propagation Magazine*, 36:65–69, 1994.

[4] D.A. Aushermann, A. Kozma, and J.L. Walker. Developments in radar imaging. *IEEE AES*, 20(4):363–398, July 1984.

[5] C.M. Bachmann, S.A. Musman, D. Luong, and A. Schultz. Unsupervised BCM projection pursuit algorithms for classification of simulated radar presentations. *Neural Networks*, 7(4):709–728, 1992.

[6] T. Balanis, C.A. and Griesser. *Radar Cross Sections of Complex Objects, W.Ross Stone (edt.)*, chapter RCS predictions of Multiple Flat-plate targets using GO–GTD (UTD) and PO–PTD Techniques: a review, pages 260–274. IEEE Inc., 1989. ISBN 0–87942–254–8.

[7] J.S. Baras and S.I. Wolk. Wavelet based progressive classification of high range resolution radar returns. In *SPIE Vol. 2242 Wavelet Applications*, pages 967–977, 1994.

[8] J.J. Beauchamp and D.S. Robson. Transformation considerations in discriminant analysis. *Communications in Statistics–Simulation and Computation*, 15(1):147–179, 1986.

[9] M.R. Bell and R.A. Grubbs. JEM modeling and measurement for radar target identification. *IEEE Transactions on Aerospace and Electronic Systems*, 29(1):73–87, January 1993.

[10] S.A. Billings and G.L.Zheng. Radial basis function network configuration using genetic algorithms. *Neural Networks*, 8(6):877–890, 1995.

[11] G.E.P. Box and D.R. Cox. An analysis of transformations. *Journal of the Royal Statistical Society, Series B*, 26:211–252, 1964.

[12] M.G.E. Brand. Radar signature analysis and prediction by physical optics and ray tracing. The RAPPORT code for RCS prediction. FEL–95–A097, TNO-FEL, 1995.

[13] D.S. Broomhead and D. Lowe. Multivariable functional interpolation and adaptive networks. *Complex Systems*, 21:321–355, 1988.

[14] R.J. Carrol. A robust method for testing transformation to achieve approximate normality. *Journal of the Royal Statistical Society, Series B*, 42(1):71–78, 1980.

[15] N.F. Chamberlain. *Recognition and analysis of aircraft targets by radar, using structural pattern representations derived from polarimetric signatures.* PhD thesis, Ohio State University, 1994.

[16] K.M. Chen, D.P. Nyquist, E.J. Rothwell, and W.M. Sun. New progress on E/S pulse techniques for noncooperative target recognition. *IEEE Transactions on Antennas and Propagation*, 40(7):829–833, July 1992.

[17] S. Chen, C.F.N. Cowan, and P.M. Grant. Orthogonal least squares learning algorithm for radial basis function networks. *IEEE Transactions on Neural Networks*, 2(2):302–309, March 1991.

[18] V. Cortes, C. and Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.

[19] T.M. Cover and P.E. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13:21–27, January 1967.

[20] P.A. Devijver and J. Kittler. *Pattern recognition, a statistical approach*. Prentice Hall International, 1982. ISBN 0–136–54236–0.

[21] M.R. DeWitt. High range resolution radar target identification using the Prony model and hidden Markov models. Master's thesis. AFIT/GE/92–d–15, Air Force Inst. of Tech., Wright-Patterson AFB, OH. School of Engineering, 1992.

[22] R.O. Duda and P.E. Hart. *Pattern classification and Scene Analysis*. John Wiley & Sons, 1973. ISBN 0–471–22361–1.

[23] R.P.W Duin and D.M.J. de Ridder, D. and Tax. Featureless classification. In *First international workshop on statistical techniques in pattern recognition, Prague*, pages 37–42, 1997.

[24] A. Farina, F.A. Studer, and R. Vitiello. High resolution radar for enhanced target detection. In *IEE International Conference on Radar*, pages 256–266, 1992.

[25] R.A. Fisher. The use of multiple measurements in taxonomic problems. *Ann. Eugenetics*, 7:179–188, 1936. Part II.

[26] K. Fukunaga and T.E. Flick. An optimal global nearest neighbour metric. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(3):314–318, May 1984.

[27] F.D. Garber, N.F. Chamberlain, and Ö. Snorrason. Time-domain and frequency domain feature selection for reliable radar target identification. In *IEEE National Radar Conference*, pages 79–84, 1988.

[28] G.H. Golub and C.F. van Loan. *Matrix Computations*. North Oxford Academic, 1983.

[29] B. Harris and S.A. Kramer. Asymptotic evaluation of the ambiguity functions of high-gain FM matched filter sonar systems. *Proceedings of the IEEE*, 56(12):2149–2157, December 1968.

[30] F.J. Harris. On the use of windows for harmonic analysis with the discrete Fourier transform. *Proc. of the IEEE*, 66(1):51–83, October 1978.

[31] R. van der Heiden. Non-constant rotation correction in ISAR-processing. FEL–92–B200, TNO-FEL, 1992.

[32] R. van der Heiden. Vriend of vijand? Het gebruik van radar voor niet-coöperatieve doelherkenning. *Tijdschrift van het Nederlands Elektronica- en Radiogenootschap*, 2(59):59–63, 1994.

[33] R. van der Heiden and J. de Vries. The ORFEO measurement campaign. FEL–96–A073, TNO FEL, 1996.

[34] R. van der Heiden, L.J. van Ewijk, and F.C.A. Groen. A comparison on radar range profiles between in-flight measurements and RCS-predictions. In *Proceedings of the symposium on non-cooperative air target identification using radar, 22–24 April 1998, Mannheim, Germany*, 1998.

[35] R. van der Heiden, L.J. van Ewijk, and F.C.A. Groen. In-flight measurements and RCS-predictions: a comparison on broad-side radar range profiles of a Boeing 737. In *Proceedings of the IEE-radar conference, 14–16 October 1997, Edinburgh, UK*, pages 444–448, 1998.

[36] R. van der Heiden and F.C.A. Groen. Distance based range profile classification techniques for aircraft recognition by radar – a comparison on real radar data. In *AGARD Conference Proceedings on "Tactical Aerospace C3I in Coming Years", 15–18 May, 1995, Lisbon, Portugal*, pages 17/1–17/7. AGARD–CP–557, 1995.

[37] R. van der Heiden and F.C.A. Groen. A metric for the radial basis function network – application on real radar data. In *Proceedings of the "World Congress on Neural Networks – 1996 International Neural Network Society Annual Meeting", San Diego, CA,*

September 15–18, 1996, pages 363–367. Lawrence Erlbaum Ass., Inc., Mahwah, NJ, 1996.

[38] R. van der Heiden and F.C.A. Groen. The Box-Cox metric for nearest neighbour classification improvement. *Pattern Recognition*, 30(2), 1997.

[39] R. van der Heiden and F.J. Willems. Classification of HRR radar profiles with radial basis functions. In *Proceedings of the Workshop on "Radar Imaging and Classification Techniques", AC/243 (Panel 10) RSG.12 TP/1", January 28–29, 1993, FGAN, Wachtberg-Werthoven, Germany.*, February 1993.

[40] J.A. Hertz, R.G. Palmer, and A.S. Krogh. *Introduction to the Theory of neural computation.* Addison-Wesley, 1991. ISBN 0–201–51560–1.

[41] S. Hudson and D. Psaltis. Correlation filters for aircraft identification from radar range profiles. *IEEE Transactions on Aerospace and Electronic systems*, 29(3):741–748, July 1993.

[42] G.A. Ioannidis. Final report for target identification using radar imagery and moment methods. FR80–23–919, Hughes Aircraft Company, Radar Systems Group, 1980.

[43] I. Jouny, F.D. Garber, and S.C. Ahalt. Classification of radar targets using synthetic neural networks. *IEEE Transactions on Aerospace and Electronic Systems*, 29(2):336–334, April 1993.

[44] I. Jouny, F.D. Garber, R.L. Moses, and E.K. Walton. Applications of the bispectrum in radar signature analysis and target identification. In *SPIE Vol. 1471 Automatic Object Recognition*, pages 142–153, 1991.

[45] E.F. Knott. A tool for predicting the radar cross section of an arbitrary trihedral corner. *IEEE SOUTHEASTCON*, pages 17–20, April 1981.

[46] E.F. Knott, J.F. Shaeffer, and M.T Tuley. *Radar Cross Section. Its prediction, measurement and reduction.* Artech House inc., 1985. ISBN 0–89006–174–2.

[47] A.N. Kolmogorov. On the representation of continuous functions of several variables by superposition of continuous functions of one variable and addition. *Doklady Akademii Nauk, USSR*, 114:953–956, 1957.

[48] E.T. Kouba. Recurrent neural networks for radar target identification. Master's thesis. AFIT/GSO/ENG/92d–02, Air Force Inst. of Tech., Wright-Patterson AFB, OH. School of Engineering, 1992.

[49] Mark Lambert, editor. *Jane's all the worlds aircraft, Eighty first edition.* 1990–1991. ISBN 0 7106 0908 6.

[50] L. Leushacke. Range profile analysis and classification techniques. In *Proceedings of the Workshop on "Radar Imaging and Classification Techniques", AC/243 (Panel 10) RSG.12 TP/1", January 28–29, 1993, FGAN, Wachtberg-Werthoven, Germany.*, February 1993.

[51] H.-J. Li, Y.-D. Wang, and L.-H. Wang. Matching score properties between range profiles of high-resolution radar targets. *IEEE Transactions on Antennas and Propagation*, 44(4):444–452, 1996.

[52] S-.H. Li, H.-J. and Yang. Using range profiles as feature vectors to identify aerospace objects. *IEEE Transactions on Antennas and Propagation*, 41(3):261–268, 1993.

[53] H. Ling, R. Chou, and S.W. Lee. Shooting and bouncing rays: calculating the RCS of an arbitrary shaped cavity. *IEEE Transactions on Antennas and propagation*, 37:194–205, 1989.

[54] M.A. Morgan. Target I.D. using natural resonances. In *IEEE potentials*, pages 11–14, 1993.

[55] M.T. Musavi, W. Ahmed, K.H. Chan, K.B. Faris, and D.M. Hummels. On the training of radial basis functions classifiers. *Neural Networks*, 5:595–603, 1992.

[56] J.P. Myles and D.J. Hand. The multi-class metric problem in nearest neighbour discrimination rules. *Pattern Recognition*, 23(11):1291–1297, 1990.

[57] V.G. Nebabin. *Methods and techniques of Radar Recognition.* Artech House, 1995. ISBN 0–89006–719–8.

[58] L.M. Novak. A comparison of 1-D and 2-D algorithms for radar target classification. In *IEEE International Conference on Systems Engineering*, pages 6–12, 1991.

[59] L.R. Pericchi. A Bayesian approach to transformations to normality. *Biometrika*, 68:35–43, 1981.

[60] T. Poggio and S. Edelman. A network that learns to recognize three-dimensional objects. *Nature*, 343:263–266, Januari 1990.

[61] M.J.D. Powell. Radial basis functions for multivariate interpolation: A review. *Algorithms for Approximation*, pages 143–167, 1987.

[62] G. Retzer. Radar signatures of helicopters. In *AGARD–CP–364*, pages 27–1 to 27–11. AGARD, 1985.

[63] G. Retzer. *Zur Detection von Zielen im Clutter unter Ausnutzung von Zielechomodulationen durch rotierende Strukturen.* PhD thesis, Universität der Bundeswehr München, 1986.

[64] L.S. Riggs and C. R. Smith. Bayesian probability theory applied to the problem of radar target discrimination. In *IEEE Antennas and Propagation Society International Symposium*, volume 3, pages 1246–1249, 1992.

[65] A.W. Rihaczek and S.J. Hershkowitz. Man-made target backscattering behaviour: applicability of conventional radar resolution theory. *IEEE Transactions on Aerospace and Electronic systems*, 32(2):809–824, April 1996.

[66] R.A. Ross. Radar cross section of rectangular flat plates as a function of aspect angle. *IEEE Transactions on Antennas and Propagation*, 14(3):329–335, January 1966.

[67] G.T. Ruck, D.E. Barrick, W.D. Stuart, and C.K. Krichbaum. *Radar Cross Section Handbook*, volume 1. Plenum Press, New York, 1970. ISBN 0–306–30343–4.

[68] R.M. Sakia. The Box-Cox transformation technique: a review. *The Statistician*, 41:169–178, 1992.

[69] R.D. Short and K. Fukunaga. The optimal distance measure for nearest neighbour classification. *IEEE Transactions on Information Theory*, 27(5):622–627, September 1981.

[70] R. Silverstein, J.D. and Bender. Measurements and predictions of the RCS of bruderhedrals at milimeter wavelengths. *IEEE Transactions on Antennas and Propagation*, 45(7):1071–1079, July 1997.

[71] Ö. Snorrason and F.D. Garber. Formulation and performance evaluation of adaptive, sequential radar target recognition algorithms. In *SPIE Vol. 1471 Automatic Object Recognition*, pages 116–127, 1991.

[72] B.D. Steinberg. Microwave imaging of aircraft. *Proceedings of the IEEE*, 76(12):1578–1592, 1988.

[73] P.N.R. Stoyle. ISAR review and progress in algorithms for image focussing. In *Proceedings of the Workshop on "Radar Imaging and Classification Techniques", AC/243 (Panel 10) RSG.12 TP/1", January 28–29, 1993, FGAN, Wachtberg-Werthoven, Germany.*, February 1993.

[74] F.T. Ulaby, R.K. Moore, and A.K. Fung. *Microwave Remote Sensing, active and passive. Volume II: Radar Remote Sensing and Surface Scattering and Emission Theory.* Addison-Wesley, 1982.

[75] Ph. van Dorp. Private communication.

[76] V.N. Vapnik. *The nature of statistical learning theory.* Springer Verlag, Berlin, 1995.

[77] J.L. Walker. Range-Doppler imaging of rotating objects. *IEEE AES*, 16(1):23–52, January 1980.

[78] D.R. Wehner. *High-Resolution Radar, 2nd edition.* Artech House, 1994. ISBN 0–89006–727–9.

[79] D.A. Zolnick. Calculating the radar cross section from multiple-bounce interactions. In *IEE International Conference on Computation in Electromagnetics*, pages 193–196. IEE London, Conference Publication 350, 1991.

[80] A. Zyweck and R.E. Bogner. Radar target classification of commercial aircraft. *IEEE Transactions on Aerospace and Electronic systems*, 32(2):598–606, April 1996.

Samenvatting

In 1988 werd in de Perzische Golf een Iraanse Airbus neergeschoten vanaf het Amerikaanse oorlogsschip *USS Vincennes*. Hierbij kwamen 298 passagiers om het leven. De oorzaak van deze ramp was de foutieve herkenning van het toestel als een F14 gevechtsvliegtuig.

Dit ongeval staat niet op zichzelf. Sinds vliegtuigen worden ingezet in oorlogs- of crisissituaties heeft de herkenning van vliegtuigen problemen opgeleverd. Waren deze in het begin niet urgent — vliegtuigen vlogen immers met relatief lage snelheden en ze droegen grote tekens die aangaven uit welk land ze kwamen — naarmate de toestellen steeds meer op elkaar gingen lijken en zich alsmaar sneller konden verplaatsen, werd het uiterst moeilijk ze tijdig op het zicht te herkennen.

In dit proefschrift is onderzocht in hoeverre *radar* gebruikt kan worden voor de herkenning van vliegtuigen. Een gunstige eigenschap van dit instrument is de grote afstand waarop objecten gezien kunnen worden. Ook werkt het zowel bij dag als bij nacht en onder vrijwel elke weersconditie. Het principe is eenvoudig: de radar zendt een geconcentreerde hoeveelheid energie uit in een bepaalde richting. Deze energie wordt door een een object gereflecteerd. Dit object kan een vliegtuig zijn, maar bijvoorbeeld ook een schip of voertuig. De weerkaatsing wordt ontvangen door dezelfde radar zodat gemeten kan worden wat de richting van en afstand tot het object is. Hiermee belanden wij bij de twee belangrijkste functies van een radar. Ten eerste dient met behulp van het instrument het antwoord te worden verkregen op de vraag: "Is er iets aanwezig?" en, ten tweede, indien het antwoord bevestigend is: "Waar is het?"

Een voor de hand liggende derde vraag die vervolgens beantwoord zou moeten worden is: "Wat is het?". Alhoewel de radar al sinds de Tweede Wereldoorlog operationeel wordt gebruikt, krijgt het onderzoek naar deze laatste vraag pas de laatste decennia aandacht. De reden hiervoor is dat pas in de loop van de zestiger jaren de eerste radars werden gebouwd die voldoende detail van een object kunnen waarnemen.

De belangrijkste methoden om met behulp van radar een "vingerafdruk" of signatuur van een vliegtuig te meten ten behoeve van classificatie worden belicht in hoofdstuk 2. Aan het type signatuur kunnen eisen worden gesteld ten aanzien van de operationele bruikbaarheid.

Vooreerst dient het signatuur snel gemeten te kunnen worden, het liefst op grote afstand en bovendien onafhankelijk van de bewegingen van het vliegtuig. Vervolgens is het van belang dat de herkenning kan plaatsvinden, ongeacht de oriëntatie van het vliegtuig. Het type signatuur dat deze wensen het dichtst benadert is het zogenaamde *radarafstandsprofiel*. Een dergelijk profiel is feitelijk niets anders dan een grafiek die laat zien op welke afstanden reflectiehaarden op het vliegtuig aanwezig zijn.

In hoofdstuk 3 wordt uitgebreid ingegaan op het meten en de eigenschappen van afstandsprofielen. In dit hoofdstuk worden tevens de twee belangrijkste kenmerken van deze signaturen voor de bouw van een vliegtuigherkennend systeem benadrukt. Profielen zijn afhankelijk van de geometrie van het vliegtuig — het is nu juist dit gegeven dat ons in staat stelt de profielen te gebruiken voor de vliegtuigherkenning. Daarnaast is de vorm sterk afhankelijk van de oriëntatie van het vliegtuig. Kleine verdraaiingen van het toestel kunnen resulteren in sterk verschillende profielen.

Dit laatste gegeven vormt een groot probleem voor operationele toepassingen — het is zeer duur en tijdrovend om, van alle relevante vliegtuigen, voldoende profielen te verkrijgen in meetcampagnes. Hierbij moet gedacht worden aan ettelijke duizenden profielen voor één vliegtuig. Bovendien zullen, uit de aard van de zaak, geen vijandelijke vliegtuigen deelnemen aan de metingen. Een oplossing wordt gezocht in computersimulaties waarin geometrische computermodellen van vliegtuigen gebruikt worden en een rekenprogramma dat de weerkaatsing op het vliegtuig nabootst.

Voor het onderzoek staan twee typen van afstandsprofielen tot onze beschikking: metingen aan vliegtuigen in volle vlucht en computerberekeningen aan modellen. Een uitgebreide beschrijving van deze gegevens alsmede de opzet van de experimenten in dit proefschrift kan worden gevonden in hoofdstuk 4.

In hoofdstuk 5 wordt aangetoond dat het daadwerkelijk mogelijk is om vliegtuigen te herkennen op basis van radarafstandsprofielen. Hierbij worden voor zowel het trainen van het herkenningssysteem als het testen gemeten afstandsprofielen gebruikt. Diverse classificatietechnieken worden beschreven, toegepast en vergeleken — hier wordt niet alleen de classificatie-nauwkeurigheid in aanmerking genomen, maar

ook de snelheid waarmee een herkenning wordt geleverd.

In hoofdstuk 6 wordt een eenvoudige schaling, de zogenaamde *Box-Cox transformatie*, toegepast op de profielen. Aangetoond wordt dat hiermee een grote winst in classificatie-prestaties wordt behaald. Tevens wordt aannemelijk gemaakt dat er voor deze techniek vele andere toepassingen zijn. Het succes van de transformatie ligt, kort gezegd, in de "aard der dingen": hoe groter de gemiddelde waarde van een aantal metingen van gelijksoortige objecten, hoe groter de spreiding rond dit gemiddelde. Een voorbeeld van deze "aard" is de variatie in lichaamslengte: deze zal groter zijn voor volwassen Nederlanders dan voor, bijvoorbeeld, volwassen Japanners, simpelweg omdat de laatsten gemiddeld kleiner zijn.

In hoofdstuk 5 en 6 zijn, zowel voor de training als voor het testen van de classificatietechnieken, gemeten profielen gebruikt. Al eerder is genoemd dat radarmetingen alléén niet voldoende zijn om een volledige bibliotheek van afstandprofielen te verkrijgen voor de herkenning van alle relevante doelen. In hoofdstuk 7 en 8 is daarom onderzocht in hoeverre computersimulaties behulpzaam kunnen zijn voor het completeren van het bestand van profielen.

Allereerst is het van belang om te onderzoeken of de berekeningen en metingen voldoende overeenkomen. In hoofdstuk 7 wordt hiertoe in detail een vergelijking uitgevoerd. De resultaten tonen aan dat vooral op zij-aanzichtshoeken een uitstekende overeenkomst bestaat tussen metingen en berekeningen. De correlaties worden slechter indien de radar tegen de neus of de staart van het vliegtuig "aankijkt". De verschillen kunnen verklaard worden uit het feit dat de radarstraling door de motoren wordt gereflecteerd hetgeen in de gebruikte rekenprogrammatuur niet is geïmplementeerd.

In hoofdstuk 8 wordt een verdere stap gezet: de classificatie van de gemeten profielen welke geschiedt op basis van de *gesimuleerde* profielen. De experimenten in dit hoofdstuk worden uitgevoerd op een verzameling gemeten profielen van vijf verschillende verkeersvliegtuigen. Tijdens deze metingen waren de vliegtuigen met hun zijkant naar de radar gericht. Van dezelfde toestellen zijn geometrische computermodellen aanwezig waarvan profielen zijn berekend, uiteraard ook op zijaanzichtshoeken.

De resultaten tonen aan dat het in beginsel mogelijk is om de

vliegtuigen te discrimineren op basis van de computerberekeningen. Een aanzienlijke verbetering hierbij blijkt te worden behaald indien meerdere metingen van hetzelfde vliegtuig simultaan voor herkenning worden gebruikt. Overigens lijkt het ons raadzaam om verder te onderzoeken of kleine details uit de profielen kunnen worden weggelaten — de resultaten zoals beschreven in dit hoofdstuk tonen vooralsnog aan dat hiermee een verbetering van de classificatie-score kan worden behaald.

Veruit de belangrijkste conclusies van dit proefschrift zijn dat herkenning op basis van radarafstandsprofielen zeer wel mogelijk is en en dat computersimulaties een mogelijke oplossing van het probleem van het structurele gebrek aan gemeten afstandsprofielen vormen. Voor de uiteindelijke toepassing in radarsystemen ligt er niettemin nog een groot, maar uitdagend en zeer intrigerend onderzoeksgebied braak. Onze verwachting is echter dat vliegtuigherkenning op basis van afstandsprofielen een wezenlijk onderdeel zal vormen van toekomstige operationele radarsystemen.