

FFIE/805/113

Godkjent
Kjeller 7 mai 2003



Torleiv Maseng
Forskningsjef

**METODER FOR PULSSORTERING OG
EMITTERGJENKJENNING I ESMEX**

SKARTLIEN Roar, SUNDGOT Rune, ØYEHAUG Leiv

FFI/RAPPORT-2003/00881


FORSVARETS FORSKNINGSINSTITUTT
Norwegian Defence Research Establishment
Postboks 25, 2027 Kjeller, Norge

FORSVARETS FORSKNINGSPINSTITUTT (FFI)
Norwegian Defence Research Establishment

UNCLASSIFIED

P O BOX 25
 NO-2027 KJELLER, NORWAY
REPORT DOCUMENTATION PAGE

SECURITY CLASSIFICATION OF THIS PAGE
 (when data entered)

1) PUBL/REPORT NUMBER FFI/RAPPORT-2003/00881	2) SECURITY CLASSIFICATION UNCLASSIFIED	3) NUMBER OF PAGES 49
1a) PROJECT REFERENCE FFIE/805/113	2a) DECLASSIFICATION/DOWNGRADING SCHEDULE -	
4) TITLE METODER FOR PULSSORTERING OG EMITTERGJENKJENNING I ESMEX METHODS FOR DEINTERLEAVING AND EMITTER CLASSIFICATION IN ESMEX		
5) NAMES OF AUTHOR(S) IN FULL (surname first) SKARTLIEN Roar, SUNDGOT Rune, ØYEHAUG Leiv		
6) DISTRIBUTION STATEMENT Approved for public release. Distribution unlimited. (Offentlig tilgjengelig)		
7) INDEXING TERMS IN ENGLISH:		IN NORWEGIAN:
a) <u>Electronic Support Measures (ESM)</u>		a) <u>Elektroniske støttetiltak</u>
b) <u>Deinterleaving</u>		b) <u>Pulssortering</u>
c) <u>Specific Emitter Identification</u>		c) <u>Spesifikk emitteridentifikasjon</u>
d) <u>Emitter databases</u>		d) <u>Emitterdatabaser</u>
e) _____		e) _____
THESAURUS REFERENCE:		
8) ABSTRACT This report describes methods for deinterleaving and emitter classification developed in FFI-project 805 "PROFIL, Luftkommando- og kontrollsystem". An experimental ESM system, ESMEX, has been developed and tested on civilian and military emitters. The hardware components of ESMEX include antennas, tuners, and sampling devices. Software components include deinterleaving, pulse feature extraction/compression and emitter databases. Implemented methods for deinterleaving and emitter classification are described, as well as current test results and plans for further development.		
9) DATE 7 May 2003	AUTHORIZED BY This page only  Torleiv Maseng	POSITION Director of Research

UNCLASSIFIED

ISBN 82-464-0738-4

SECURITY CLASSIFICATION OF THIS PAGE
 (when data entered)

FFI-B-22-1982

INNHOOLD

	Side	
1	INNLEDNING	7
2	PROGRAMVAREKOMPONENTER I ESMEX	7
3	BETRAKTNINGER OM PULS-SORTERING I ESM	8
3.1	Generelt emitterscenario	8
3.2	Problemdefinisjon for puls-sortering	8
3.3	Problemløsning med AOA	9
3.4	Problemløsning uten AOA	9
3.5	To alternative løsninger uten AOA	10
4	PULS-SORTERING I ESMEX	10
4.1	Problemdefinisjon	10
4.2	Problemløsning	10
4.3	Egenskaper ved clustring	12
4.4	PDW-rate fra kanalkortet.	12
5	EDW-BEREGNING I ESMEX	13
6	CLUSTRING OG PRI MØNSTERGJENKJENNING	14
6.1	Hierarkisk clustring	14
6.1.1	"K-means" iterativ clustring	14
6.1.2	Adaptiv oppløsning ved hierarkisk clustring	15
6.1.3	Initialtilstand (frøpunkter) fra glattet histogram	16
6.2	PRI mønstergjenkjenning	16
6.2.1	Mønstergjenkjenning med andre ordens differanse histogram	16
7	PULS-SORTERING PÅ DELVIS OVERLAPPENDE EMITTERE	17
8	IDENTIFIKASJON AV EMITTERINDIVIDER	20
8.1	UMOP	20
8.2	Pulsdeteksjon og preprosessering	21
8.2.1	Deteksjon av puls	21
8.2.2	Beregning av amplitude og fase	22
8.2.3	Midling over pulser	23
8.3	Pulskomprimering ved ortonormal approksimasjon	23
8.4	Statistiske metoder for ortonormale representasjoner	24
8.4.1	Statistikk for pulsrepresentasjoner	25
8.4.2	Fordelingsfunksjoner for avstanden når $\tilde{f} = f + \text{støy}$	25
8.4.3	Fordelingsfunksjoner for avstanden når $\tilde{f} = h + \text{støy}, h \neq f$	26

8.5	Multipath	27
8.6	Resultater	27
8.7	Konklusjoner og utfordringer	29
9	EMITTERGJENKJENNING VED DATABASEKORRELASJON	30
9.1	Databaseinnhold	31
9.2	Intern databasestruktur	32
9.3	Oppdatering	33
9.4	Korrelasjon	35
9.4.1	Avstandsmål	35
9.4.2	Overlappende klasser	37
9.4.3	Modifisert match	39
9.5	Oppsummering og resultater	41
	APPENDIKS	43
A.1	Parametersett i LABVIEW-koden	43
A.2	Sortering på amplitude	43
A.3	Legendre-polynomer	44
A.4	Sentrale og ikke-sentrale kji-kvadrat-fordelinger	44
A.5	Emittergjenkjenning	45
A.6	Følsomhet for multipath	46
	LITTERATUR	48
	Fordelingsliste	49

METODER FOR PULSSORTERING OG EMITTERGJENKJENNING I ESMEX

1 INNLEDNING

Prosjekt 805 PROFIL ved FFI har bygget opp første versjon av et ESM-eksperimentsystem, ESMEX. ESMEX består foreløpig av en sensor (med antenne, mottaker, digitalisering, puls-sortering og pulskomprimering), samt programvare og databaser for emittergjenkjenning. Den siste delen av funksjonskjeden kjøres på en egen "AnalysePC" tilkoblet sensoren via ethernet (TCP/IP).

En tidlig spesifisering av ESMEX finnes i (1). Kanalkortet (DIPP – Digital Intermediate frequency Pulse Processor), som er en sentral komponent i sensoren, er beskrevet i (2). I (3) gis en oppsummering av aktuelle metoder for puls-komprimering og metoder for emittergjenkjenning. For en mer overordnet oversikt over ESMEX og hele PROFIL-prosjektet henvises det til sluttrapporten (15). Denne rapporten er en konkretisering av (3), utvidet med beskrivelse av puls-sortering, og foreløpige testresultater/evaluering av implementerte metoder. Vi har testet ESMEX med data fra værradarer på sivile passasjerfly og F16-radar.

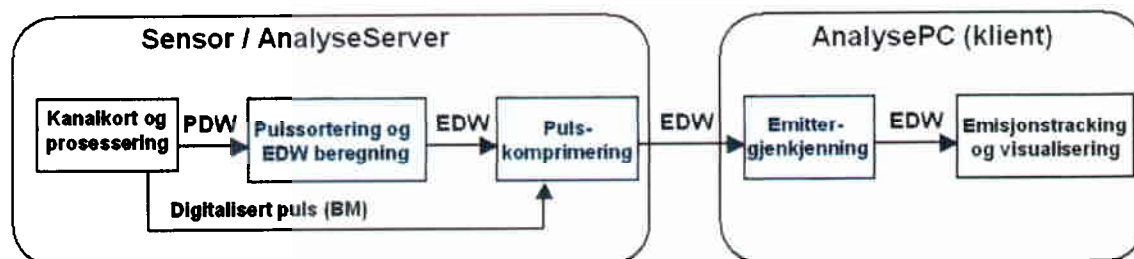
2 PROGRAMVAREKOMPONENTER I ESMEX

Etter digitalisering av pulser beregner DIPP-kortet pulsbeskrivelsesord (PDW - Pulse Description Word). Funksjonskjeden deretter kan deles inn i:

- Puls-sortering og EDW-beregning
- Pulskomprimering
- Emmittergjenkjenning ved databasekorrelasjon
- Emisjonstracking og visualisering

Puls-sortering forsøker å skille fra hverandre pulser fra forskjellige emittere. Etter pulssortering beregnes middelerverdier og standardavvik for en rekke parametere som beskriver et pulstog. Disse parameterne samles i et emisjonsbeskrivelsesord (EDW – Emission Description Word). Pulskomprimering utvider EDWen med nye parametere (koeffisienter) som koder for midlet pulsform gjennom pulstog. Emmittergjenkjenning sammenligner EDW-parametere med databaser for å avgjøre hvilken emittertype og emitterindivid som er opphav til EDWen. Navn på gjenkjente databaseelementer legges til EDWen. Siste ledd i funksjonskjeden er emisjonstracking som går ut på å gruppere sammen EDWer fra samme emitter og å følge emittere over tid.

De neste kapitlene beskriver de tre første punktene ovenfor. Det siste punktet vil ikke beskrives nærmere fordi endelig valg av metode for emisjonstracking ennå ikke er avklart. Foreløpig implementert metode er basert på vinkelangivelsen fra sensoren, men kvaliteten på denne parameteren er ikke god nok. Siden pulser også mottas i antennens sidelover er det ikke antennens pekeretning vi er interessert i, men retning til emitteren. Videreutvikling av emisjonstracking avhenger av pågående arbeid med metoder for retningsbestemmelse.



Figur 2.1 Skisse av funksjonskjeden etter digitalisering og PDW-beregning i kanalkortet. Hvert av leddene som produserer EDW'er legger til nye parametere.

For hvert enkelt ledd i funksjonskjeden i Figur 2.1 vil oppsummering av resultater, erfaringer og planer for videre arbeid finnes i de respektive kapitler.

3 BETRAKTNINGER OM PULS-SORTERING I ESM

3.1 Generelt emitterscenario

I et generelt scenario består mottatt signal på ESM-sensoren av overlappende pulstog fra flere emittere. Pulsene fra forskjellige emittere er "stokket" i hverandre ("interleaved" på engelsk), men vi antar at pulsene ikke er fysisk overlappende. Målet for en puls-sorterer er å trekke ut ("de-interleave") grupper av pulser som tilhører samme emitter.

Generelt sett kan pulsens bærefrekvens (f) variere gjennom pulstog, og PRI mønsteret kan karakteriseres som *konstant PRI*, *Stagger*, eller *Switched Dwell*. For å sortere pulstog mhp emitter, er det naturlig å ta utgangspunkt i de signaturene som er emitterspesifikke. Hverken frekvens eller PRI er kontrollerbare fra ESM sensorens side, og kan betraktes som vilkårlige variable uten forhånds- eller tilleggsinformasjon. De størrelser som ikke er vilkårlige (men emitterspesifikke) er den romlige posisjonen, emitterens grunnfrekvens (eventuelt krystall-oscillator), og emitterens pulsform.

3.2 Problemdefinisjon for puls-sortering

Observerbare størrelser er normalt sett PDW-data av typen TOA (Time Of Arrival), f , A (Amplitude), pulsform (instantan amplitude og fase) og eventuelt AOA (Angle Of Arrival). Puls-sorteringsproblemet defineres på grunnlag av en undergruppe av de observerbare størrelsene.

Vi antar at (emitterspesifikke) pulsformdata ikke er tilgjengelig i utgangspunktet, før vi har sortert pulsene. Generelt sett må man midle over flere pulser fra samme emitter for å oppnå et sikkert estimat av pulsform i situasjoner der signal/støyforholdet er lavt (for eksempel støy-RMS på mer enn 5% av pulsamplituden). Dette medfører at puls-sortering må utføres før midling av pulsformdata kan finne sted. Det er derfor naturlig å anta at "sikre" pulsformdata ikke er tilgjengelig før puls-sortering.

Derimot antar vi at PDW-dataene TOA, f , A , og eventuelt AOA er tilgjengelig for sortering. På grunnlag av denne undergruppen er målet å trekke ut grupper av pulser som tilhører samme emitter.

Hvis AOA er tilgjengelig, har vi delvis løst opp puls-sorterings problemet, hvis ikke flere emittere ligger langs samme synslinje. Vi må derfor skille mellom puls-sortering med og uten AOA, men samtidig være klar over at puls-sortering kan være nødvendig i begge tilfeller.

3.3 Problemløsning med AOA

Vi antar at vi har en direktiv antenne (eller et fasestyrt antenne-array) med stort gain (smal hovedlobe og lave sidelober) i en styrbar vinkel (elevasjon og asimut), slik at vinkelopløsningen er "god" (for eksempel 1-2 grader). Sannsynligheten for at to emittere befinner seg i samme retning er naturligvis mindre jo mer direktiv antennen er, forutsatt at emitterne har en homogen romlig fordeling. Dermed kan det argumenteres for at puls-sorteringen kan utføres på AOA alene. Ulempen med roterende direktiv antenne er lavere POI (Probability of Intercept). Alternativt til direktiv antenne kan man bruke fasedifferansemålinger i data fra et antenne-array med to eller flere antenner som i FFI-prosjekt KREST (14).

3.4 Problemløsning uten AOA

I nåværende ESMEX er ikke AOA estimert nøyaktig nok (grunnet bred hovedlobe og markante sidelober i antennediagrammet), og vi må betrakte puls-sorterings problemet fra et deinterleavings-synspunkt. Det er sannsynligvis problematisk å formulere en generell deinterleavings-prosedyre uten å splitte opp et generelt scenario i flere delproblemer (eller spesifisere hvilket scenario man skal "løse opp"). Man kan tenke seg en hierarkisk deinterleaver der de "ytre lag" bygger på enkle sorteringskriterier basert på PDW-ord. Først løser en opp enkle bestanddeler av signalet, for deretter å løse opp mer komplekse sider ved signalet i de neste "lagene" av deinterleaveren.

En slik tankegang er beskrevet i (7), der man sorterer først på PDW parametere (hvis mulig) før puls-sorteringen foretas ved hjelp av TOA-deinterleaving. Deinterleaving gjør utstrakt bruk av histogrammer basert på TOA tidsdifferanser, for å finne sannsynlige PRI-nivåer (7,8,9,10,11). Dette danner grunnlaget for et sekvensielt søk etter bestemte PRI-sekvenser. Finnes disse sekvensene, dannes et nytt histogram som initierer et nytt sekvensielt søk, osv.

3.5 To alternative løsninger uten AOA

- Normalt sett vil PRI-mønsteret gjentas med en grunnperiode T , og to forskjellige emittere vil prinsipielt ha forskjellig periode T (eller midlere T hvis vi tar hensyn til jitter). To eller flere serier med ulik periode T (og vilkårlige PRI-mønstre) bør kunne skilles fra hverandre så lenge minst to perioder fra hver emitter finnes i pulstoget (slik at repetisjonsintervallene kan påvises). En relatert metode som bruker Fourierspektra er beskrevet i (10).
- Amplitude (A) fra en emitter har normalt sett et tidsforløp med kontinuitet i amplitude og kontinuitet i tid (se appendiks A.2). To eller flere overlappende pulstog fra forskjellige emittere bør derfor kunne skilles på individuelle trekk i amplitude som funksjon av tid $A(t)$ (12). Det er imidlertid observert at frekvenshopping medfører "amplitudehopping"/amplitudesprang når bærefrekvensen skifter, på grunn av emitterens frekvensrespons. Slike effekter må derfor betraktes i en eventuell mønstergjenkjenningsprosedyre for amplitude.

4 PULS-SORTERING I ESMEX

4.1 Problemdefinisjon

- Vi antar at emittere har konstant frekvens over begrensede tidsintervaller (flere pulser etter hverandre har samme frekvens), men med en generell usikkerhet uttrykt i et standardavvik.
- Vi antar delvis frekvensoverlapp mellom de ulike frekvensene (overlappende sannsynlighetsfordelinger).
- Vi antar et vilkårlig PRI mønster, med en generell usikkerhet (jitter).

4.2 Problemløsning

Det er nå tilstrekkelig å sortere og karakterisere pulstogene på følgende måte:

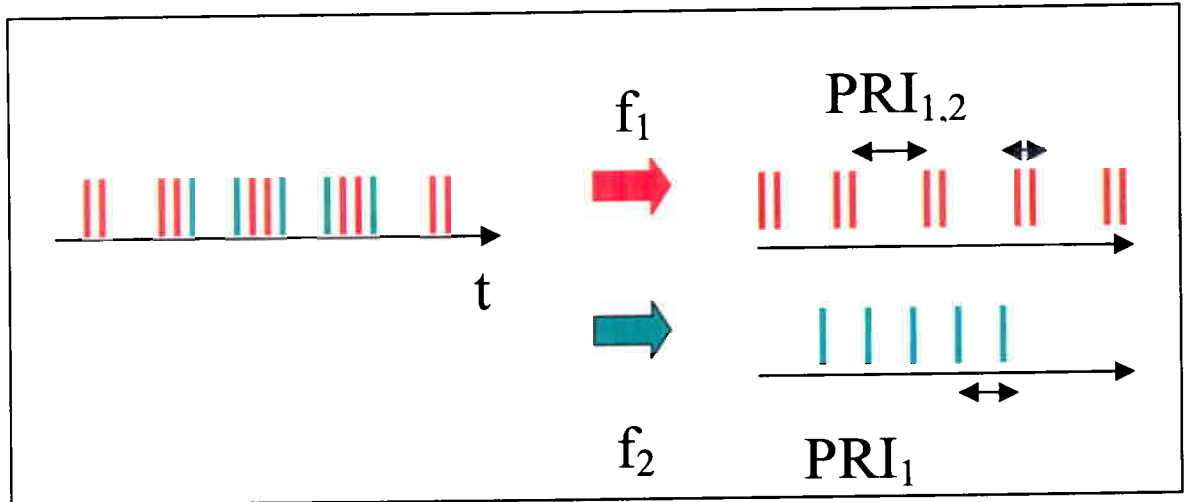
- Vi sorterer først pulsene med hensyn på frekvens
- Deretter beregner vi PRI innenfor hver frekvensgruppe

Den samme clustringsmetoden brukes i begge tilfeller ovenfor. Vi merker oss at puls-sorteringen utføres ved clustring av frekvensinformasjonen i PDW dataene (se Figur 4.1 og Figur 4.2). Pulssorteringen i ESMEX kan derfor betraktes som "første lag" i en mer generell deinterleaver som etter hvert kan bygges ut (11). Vår forenklete puls-sorterer genererer imidlertid korrekte EDW-data for ca 95% av alle emittersignaler målt i luftrommet omkring FFI i siste halvdel av 2002.

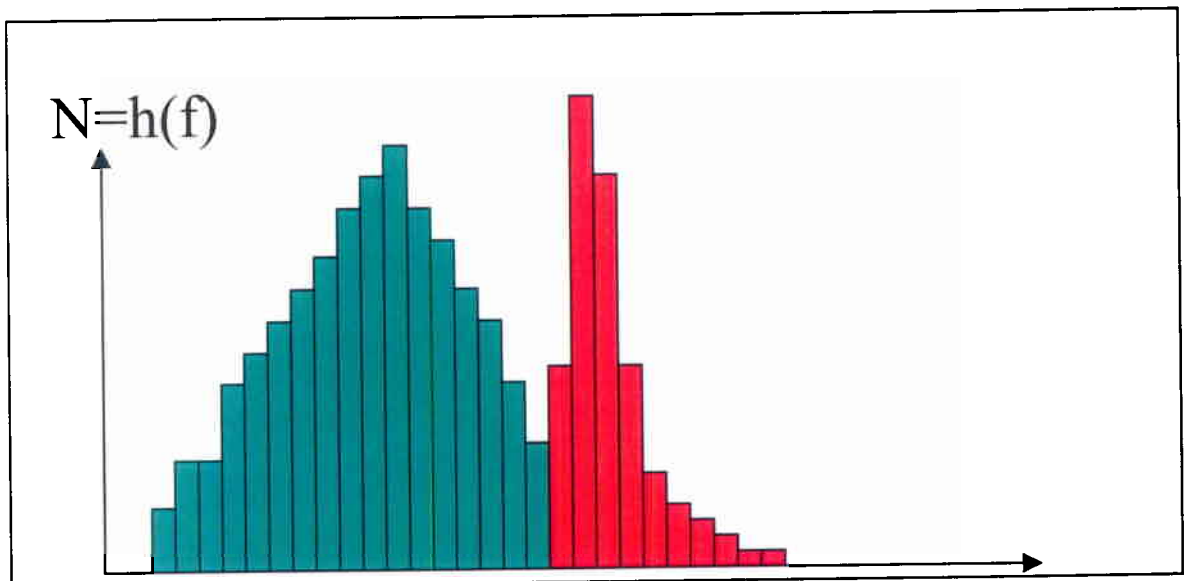
TOA brukes her bare i etterhånd når PRI beregnes. Vi merker oss at todimensjonal clustring i (f , PRI)-planet ikke er mulig, da PRI kun kan beregnes *etter* puls-sortering i vårt tilfelle.

Hovedegenskapen ved clustring er at vi behandler alle tilgjengelige data samtidig, på statistisk vis. Et cluster kan betraktes som en naturlig gruppering av data som "hører sammen", for

eksempel tilkjennegjort ved et lokalt maksimum i histogramfordelingen til dataene. Clustrings-metoden har den fordel at vi kun bruker en enkel "kjernealgoritme" som kan optimaliseres med hensyn på CPU-tid. Algoritmen er også fleksibel for utvidelse med hensyn på tidsstempling av "eventer" i et pulstog (for eksempel tidspunktet for dwell-skift, eller tidspunktet for en spesifikk stagger kombinasjon). Clustrings-proseduren er beskrevet i detalj i kapittel 6.1.



Figur 4.1 Enkel puls-sortering på frekvens



Figur 4.2 Overlappende frekvensfordelinger, vist ved frekvenshistogram. Rød og grønn viser de to resulterende cluster. I rødt cluster vil punkter fra grønt cluster også finnes, men de er relativt få (og vice versa). Dette kan betraktes som støy. Korrekte PRI-nivåer kan likevel finnes for rødt cluster, hvis antall grønne støypunkter er relativt få.

4.3 Egenskaper ved clustring

For sortering på frekvens kan man generelt konkludere med at:

- Når vi har frekvensoverlapp mellom to clusterer, vil korrekte PRI nivåer oppnås statistisk sett, hvis relativt få pulser til naboemitteren blandes sammen med pulsene i clusteret fra den aktuelle emitteren (se Figur 4.2).
- Hvis frekvensfordelingen til en emitter ikke har et veldefinert maksimum, kan tilhørende pulser bli inkludert i clusteret til en overlappende emitter. Man kan da risikere at ett eller ingen av PRI-nivåene fra de to emitterne blir beregnet korrekt.
- Hvis en emitter har en bi-modal (eventuelt multi-modal) frekvensfordeling kan man risikere feilberegning av emitterens PRI-nivåer. Slike fordelinger er imidlertid svært sjeldent observert.

Hvis en emitter bruker et bestemt sett frekvenser under emisjonen (frekvenshopping), vil ikke bare ett, men flere diskrete frekvensclusterer representere emitteren. I dette tilfellet kan det være mulig å assosiere *et sett* frekvensclusterer til en emitter hvis vi benytter tidsinformasjon:

- Hvis to overlappende emittere hopper i frekvens relativt sjelden (flere etterfølgende pulser har samme frekvens), kan de skilles fra hverandre ved å identifisere PRI-mønstre innenfor hvert frekvenscluster.
- PRI beregning innenfor hvert frekvenscluster gir ikke korrekte PRI-verdier om emitterne skifter frekvens fra puls til puls. I dette tilfellet kan vi ikke bruke frekvensbasert puls-sortering.

PRI-clustring:

- Erfaringen har vist at vi må kreve mer enn ca. 30 pulser per PRI-nivå, for statistisk stabilitet.
- PRI-clustring er ikke følsom for manglende pulser. Dette genererer clusterer med middel PRI lik et multiplum av korrekt PRI, men slike clusterer inneholder relativt få tellinger.
- PRI-clustring er ufølsom for jitter. Ved økende jitter blir tilhørende cluster "bredere" (større standardavvik).
- PRI-clustring blir statistisk ustabil hvis få pulser kombineres med stort jitter, slik at clusterer ikke blir veldefinert.

4.4 PDW-rate fra kanalkortet.

I sanntid leses det normalt inn en samling PDWer fra kanalkortet (per kall fra software). Lengden L på PDW bufferet kan settes av brukeren. Hvor ofte bufferet leses (τ_o) kan også bestemmes av brukeren. For et pulstog med gitt PRI, vil kriteriet for å lese alle innkommende PDW uten avbrudd være gitt ved (13):

$$N_r = \frac{\tau_o + \tau_c}{PRI} \leq L \quad (4.1)$$

Her er N_r lik antall PDWer som mottas innenfor tidsintervallet $\tau = \tau_o + \tau_c$. Vi merker oss at τ har en nedre grense satt av syklustiden til softwaren τ_c :

$$\tau > \tau_c \quad (4.2)$$

Uten henting av ”burst”-data for pulsformberegninger, er $\tau_c \sim 50 \text{ ms}$, og omtrent dobbelt så stor med burst-prosessering. Denne tiden øker monotont med antall prosesserte pulser N_T i foregående pulstog. Om deler av pulstoget forkastes, kan en begrense syklustiden τ_c .

Vi merker oss at i et tett signalmiljø (lav effektiv PRI), er minst τ_c / PRI pulser tilgjengelig i bufferet. Vi konkluderer med at vi kan forvente mange pulser i bufferet i et ”tett signalmiljø”, selv om utlesningstiden τ_o settes til et minimum. Om syklustiden er ”for lang”, kan en risikere at bufferet blir overskrevet, slik at vi mister PDW-data. Antall tilgjengelige pulser er i de fleste tilfeller sannsynligvis nok til at en statistisk behandling av dataene er mulig.

Ved innlesing av PDWer benyttes API-funksjonen ”GetRealTimePDWs”. I kallet oppgis L , og hvilken av de to kanalene på kanalkortet en ønsker å få data fra. Videre må en definere et buffer som definerer hvilke PDW-data som skal hentes ut (se spesifikasjonene til kanalkortet). Foruten PDW-data returnerer API-funksjonen antall nye PDWer som er samlet inn av kanalkortet siden forrige gang en leste APIen.

I nåværende form er utlesningssoftwaren konstruert slik at vi søker etter å gruppere PDW-dataene (før de sendes videre til puls-sortering), hvis PDWene hører til et sammenhengende pulstog. Man tester om APIen returnerte PDWer ved forrige lesing. Ved tilfeller hvor APIen returnerer nye PDWer i to etterfølgende kall, antar vi at pulsene fra disse to bufferne hører til et pulstog og slår dem sammen til en datablokk.

Det er lagt inn en maksimum grense på 5 PDW-buffer. Det vil si at om det tas inn PDWer i fem kall på rad, sendes datablokken videre (med lengde $5L$) for puls-sortering. Om pulsdata mangler for et kall sendes dataene videre i en blokk med lengde mindre enn $5L$.

5 EDW-BEREGNING I ESMEX

Etter puls-sortering beregnes et beskrivelsesord (EDW) for hvert pulstog (tilsvarende et bestemt frekvenscluster). På grunnlag av vår antagelse, vil et pulstog komme fra en bestemt (men fortsatt ukjent) emitter.

Vi midler alle PDW parametere gjennom pulstoget, samt beregner standardavvik. Middel PRI og tilhørende standardavvik beregnes innenfor hvert PRI-cluster. I tillegg detekteres PRI-mønsteret, som beskrevet i avsnitt 6.2. En oversikt over EDW-dataene finnes i nedenforstående Tabell 5.1. Disse dataene går videre til databasen, for klassifikasjon og identifikasjon.

Feltnavn	Beskrivelse
F	Bærefrekvens [GHz]
PRI	Vektor med alle PRI [ms] ([pri1 pri2 ...])
PRI-mønster	Typenummer for PRI-mønster (1=konstant, 2= stagger, 3= switched dwell)
Tidsinformasjon	Tidspunkt for første PDW. Vektor med tidspunkt for "pri-eventer" [absolutt tid (UT)], for eksempel dwell-skift, kan legges inn.
PW	Pulsbredde [ms]
A	Pulsamplitude [dBm]
AOA	Vinkel [grader] (foreløpig ikke nøyaktig nok)
Pulsform-koeffisienter	Vektor med polynomkoeffisienter for amplitude og fase. Beskriver midlere pulsform.
Multipath ID	Indikator for puls kvalitet, mhp multipath-forvrengning (skal implementeres).
Standardavvik for alle størrelser	Standardavvik gjennom pulstoget, for alle størrelser over. For å estimere usikkerheten (gjennomsnittlig avvik fra middelerdi).

Tabell 5.1 EDW-data. AOA og pulsform regnes som emitter-spesifikke. Alle størrelser er middelerdier over pulstoget, samt tilhørende standardavvik. Multipath ID er en indikator som viser i hvilken grad pulsene er forvrengt av multipath effekter. Metoder for denne beregningen er på utviklingsstadiet.

6 CLUSTERING OG PRI MØNSTERGJENKJENNING

6.1 Hierarkisk clustering

Clustrings-proseduren må løse opp alle naturlig forekommende clusterer automatisk. Dette gir opphav til hierarkisk clustering, som foregår ved at man utfører "sub-clustering" et vilkårlig antall ganger innenfor foregående beregnede clusterer, inntil man har løst opp alle clusterer (Figur 6.1). Det viser seg at hierarkisk clustering er absolutt nødvendig på generelle PRI-data der vi har stor PRI-variasjon innenfor modene *stagger* eller *switched dwell*.

6.1.1 "K-means" iterativ clustering

Vi bruker en relativt enkel form for clustering. Denne er basert på tyngdepunktsberegninger, og et nærmeste nabo kriterium på følgende måte:

1. Velg et sett med tyngdepunkter C_j som initialtilstand
2. Tilordne datapunktet Y_i til C_j basert på nærmeste nabo: $MIN \{ |Y_i - C_j| \}$
3. Tyngdepunkter som ikke er nærmeste nabo til noe datapunkt, fjernes
4. Oppdater alle tyngdepunkter C_j , ved å midle datapunktene over tilhørende cluster j
5. Gå til punkt 2, eller avslutt om tyngdepunktene har konvertert

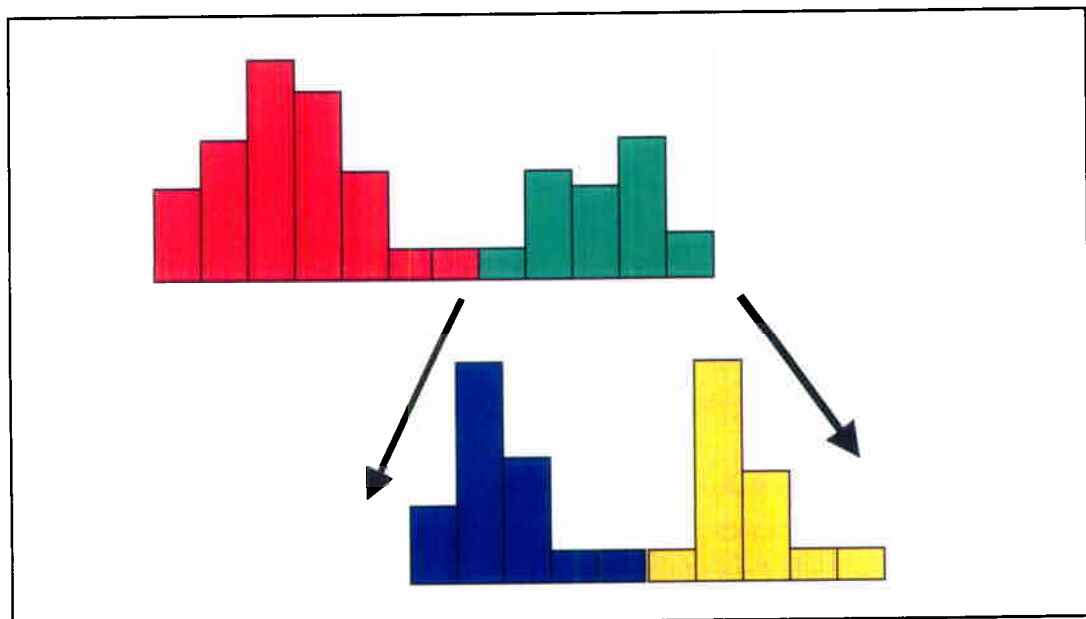
Denne formen for clustering kalles "K-means clustering". Mer informasjon finnes i (6). Iterasjonen flytter tyngdepunktene C_j til koordinater nær lokale maksima i fordelingsfunksjonen til dataene Y (der tettheten av dataene er stor). Eventuelt overflødige initialpunkter blir fjernet.

Til hvert cluster har man array-indeks over tilhørende datapunkter. På denne måten er det trivielt å adressere en hvilken som helst PDW-parameter som tilsvarer et bestemt cluster (frekvens eller PRI). For eksempel vil alle PDW data som tilhører et bestemt frekvenscluster kunne behandles separat, for eksempel når vi beregner PRI. Et annet eksempel er å behandle et bestemt PRI-dwell separat.

6.1.2 Adaptiv oppløsning ved hierarkisk clustering

En tilsynelatende svakhet ved K-means clustering er at naturlig forekommende clusterer som ligger nærmere hverandre enn tettheten til initialpunktene kan bli slått sammen til ett cluster. Dvs, man kan ikke vite på forhånd om oppløsningen er tilstrekkelig. En naiv løsning på problemet er da å øke antall initialpunkter, på bekostning av drastisk økende prosesseringstid, og fortsatt uten garanti for at vi løser opp alle naturlige clusterer. En bør i stedet benytte adaptiv clustering, som kan garantere tilstrekkelig oppløsning.

En løsning er implementert ved først å minimalisere antall initialpunkter (beskrevet i avsnitt 6.1.3), og i tillegg utføre hierarkisk clustering i flere lag (trestruktur-clustering) (6). Dette foregår ved at man utfører clustering et vilkårlig antall ganger innenfor foregående beregnede clusterer. Den mest generelle fremgangsmåten er å stoppe clusteringen når kun ett cluster gjenstår. Dette medfører en rekursiv algoritme. Det viser seg at man ikke kan programmere rekursivt i LABVIEW, så vi har brukt en "hardkodet" 2 nivåers clustering, som løser opp alle PRI nivåer i testene på sivile og militære fly. For frekvens-clustering holder det med ett nivå.



Figur 6.1 Grønt cluster har her blitt splittet opp i to clusterer, etter 2 gangs clustering. Søylen forestiller histogrammet til dataene.

6.1.3 Initialtilstand (frøpunkter) fra glattet histogram

For å velge få, men effektive, initialpunkter som leder til rask konvergens, brukes lokale maksima til histogrammet av dataene. Glattingen utføres enkelt med for eksempel et løpende middel over 3 til 5 punkter i histogrammet. Initialpunktene treffer omtrent de aktuelle tyngdepunkter, og man trenger da bare 2-3 iterasjoner per clustringsoperasjon. Parametere for clustringsalgoritmene er gitt i appendiks A.1.

6.2 PRI mønstergjenkjenning

Målet var her å utvikle en enkel algoritme som detekterer PRI-mønsteret. Ideen er å teste hyppigheten av PRI-skift. Vi ønsker å skille switched dwell fra stagger, med følgende rimelige antagelse som utgangspunkt:

Stagger skifter oftere PRI enn Switched Dwell.

6.2.1 Mønstergjenkjenning med andre ordens differanse histogram

For å teste ovennevnte utsagn er det naturlig å undersøke histogrammet til andre ordens TOA differanse, Δ_j^2 , som uttrykker den tidsderiverte til TOA:

$$\begin{aligned} PRI_j &= \Delta t_j = ToA_j - ToA_{j-1} \\ \Delta_j^2 &= \Delta t_j - \Delta t_{j-1} \end{aligned} \quad (6.1)$$

Karakteristiske trekk i histogrammet til Δ_j^2 indikerer hvilket PRI-mønster vi har:

PRI mønster	Karakteristiske trekk
Stagger	Bredt histogram, uniform fordeling
Switched dwell	Dominans av verdier rundt null
Konstant	Bare verdier rundt null

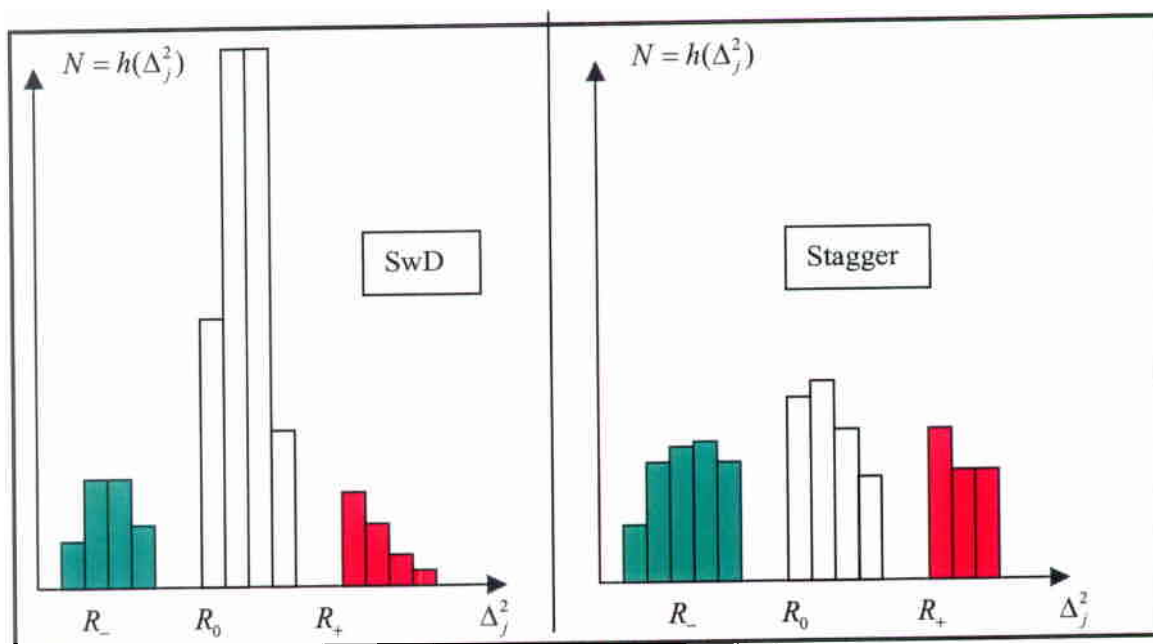
Disse trekkene kunne detekteres ved å måle "kurtosis" i fordelingene som er mål på hvor konsentrert histogrammet er rundt null (Kurtose bygger på 4. ordens momentet til en vilkårlig sannsynlighetsfordeling, og måler i hvilken grad dataene er super eller sub-gaussiske).

Vi har i stedet brukt clustring også i dette tilfellet, men nå på dataene Δ_j^2 . Vi beregner tre clusterer R_+ , R_- og R_0 , med antall tellinger N_+ , N_- og N_0 . Clusteret R_0 ligger rundt null, og R_+ og R_- hhv. på positiv og negativ side (Figur 6.4). Frøpunktene for R_+ , R_- og R_0 er henholdsvis satt til $\tilde{\mu} + \tilde{\sigma}$, $\tilde{\mu} - \tilde{\sigma}$ og $\tilde{\mu}$, der $\tilde{\mu}$ er middelværdi og $\tilde{\sigma}$ er standardavvik (for Δ_j^2). Tildesymbol indikerer at $\tilde{\mu}$ og $\tilde{\sigma}$ er estimater av middelværdi og standardavvik av den underliggende stokastiske prosessen.

Switched dwell detekteres (per definisjon) hvis vi oppfyller kriteriet

$$\alpha(N_+ + N_-) \leq N_0 \quad (6.2)$$

I motsatt fall detekteres stagger. Her er α en fri parameter. Det kreves at antall tellinger i senter-clusteret må være α ganger høyere enn tellingene i "vingene" av fordelingen, for at vi definerer signalet som switched dwell. Dette vil si at vi krever at switched dwell skifter nivå ikke oftere enn ca. hver α radarpuls. I motsatt fall detekteres stagger. Valget falt på $\alpha = 5$ under eksperimenter på mottatte radarsignaler fra fly.



Figur 6.2 Eksempler på Δ_j^2 -histogrammer. Switched dwell (venstre), stagger (høyre).

7 PULS-SORTERING PÅ DELVIS OVERLAPPENDE EMITTERE

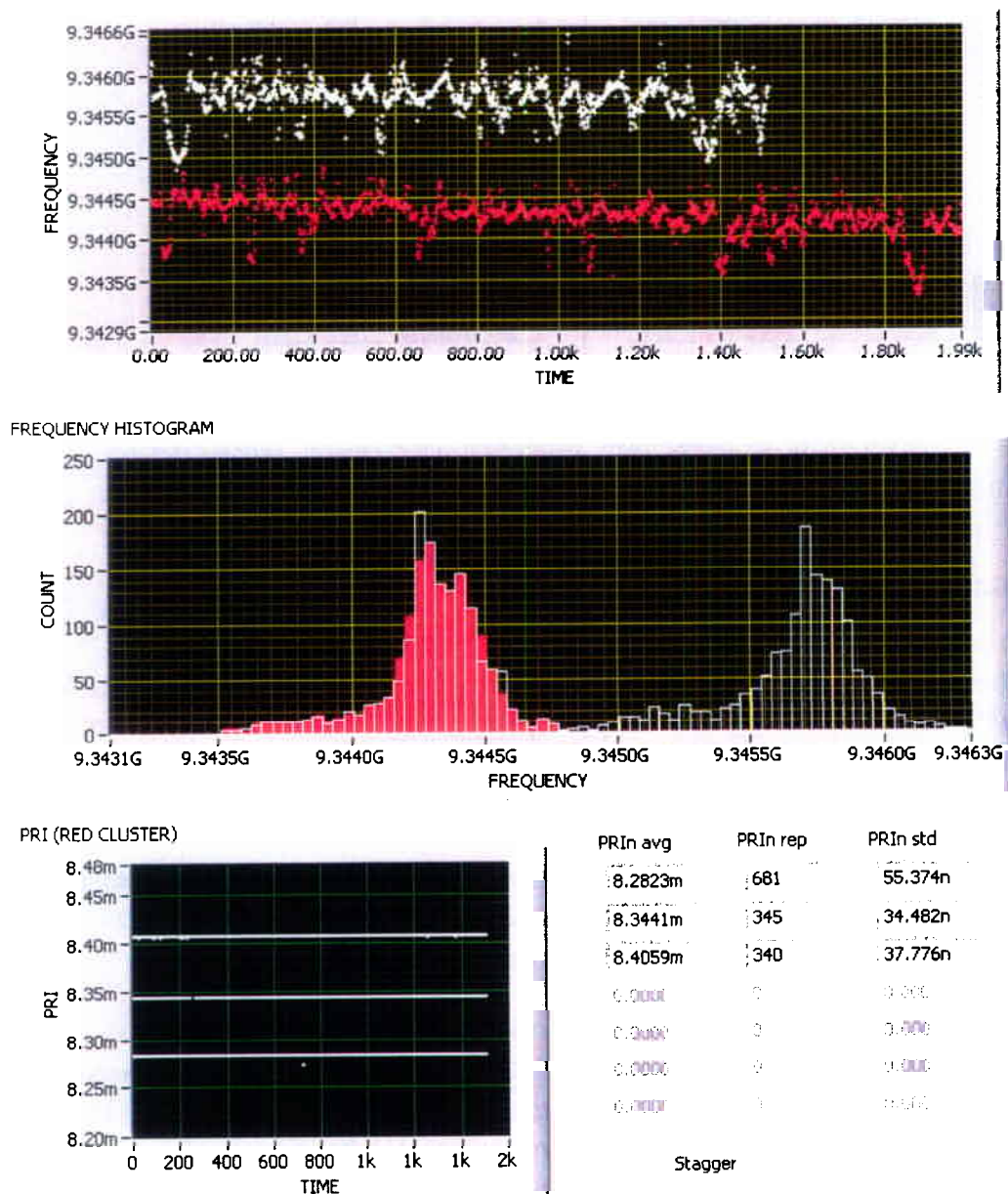
Vi vil her demonstrere at frekvensbasert puls-sortering, selv på delvis overlappende frekvensfordelinger, er et tilstrekkelig utgangspunkt for å oppnå et godt estimat av PRI. Vi tar for oss to PDW-sekvenser (A og B) tatt opp på Kjeller med ESMEX. De to sekvensene kommer fra samme emittertype med samme 3-nivå stagger mønster, og samme PRI-nivåer. Vi bruker en fritt valgt frekvensoffset på sekvens B slik at vi kan justere graden av overlapp til frekvensmålingene for sekvens A. Disse to sekvensene "stokkes" sammen med en liten tidsoffset på 10 mikrosekunder. I tillegg til en stokastisk komponent, viser frekvenssignalet til dels systematisk frekvensvariasjon over tid (oscillasjoner). De systematiske variasjonene skyldes sannsynligvis ytre effekter som multipath, og ikke selve emitteren.

Figur 7.1 viser frekvensene i sekvens A (rød) og sekvens B (hvit), i øverste figur. Her er graden av frekvensoverlapp liten, som også kan sees i histogrammet i midterste figur. Histogrammet som er merket rødt indikerer data for clusteret som tilsvarende sekvens A. Nederst til venstre vises PRI for "rødt cluster" og tabellene nederst til høyre viser middelerverdi [millisekunder] (PRI_n avg), antall tellinger (PRI_n rep), og standardavvik [nanosekunder] (PRI_n std). Disse PRI-dataene er funnet ved hjelp av clustrings-proseduren. Vi oppnår her korrekte PRI-nivåer.

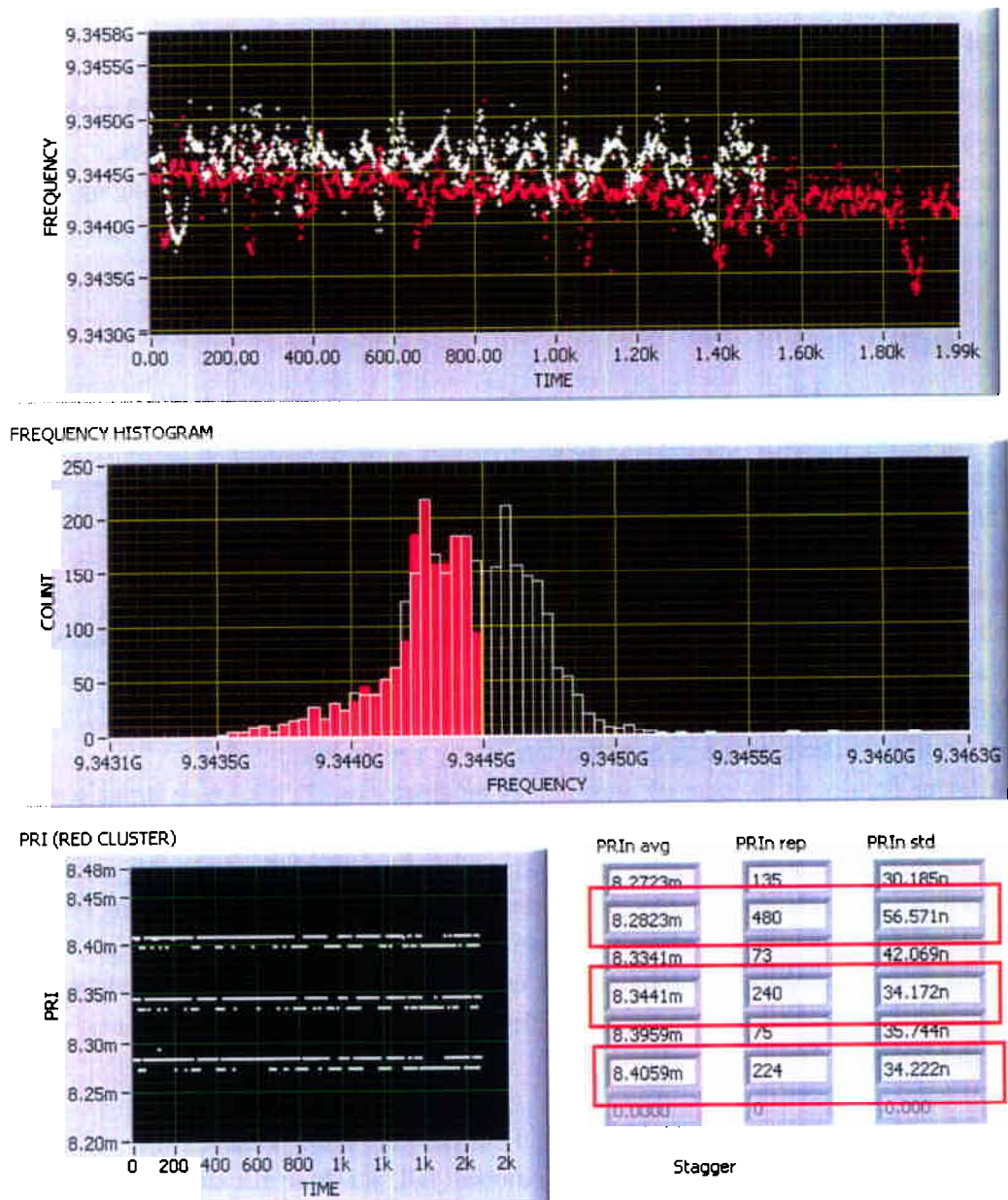
Figur 7.2 viser tilsvarende figur, men med delvis frekvensoverlapp. Middelerverdien til de to

frekvensfordelingene ligger nå bare ca. ett standardavvik fra hverandre. "Rødt cluster" inneholder nå flere punkter fra "hvitt cluster". Vi ser av PRI-plottet at vi sporadisk introduserer PRI med 10 mikrosekunders offset som skyldes data fra "hvitt" frekvenscluster. Disse falske PRI-nivåene inneholder imidlertid relativt få tellinger i forhold til de korrekte PRI-nivåene som er merket med røde rammer i Figur 7.2.

Forkastes PRI-clustere med få tellinger, vil vi eliminere de fleste falske PRI-nivåene. Alternativt vil et godt samsvar med PRI-data i en emitterdatabase (der de fleste PRI-nivåene stemmer overens med databasen) indikere korrekt emitter.



Figur 7.1 Sekvens A (rød) og sekvens B (hvit), er vist øverst. Sekvensene overlapper ikke i frekvens. Frekvenshistogrammet er vist i midterste figur. Rødt histogram indikerer sekvens A. Nederst til venstre vises PRI for sekvens A (som funksjon av tidsindeks) og tabellene nederst til høyre viser middelverdi [millisekunder] (PRIn avg), antall tellinger (PRIn rep), og standardavvik [nanosekunder] (PRIn std). Detektert PRI-mønster (fra mønstergjenkjenning) er også vist som "Stagger".



Figur 7.2 Se foregående figur for betydningen av plottene. Vi ser av PRI-plottet at vi sporadisk introduserer PRI med 10 mikrosekunders offset som skyldes data fra "hvitt" frekvenscluster. Disse falske PRI-nivåene inneholder imidlertid relativt få tellinger i forhold til de korrekte PRI-nivåene som er merket med røde rammer i tabellen.

8 IDENTIFIKASJON AV EMITTERINDIVIDER

Emitteridentifikasjon, dvs gjenkjenning av emittere på individnivå, kan gjøres ved å analysere emitterens pulsform. Dette kapitlet tar for seg de fleste aspekter ved pulsformanalyse; preprosessering, uttrekning av pulsens karakteristiske egenskaper, datakomprimering og pulsrepresentasjon og metoder for korrelasjon med emitterdatabase. Vi presenterer også noen resultater og erfaringer og vurderer veien videre.

8.1 UMOP

UMOP ("Unintentional Modulation On Pulse") refererer til utilsiktede modulasjoner i pulsform. Ideen bak å gjøre identifikasjon basert på UMOP er at disse modulasjonene er unike og konstante over tid for hver enkelt emitter. Pulsformen kan ses på som emitterens fingeravtrykk. Pulsformanalyse blir derfor ofte kalt "emitter fingerprinting".

UMOP er et resultat av en radars transmitter. Alle transmittere består av en modulator, høyspennings strømforsyning (HVPS) og RF-forsterkerenhet. De viktigste UMOP-effektene er

- Pushing-effekter
- Pulling-effekter
- Andre effekter (temperatur, gammelt utstyr, dårlig vedlikehold)

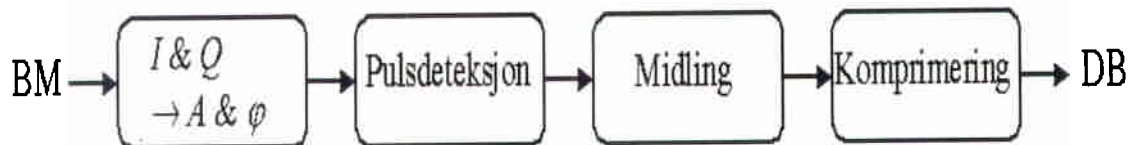
Pushing-effekter skyldes variasjoner på utgangen av modulator og HVPS. Disse variasjonene kommer til uttrykk som frekvens-, fase- eller amplitudemodulasjon. Som eksempel vil ripple i HVPS i en magnetrontransmitter komme til uttrykk som frekvensmodulasjon i pulsen. Pulling er normalt et problem for pulsede oscillatortransmittere (magnetroner). En dårlig kobling i en roterende bølgeleder vil kunne gi frekvens- og fasevariasjoner gjennom pulsen. Dårlig vedlikehold og temperaturdrift i RF-forsterkeren er to av flere elementer som vil påvirke pulsmodulasjonen. En radar som ikke vedlikeholdes vil kunne få drift i frekvens over tid som følge av drift i HVPS.

Dersom de ovennevnte effektene er stabile over tid bør det være mulig å utnytte disse til å gjenkjenne en emitter ved dens unike pulsform. I motsatt fall, når disse effektene fører til at emitteren er ustabil og sender ut pulser med tidsvarierende pulsform blir denne oppgaven svært vanskelig. Andre effekter som kan påvirke pulsform, så som temperatur, vil selvsagt også vanskeliggjøre en nøyaktig karakterisering. Dersom emitteren ellers er stabil vil endring av pulsform pga slitasje og dårlig vedlikehold skje på en relativt langsom tidsskala. En karakterisering av pulsformen til slike emittere er derfor gyldig over korte tidsrom. Går det for lang tid mellom hver gang signalene til en emitter blir tatt opp vil derfor gjenkjenning av signalene være vanskelig. Hvis signalene *blir* gjenkjent, er det viktig at det tilhørende database-elementet oppdateres.

Det bør nevnes at det er svært stor forskjell på ulike typer emittere. Værradarer på sivile passasjerfly f eks, er sannsynligvis mer utsatt for slitasje og dermed variasjon i pulsform enn de fleste militære radarer siden kravene de sistnevnte skal tilfredsstillende er langt høyere enn kravene som blir stilt til sivile radarer. Dette kan tyde på at karakterisering og gjenkjenning av militære emittere i større grad er gjennomførbart enn for sivile emittere.

8.2 Pulsdeteksjon og preprosessering

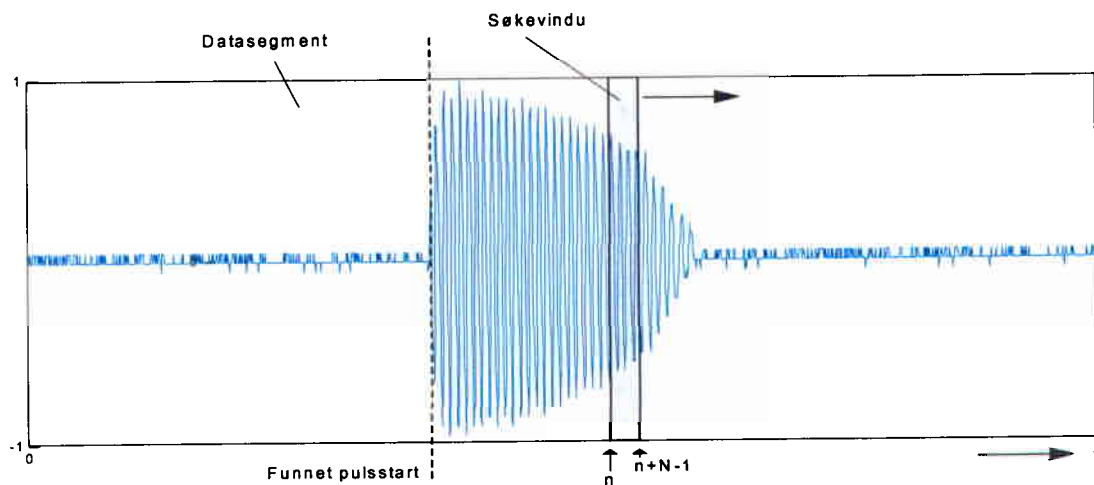
En av oppgavene til ESMEX er beregning av parametere som beskriver pulsform. Pulsene detekteres og pulsdataene legges i datasegmenter som typisk inneholder 1024 målinger av I og Q . Fra disse beregnes amplitude og fase for pulsen, amplituden brukes så til å lokalisere pulsen i segmentet, fra de lokaliserte pulsene beregnes den gjennomsnittlige pulsen (amplitude og fase). Til sist bestemmes komprimerte representasjoner av pulsens amplitude og fase. Disse blir korrelert mot tilsvarende representasjoner i en emitterindivid-database. Figur 8.1 viser et flytskjema over denne prosessen.



Figur 8.1 Flytskjema for pulsformanalyse i ESMEX. I første blokk beregnes amplitude og fase fra I og Q i rådata (BM – Burst Memory). Deretter følger pulsdeteksjon ved terskling - og pulsmidling. Ved komprimering erstattes pulsrepresentasjonen med et mindre sett koeffisienter som koder for pulsform. Til slutt sammenlignes disse koeffisientene med verdier i en database for om mulig å gjenkjenne emitterindivid.

8.2.1 Deteksjon av puls

Pulsdeteksjonen som utføres av kanalkortet gir en svært unøyaktig lokalisering av pulsen. For å bestemme hvor pulsen starter og stopper brukes en mer nøyaktig metode. Deteksjonen blir gjort med energiterskling som prinsipp. Gjennomsnittsennergien i et søkevindu blir målt og søkevinduet blir så flyttet i diskrete steg gjennom datasegmentet (Figur 8.2).



Figur 8.2 Deteksjon av puls i et datasegment.

Deteksjonen styres av tre parametere: En terskel Θ_{start} for å finne pulsstart, en annen terskel Θ_{stop} for å finne slutten på pulsen og størrelsen på søkevinduet (antall sampler). Gjennomsnittsennergien $E(n)$ er definert som

$$E(n) = \frac{1}{N} \sum_{i=n}^{n+N-1} |x_i|^2,$$

hvor x_i betegner verdien til sample i og n er samplenummer. Når denne energien overstiger Θ_{start} blir samplet i midten av søkevinduet merket som starten av pulsen. Når energien faller under Θ_{stop} blir samplet i midten av søkevinduet merket som slutten av pulsen.

8.2.2 Beregning av amplitude og fase

Vi ønsker å trekke ut tidsavhengig amplitude $A(t)$ og fase $\varphi(t)$ fra hver enkelt puls. Signalet (puls) er representert ved funksjonen $f = f(t)$ hvor f er på formen

$$f(t) = A(t)\exp(i\varphi(t)).$$

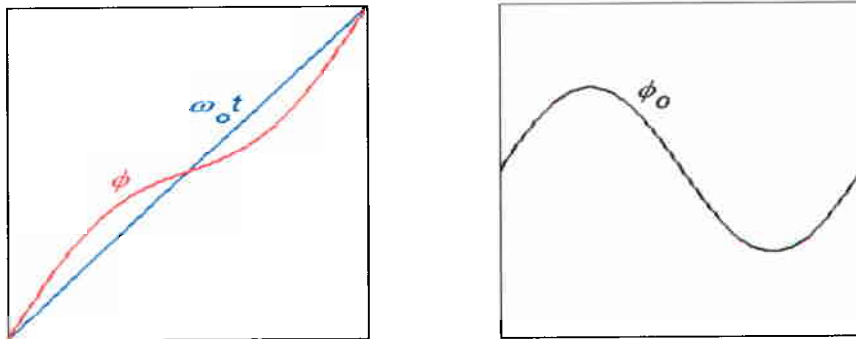
De reelle og imaginære komponentene i det komplekse signalet er output fra systemet (kanal-kortet) og vil bli referert til som henholdsvis $I = I(t)$ og $Q = Q(t)$. Dermed kan amplituden $A(t)$ og fasen $\varphi(t)$ bestemmes ved hjelp av formlene

$$A(t) = \sqrt{I(t)^2 + Q(t)^2}, \quad \varphi(t) = \arctan\left(\frac{Q(t)}{I(t)}\right).$$

Arctan-funksjonen gir verdier i intervallet $(-\pi/2, \pi/2)$. Siden fasen er tilnærmet lineær betyr dette at $\varphi(t)$ er en tilnærmet sagtannfunksjon. For å få fasen på formen

$$\varphi(t) = \omega_0 t + \varphi_0(t), \quad (8.1)$$

hvor ω_0 er senterfrekvensen og $\varphi_0(t)$ er den ikkelineære komponenten av fasen, "limes" de forskjellige sagtennene sammen ("unwrapping"). Siden vi ikke kan bruke frekvensen ω_0 til individgjenkjenning er vi interessert i å bestemme komponenten $\varphi_0(t)$. Dette gjøres ved å bruke lineær regresjon på den sammenlimte fasen til å bestemme ω_0 og deretter finne $\varphi_0(t)$ ved å trekke den lineære komponenten $\omega_0 t$ fra $\varphi(t)$ i uttrykket i ligning (8.1) (se Figur 8.3). Vi bruker $A(t)$ og $\varphi_0(t)$ som representasjoner av pulsen.



Figur 8.3 Beregning av fasens ikkelineære komponent. Den lineære komponenten av fasen $\omega_0 t$ beregnes ved lineær regresjon og trekkes fra fasen $\varphi(t)$ (venstre). Resultatet er den ikkelineære fasen $\varphi_0(t)$ (høyre).

Tidligere i prosjektet ble instantanfrekvens (IFV) brukt som representasjon av pulsen (10), (17). IFV er definert som den tidsderiverte av fasen $\varphi(t)$. Fra ligning (8.1) ser vi at $\text{IFV} = \omega_0 + \varphi_0'(t)$. Bærefrekvensen ω_0 trekkes fra denne størrelsen slik at vi får $\text{IFV}_{\text{korrigeret}} = \varphi_0'(t)$. Beregning av IFV innebærer numerisk derivasjon av fasen. Dette medfører en uønsket forsterkning av støyenergien i signalet.

8.2.3 Midling over pulser

Den gjennomsnittlige pulsen i et pulstog bestemmes ved å midle over alle pulsene i pulstoget. Dette er ikke-trivielt siden posisjonen til de ulike pulsene i datasegmentet ikke er lik. Det er derfor nødvendig å forskyve pulsene slik at tilsvarende datapunkter har riktig posisjon i forhold til hverandre. Når dette er gjort bestemmes den gjennomsnittlige pulsen ved midling.

For å oppnå den ønskede midlingen benyttes følgende metode. La oss anta at $e_k(t_j)$ representerer amplituden til puls k i tidspunktet t_j , hvor $j = 1, 2, \dots, N$ og N er antall sampler i den relevante delen av datasegmentet (den delen av segmentet som ikke inneholder signal fjernes i alle pulser i pulstoget ved å bruke informasjon om gjennomsnittlig pulsbredde). Tyngdepunktet til pulsen defineres da som

$$c_k = \frac{\sum_{j=1}^N t_j e_k(t_j)}{\sum_{j=1}^N e_k(t_j)}.$$

Vi antar at tyngdepunktet representerer et punkt som relativt til pulsen ikke fluktuerer fra puls til puls. En forskyvning av pulsene i forhold til hverandre vil derfor kunne bety at sample j i ulike pulser representerer det samme punktet i pulsene uansett hvilken puls vi ser på. Når tyngdepunktet c_k for puls k er beregnet for alle verdier av k bestemmes det gjennomsnittlige tyngdepunktet $\bar{c} = (\sum_{k=1}^M c_k) / M$ (vi antar at det er M pulser i pulstoget). Til sist forskyves alle datasegmenter $]c_k - \bar{c}[$ hakk til venstre ($] [$ betegner avrunding til nærmeste hele tall). Til sist tar vi gjennomsnittet av amplitude og fase.

Det er i hovedsak to ulike måter å komme fram til en representasjon av pulsene i et pulstog på. Den første er den som er nevnt over, altså først midling, så beregning av komprimert representasjon. Den andre metoden beregner først den komprimerte representasjonen for hver puls i pulstoget og regner deretter ut de gjennomsnittlige parametrene for pulsene samt standardavvik. Fordelen med denne metoden er at standardavvik for pulsparametere kan regnes ut, i metoden beskrevet over er dette ikke mulig.

8.3 Pulskomprimering ved ortonormal approksimasjon

Representasjon av pulsform ved amplitude og fase betyr at pulsen representeres med et stort antall datapunkter ($2 \times$ antall sampler i pulsen). Videre vil en slik direkte representasjon ikke bidra til reduksjon av støy. Det er derfor ønskelig å bruke en metode som fører til reduksjon både i antall parametere og i støy. Til dette formålet har vi brukt representasjon ved ortogonale polynomer. Vi kommer tilbake til denne klassen funksjoner senere. Først presenteres generell teori for representasjon av signaler ved såkalte ortonormale approksimasjoner.

Gitt et rent (støyfritt), kontinuerlig signal $f(t)$ definert på $[-1, 1]$ (dersom signalet er definert på et annet intervall kan vi bruke en lineær transformasjon for å skifte intervall til $[-1, 1]$). Vi ønsker å representere f ved en lineær kombinasjon av basisfunksjoner, dvs

$$f(t) = \sum_{k=1}^{\infty} a_k E_k(t), \quad -1 \leq t \leq 1,$$

hvor rekka kan trunkeres etter N ledd for å gi approksimasjonen $f_N(t)$. Dersom basisen $\{E_k\}_{k=1}^{\infty}$ er ortogonal er koeffisientene gitt ved

$$a_k = \int_{-1}^1 w(t) f(t) \overline{E_k(t)} dt. \quad (8.2)$$

I ligning (8.2) er $w(t)$ en vektfunksjon og overstreking betegner kompleks konjugert. Vi antar at basisen er ortonormal, dvs. $\langle E_k, E_\ell \rangle = \delta_{k\ell}$, hvor $\delta_{k\ell}$ er 1 når $k = \ell$ og 0 ellers og indreproduktet $\langle \cdot, \cdot \rangle$ er definert som

$$\langle f, g \rangle = \int_{-1}^1 w(t) f(t) \overline{g(t)} dt.$$

Signalet er digitalisert i punktene $t_j, j = 0, 1, \dots, M+1$, hvor $t_0 = -1$ og $t_{M+1} = 1$. I disse punktene har det rene signalet verdien $f(t_j) = f_j$. Vi antar at signalet inneholder hvit støy, dvs avviket fra det rene signalet i hvert samplingspunkt er gaussisk med middelvei 0 og standardavvik σ_j .

Signalet som blir tatt opp i $t = t_j$, er dermed

$$\tilde{f}_j = f_j + s, \quad s \sim N(0, \sigma_j^2).$$

Vi definerer det kontinuerte signalet som funksjonen vi får når vi lineært interpolerer mellom samplingspunktene. Koeffisientene \tilde{a}_k bestemmes tilnærmet ved bruk av trapesmetoden til å beregne integralet i ligning (8.2), dvs

$$\tilde{a}_k \approx \Delta \left[\frac{1}{2} w(t_0) \tilde{f}_0 E_k(t_0) + \sum_{j=1}^M w(t_j) \tilde{f}_j E_k(t_j) + \frac{1}{2} w(t_{M+1}) \tilde{f}_{M+1} E_k(t_{M+1}) \right], \quad (8.3)$$

hvor Δ er (den konstante) avstanden mellom samplingspunktene. For det rene signalet $f(t)$ gjelder ligning (8.3) uten tildene.

Hvilke krav stilles til basisen $\{E_k\}_{k=1}^\infty$? Som sagt må basisen være ortonormal. I tillegg må den være en basis for et rimelig valgt funksjonsrom, vanligvis for $L_2[-1,1]$, rommet av funksjoner som er kvadrert integrerbare over $[-1,1]$ (dvs kvadratet av funksjonene har endelig integral over $[-1,1]$). Funksjonene i basisen må kunne utspenne rommet $L_2[-1,1]$. I praksis vil dette si at enhver funksjon $g \in L_2[-1,1]$ kan skrives som en lineær kombinasjon av funksjoner i basisen $\{E_k\}_{k=1}^\infty$, dvs

$$g(t) = \sum_{k=1}^{\infty} b_k E_k(t).$$

Hva slags funksjoner er E_k ? Et enkelt eksempel er den trigonometriske funksjonsbasisen

$$\left\{ \frac{1}{\sqrt{2}} \exp(i2\pi kt) \right\}_{k=0}^\infty \text{ definert på } [-1,1] \text{ som er ortonormal gitt vektingsfunksjon } w(t) \equiv 1. \text{ Det}$$

fins en rekke ortonormale basiser som består av, eller er utledet fra, polynomer. Eksempler på disse er de to klassene Chebyshev-polynomer (med vektingsfunksjoner $w(t) = (1-t^2)^{1/2}$ og $w(t) = (1-t^2)^{-1/2}$, respektive) og Legendre-polynomer (vektingsfunksjon $w(t) \equiv 1$) (4), (16).

Legendre-polynomer er brukt til pulsrepresentasjon i ESMEX. En annen viktig klasse av ortonormale approksimasjoner er wavelets (3), (5), (17).

8.4 Statistiske metoder for ortonormale representasjoner

Signalene vi studerer vil alltid være forurenset av støy. Vi betrakter derfor alle observasjoner som stokastiske variable.

8.4.1 Statistikk for pulsrepresentasjoner

For å kunne identifisere emittere er det nødvendig å sammenligne ulike pulsrepresentasjoner med hverandre. Dette gjøres ved å måle avstanden mellom to sett koeffisienter i en eller annen avstandsnorm, hvor det ene settet korresponderer til emitteren i emitterdatabasen og det andre til de observerte pulsene. Er avstanden mindre enn en gitt terskel er emitteren identifisert.

Gitt de to signalene $\tilde{f}(t)$ og $f(t)$ som er representert ved henholdsvis $\tilde{f}_N(t) = \sum_{k=1}^N \tilde{a}_k E_k(t)$ og

$f_N(t) = \sum_{k=1}^N a_k E_k(t)$. Pga ligning (8.3) blir differensen mellom koeffisientene i disse to

representasjonene

$$\tilde{a}_k - a_k \approx \Delta \left[\frac{1}{2} (\tilde{f}_0 - f_0) E_k(t_0) + \sum_{j=1}^M (\tilde{f}_j - f_j) E_k(t_j) + \frac{1}{2} (\tilde{f}_{M+1} - f_{M+1}) E_k(t_{M+1}) \right], \quad (8.4)$$

$$k = 1, 2, \dots, N.$$

Dette er en lineær kombinasjon av de stokastiske variablene $\tilde{f}_j - f_j$ som har middel 0 og varians σ_j^2 . Hvert av leddene i ligning (8.4) er normalfordelt med middel 0 og varians $(\sigma_j \sqrt{\Delta})^2 E_k(t_j)^2 \Delta$. Summen blir da normalfordelt med middel 0 og varians

$$\left[\frac{1}{4} (\sqrt{\Delta} \sigma_0)^2 E_k(t_0)^2 \Delta + \sum_{j=1}^M (\sqrt{\Delta} \sigma_j)^2 E_k(t_j)^2 \Delta + \frac{1}{4} (\sqrt{\Delta} \sigma_{M+1})^2 E_k(t_{M+1})^2 \Delta \right]. \quad (8.5)$$

Dersom støyen har konstant standardavvik lik σ kan størrelsen $(\sqrt{\Delta} \sigma)^2$ settes foran uttrykket i (8.5) og uttrykket innenfor parentesene blir tilnærmet lik $\int_{-1}^1 E_k(t)^2 dt$ som er 1 pga

ortonormaliteten. Hvis dette er tilfellet kan vi konkludere at for differensen $\tilde{a}_k - a_k$ gjelder tilnærmet

$$\tilde{a}_k - a_k \sim N(0, (\sigma \sqrt{\Delta})^2), \quad k = 1, 2, \dots, N, \quad (8.6)$$

mao er størrelsene $(\tilde{a}_k - a_k) / (\sigma \sqrt{\Delta})$ standardnormalfordelt.

Det er vanlig å måle avstand mellom to funksjoner ved å bruke L_2 -normen. Avstanden i L_2 -norm mellom \tilde{f}_N og f_N er

$$\|f_N - \tilde{f}_N\|_2^2 = \int_{-1}^1 (f_N(t) - \tilde{f}_N(t))^2 dt = \int_{-1}^1 \left[\sum_{k=1}^N (a_k - \tilde{a}_k) E_k(t) \right]^2 dt = \sum_{k=1}^N (a_k - \tilde{a}_k)^2, \quad (8.7)$$

hvor den siste likheten følger av ortonormaliteten til basisen $\{E_k\}_{k=1}^{\infty}$. Av ligning (8.7) ser vi at avstanden i L_2 -norm tilsvarer den euklidske avstanden mellom vektorene som inneholder koeffisientene som de to signalene er representert ved.

8.4.2 Fordelingsfunksjoner for avstanden når $\tilde{f} = f + \text{støy}$

La oss for enkelhets skyld anta at referansesignalet f ikke er stokastisk, mens samplene i det innkommende signalet \tilde{f} er normalfordelt rundt samplene i referansesignalet med varians σ^2 . Hva slags statistikker kan vi da utlede som det innkommende signalet skal tilfredsstillere?

Fra ligning (8.6) følger det at $[1/\sigma \sqrt{\Delta}](\tilde{a}_k - a_k)$ er standardnormalfordelt. Vi antar at de ulike størrelsene for ulike verdier av k er statistisk uavhengige. Sannsynlighetsteorien gir da at

summen av kvadratet til disse størrelsene er kji-kvadrat-fordelt (se Appendiks A.4) med N frihetsgrader, dvs

$$\frac{1}{\sigma^2 \Delta} \sum_{k=1}^N (a_k - \tilde{a}_k)^2 \sim \chi_N^2.$$

Forventningen og variansen til denne størrelsen er henholdsvis N og $2N$. Den forventede verdien for avstanden mellom f og \tilde{f} (eller riktigere, mellom f_N og \tilde{f}_N) er dermed $\sigma^2 \Delta N$, mens variansen er $\sigma^2 \Delta 2N$. Dersom støynivået er høyt og samplingsfrekvensen lav (σ og Δ er store) bør antall koeffisienter (N) ikke være stort. I motsatt fall bør antall koeffisienter være større. Siden representasjon av signal innebærer en approksimasjon inneholder avstanden mellom f og \tilde{f} i tillegg ledd som representerer approksimasjonen av den støyfrie funksjonen f ved en lineær kombinasjon av basisfunksjoner. I Appendiks A.3 er uttrykk for avstanden mellom f og \tilde{f}_N gitt når basisfunksjonene er Legendre-polynomer.

8.4.3 Fordelingsfunksjoner for avstanden når $\tilde{f} = h + \text{støy}$, $h \neq f$

Hva slags statistikker kan vi forvente dersom det innkommende signalet ikke kommer fra samme emitter som referansesignalet? I dette tilfellet vil koeffisientene \tilde{a}_k ikke være normalfordelt omkring a_k . Anta isteden at de er normalfordelt rundt koeffisientene b_k med varians $\sigma^2 \Delta$. Dermed blir størrelsen $(a_k - \tilde{a}_k)$ normalfordelt med middel $(a_k - b_k)$ og varians $\sigma^2 \Delta$. Av dette følger det at

$$(1/\sigma\sqrt{\Delta})(\tilde{a}_k - a_k) \sim N(a_k - b_k, 1).$$

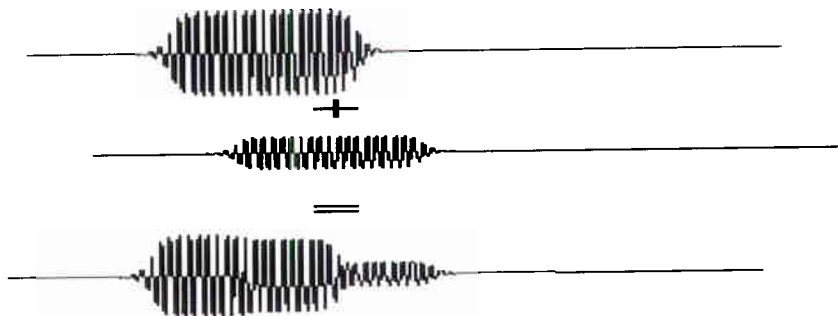
Da får vi

$$\frac{1}{\sigma^2 \Delta} \sum_{k=1}^N (a_k - \tilde{a}_k)^2 \sim \chi_{N,\lambda}^2,$$

hvor $\chi_{N,\lambda}^2$ betegner den såkalt ikke-sentrale kji-kvadrat-fordelingen med frihetsgrad N og parameter λ (se Appendiks A.4). Forventning og varians for størrelser som er ikke-sentralt kji-kvadrat-fordelte er henholdsvis $N + \lambda$ og $2(N + 2\lambda)$, dvs fordelingen flyttes mot høyre og smøres ut når λ vokser (se Figur A.2 i Appendiks A.4 for et eksempel). I vårt tilfelle er parameteren λ et mål på avstanden mellom f og \tilde{f} , nærmere bestemt er

$$\lambda = \frac{1}{\sigma\sqrt{\Delta}} \sum_{k=1}^N (a_k - b_k)^2 \approx \frac{1}{\sigma\sqrt{\Delta}} \|f - \tilde{f}\|_2^2.$$

Når vi får inn et stort antall pulser på mottageren vår kan teorien over anvendes i emittergjenkjenning, se Appendiks A.5.



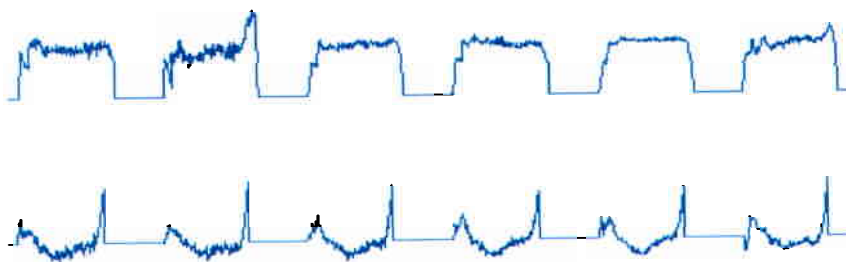
Figur 8.4 Skisse av pulsforsvigning ved multipath. Den direkte pulsen (øverst) og den reflekterte pulsen (midten) adderes i mottageren og resulterer i en forvrent puls (nederst).

8.5 Multipath

Når antennen mottar refleksjoner av signalet fra bakke og bygninger og disse adderes til direktesignalet, kan dette føre til forvrengte pulsformer, se Figur 8.4. Siden emitterens posisjon endres i løpet av den tiden det tar å emitte et pulstog, vil de reflekterte pulsene i ulik grad være forsinket i forhold til de direkte pulsene. Dette fører til at pulsformen vil variere fra puls til puls i et pulstog, særlig når emitteren er nær mottageren. Vi kan derfor finne ut om signalet er forurenset av multipath ved å studere variasjonen i pulsform mellom de ulike pulsene i pulstoget.

8.6 Resultater

De første to årene av PROFIL-prosjektet ble det gjort studier av pulser prosessert på oscilloskop. Siden DIPP-kortet ble tatt i bruk høsten 2002 har vi hatt bedre muligheter til å studere pulsformer fra mange emittører, fortrinnsvis fra værradarer på sivile passasjerfly og det er blitt mulig å gjøre klassifisering og identifikasjon i sanntid. Algoritmene for komprimering av pulsform er implementert i ESMEX. Foreløpig (mars 2003) er de implementerte metodene for korrelasjon med database relativt primitive. I tillegg er individ-databasen svært begrenset. Vi har derfor få eksempler på emittergjenkjenning i sanntid, men vi har gjort en rekke interessante observasjoner.



Figur 8.5 Amplitude (øverst) og fase (nederst) i et pulstog som inneholder varierende grad av multipath.

Multipath forekommer ofte: Ut fra pulsformene på pulsene vi mottar ser det ut til at multipath forekommer svært ofte (se Figur 8.5). Dette gjør emittergjenkjenning svært vanskelig.

Fasen er mindre påvirket av multipath enn amplituden: Vi har observert at fasen i betydelig mindre grad enn amplituden blir forvrengt av multipath (se for eksempel Figur 8.5). En enkel analyse (Appendiks A.6) viser at dette er rimelig. Hvis vi antar at forsinkelsen τ er liten og at amplituden og fasen endres lite over små tidsintervaller får vi disse tilnærmingene:

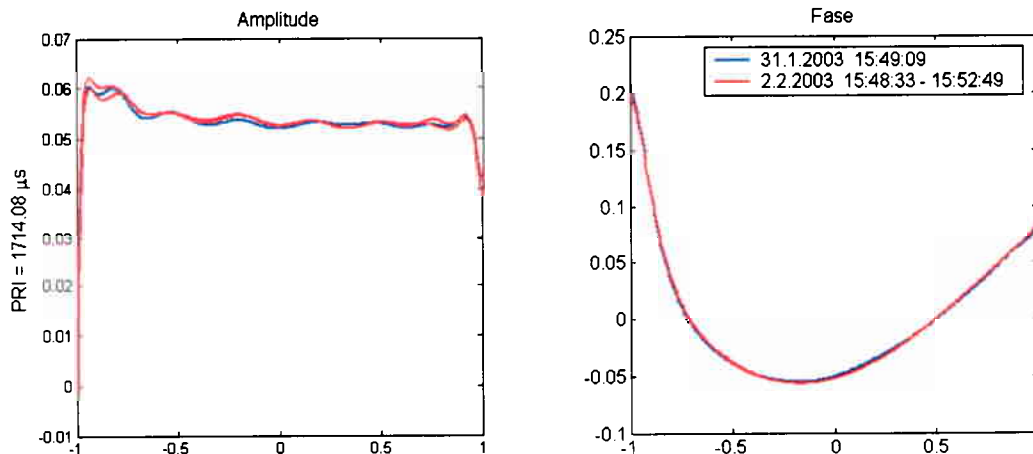
$$\begin{aligned} \text{Amplitude} &\approx A(t) \{1 + 2\alpha \cos(\omega_0 \tau) + \alpha^2\}^{1/2}, \\ \text{Fase} &\approx \omega_0 t + \varphi_0(t) + \arctan \left\{ \frac{\alpha \sin(\omega_0 \tau)}{1 + \alpha \cos(\omega_0 \tau)} \right\}. \end{aligned} \quad (8.8)$$

Arctan-leddet i uttrykket for fasen er konstant siden forsinkelsen antas å være tidsuavhengig. Den ikkelineære komponenten til fasen er derfor lik $\varphi_0(t)$. Mao ser det ut til at det ikke er noen forskjell i den ikkelineære fasen når signalet er forurenset av multipath i forhold til når signalet er rent. Vi minner om at uttrykkene i ligning (8.8) kun gjelder for små verdier av τ .

Robustheten til fasen gjør denne til en nyttig egenskap til bruk i emittergjenkjenning.

Ofte små variasjoner i pulsform mellom emittere av samme type: Mange emittere av samme type har pulsformer som er svært like. Det er derfor usikkert om systemet vårt har tilstrekkelig dynamikk til å skille slike emittere.

Emittergjenkjenning i ESMEX: Profil har gjennom sitt eksperimentelle ESM-system ESMEX tatt opp signaler som i hovedsak er emittert av værradarer på sivile fly på vei til eller fra Oslo lufthavn Gardermoen. Programvare for identifikasjon av emitterindivid er implementert i systemet, men så langt (mars 2003) er denne komponenten fremdeles på test- og forsøksstadiet. Hovedgrunnene til dette er at databasen foreløpig er svært begrenset og at vi er usikre på om de elementene som nå utgjør individdatabasen representerer emitterne de er ment å representere på en god måte.



Figur 8.6 *Eksempel på gjenkjent puls. Legg merke til at opptaket er gjort på omtrent samme tidspunkt på to ulike dager. Dette sannsynliggjør at signalet kommer fra samme emitter.*

Figur 8.6 viser høyst sannsynlig et eksempel på emittergjenkjenning. I loggen tilhørende dette opptaket observerer vi:

```

10  15:49:09.46  9346  332.8  157  1(0)  30  --E 1713.9  0.06  0.000 D 0.053 E
:      :      :      :      :      :      :      :      :      :
1166 15:48:32.74 9347  316.1  163  1(0)  30  --E 1713.9  0.062  0.030 D 0.060 E
1167 15:52:48.82 9347  337.4  116  1(0)  24  --E 1713.9  0.058  0.026 D 0.054 E

```

Hver linje i loggen representerer pulsene i et pulstog og viser tidspunkt (ikke dato) og emittertype (1713.9 i loggen ovenfor) i tillegg til en hel del andre data. I den øverste linja (pulstog med nummer 10) har vi treff i individdatabasen på element "D" med score 0 (dette betyr 100 % treff, noe som kommer av at elementet "D" ble definert fra disse dataene). Pulsene i de to pulstogene beskrevet i de to nederste linjene i loggen (nummer 1166 og 1167) har også treff i individdatabasen på element "D". Disse to pulstogene ble tatt opp omtrent samtidig på dagen, men på en annen dag enn pulstog med nummer 10 (dette går ikke frem av loggen, men kan ses i Figur 8.6). Kolonnen lengst til høyre i loggen antyder hvilket databaseelement som gir nest best treff, her elementet "E".

8.7 Konklusjoner og utfordringer

Erfaringene vi har hatt så langt innbyr ikke til å trekke sikre konklusjoner. Tvert i mot har mange nye utfordringer og spørsmål åpenbart seg gjennom arbeidet. Noen av de viktigste utfordringene er de følgende.

Multipath: Det er ikke realistisk å fjerne effektene av multipath kun ved bruk av programvare, men det bør være mulig å detektere om en puls er forvrengt pga multipath-effekter og dermed forkaste denne pulsen. De resterende pulsene er da forhåpentligvis "gode" nok til å brukes i gjenkjenning. Som beskrevet i avsnitt 8.6 er fasen mindre påvirket av multipath enn amplituden, særlig når tidsforsinkelsene i forhold til reflektert komponent er små. Dette betyr at det kan være mulig å bruke fasen til gjenkjenning også når signalet er delvis kontaminert av multipath.

Ut fra formen på amplituden i et typisk pulstog ser det ut som om det er små variasjoner fra puls til puls som ikke kan forklares av vanlig gaussisk støy. Kan dette være forårsaket av atmosfæriske forhold som for eksempel spredning? Bør vi i så fall ta hensyn til dette fenomenet når vi velger egenskaper til bruk i gjenkjenning og metoder for representasjon og korrelasjon? Både dette fenomenet og multipath vil bli gjenstand for teoretiske studier og pulsformanalyse.

Pulsrepresentasjon og korrelasjon: Legendre-polynomer er effektive for komprimering og praktiske for korrelasjon. Wavelets ble brukt og implementert i Matlab-programvare tidligere i prosjektet (17). For glatte og fine pulser kan komprimeringen være mer effektiv ved approksimasjon med polynomer enn med wavelets. For mindre glatte pulsformer kan det motsatte være tilfelle. For å klargjøre hvilke metoder som egner seg best, bør studier som sammenligner ulike metoder for komprimering og korrelasjon utføres. Deretter bør de beste metodene implementeres i ESMEX. I tillegg til wavelets og ortogonale polynomer bør andre metoder, som diskrete polynomer, trigonometriske funksjoner, interpoleringer og splines være med i en vurdering av representasjonsmetoder.

Et annet spørsmål knyttet til pulsrepresentasjon er hvilke egenskaper ved pulsene som skal representeres. Istedenfor, eller i tillegg til, amplitude og fase kan en tenke seg å benytte bestemte særtrekk ved en emitters pulsform som kan tenkes å skille pulsene til denne emitteren fra andre pulser.

Foreløpig er korrelasjonsmålet i identifikasjonsmodulen i ESMEX svært primitivt; vi bruker et relativt avstandsmål. Når en innkommende pulsform ligger i mindre enn 5% avstand fra pulsformen i databasen sier vi at vi har treff. Siden stabiliteten til ulike emittere varierer, betyr dette at vi får svært mange feilidentifiseringer (fordi kravet ikke er strengt nok) og mange

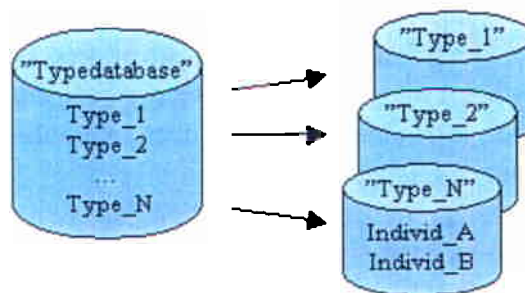
tilfeller hvor vi ikke identifiserer (fordi kravet er *for* strengt). Bruk av den såkalte Mahalanobis-avstanden vil sannsynligvis bøte på dette siden dette målet tar hensyn til hvor stor variasjonen i data er. Siden de forskjellige pulsformene ofte ligger svært nær hverandre kan det være helt avgjørende å velge et avstandsmål som er i stand til å diskriminere mellom overlappende emittere.

Deteksjon av små forskjeller i pulsform: Analyser bør gjøres for å vurdere hvor ulike to pulsformer må være for at vi skal være i stand til å skille dem gitt signal-til-støy-nivå, antall pulser og dynamikk i mottageren.

Database: Elementene i databasen er representasjoner av signalene fra ulike emittere. For å unngå feilidentifiseringer er det kritisk at vi har nøyaktige representasjoner i databasen. I den forbindelse bør vi studere hvordan vi kan forsikre oss om at signalene som brukes til å bygge databaseelementer er gode representasjoner av signalet fra den aktuelle emitteren.

9 EMITTERGJENKJENNING VED DATABASEKORRELASJON

Vi deler gjenkjenning av emittere opp i typegjenkjenning (klassifikasjon) og individgjenkjenning (identifikasjon). For typegjenkjenning brukes foreløpig kun PRI. Aktuelle tilleggsparametere er frekvens og pulsbredde. Verdiene for type-parametere beregnes rett etter puls-sorteringen. Gjenkjent emittertype brukes for å velge riktig individdatabase. EDWens pulsform-parametere brukes for korrelasjon med tilsvarende parametere i den valgte individdatabasen. Pulsform-parametere (koeffisienter for omhyllingskurve og fase) beregnes og legges til EDWen under pulskomprimering. Resultatet av gjenkjenningsprosessen legges til EDWen i form av referanser til gjenkjent emittertype og emitterindivid i databasene.



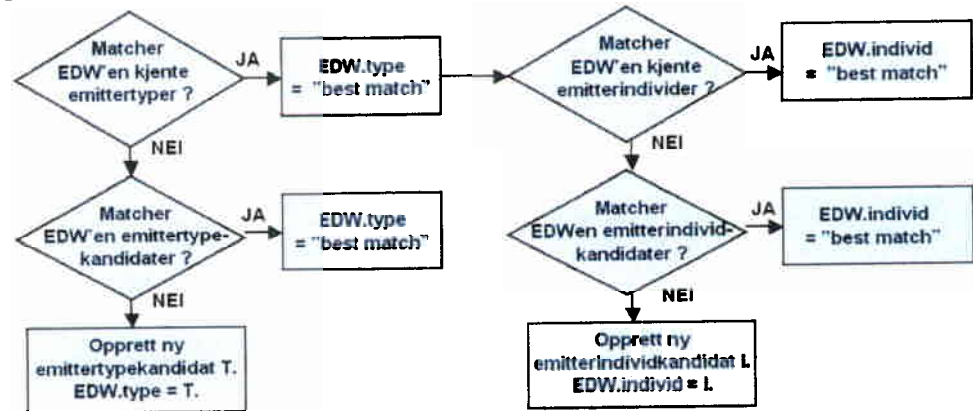
Figur 9.1 Emittertypedatabase og tilhørende emitterindividdatabaser.

Vi har en database for typegjenkjenning, og for hver radartype kan vi ha en egen database for individgjenkjenning. Ved oppstart av "AnalysePC" velges manuelt hvilken typedatabase som skal brukes. Denne databasen leses fra fil og legges inn i passende datastruktur i internminnet. For alle emittertypene i denne databasen lastes deretter automatisk inn tilhørende individdatabaser - dersom disse eksisterer. Databasene er ikke større enn at det er raskest og enklest å lagre alt i internminnet under kjøring.

For både emittertyper og emitterindivider inneholder databasene en angivelse av navn og en

(valgfri) beskrivelse. Dette legges inn manuelt gjennom programmets brukergrensesnitt. De øvrige parameterne i databasene hentes/beregnes automatisk ut fra mottatte EDW'er.

Emittergjenkjenningen starter med typegjenkjenning. Dersom vi klarer å bestemme hvilken emittertype EDW'en tilhører, fortsetter vi med individgjenkjenning. Gjenkjenningsprosedyren er skissert i Figur 9.2.



Figur 9.2 Grovskisse av prosedyren for databasekorrelasjon - starter øverst til venstre. Kandidater opprettes når EDW'en ikke matcher noen av de kjente databaseelementene. Kandidatene finnes kun i internminnet, ikke på fil. Individgjenkjenning forsøkes kun for de kjente emittertypene.

De neste avsnittene beskriver mer detaljert kriterier for når en EDW "matcher" databaseelementer, hvordan "best match" beregnes, hvordan databaseparametere oppdateres og til slutt hvordan implementerte metodene har fungert på testdata, samt hva det bør jobbes videre med.

9.1 Databaseinnhold

Databaseelementene på fil og i internminnet inneholder de samme parameterne. På fil er alle databasene lagret som vanlig tekst (ASCII-format). Det er gjort for å ha muligheten til å editere databasen direkte i en teksteditor, og for å gjøre det enkelt å lese databasen i andre programmeringsspråk som f eks Matlab. Tabellene nedenfor viser innholdet i type- og individdatabaser. Kolonnen "Datatype" angir hvilken datatype parameteren er representert ved internt i programmet (i Java).

Parameter	Datatype	Beskrivelse
name	String	Navn på radarmodell
description	String	Beskrivelse av radarmodell (valgfritt)
NoModes	int	Antall moder
... hver mode er representert ved ...		
freq_avg	double	Frekvens middel
freq_std	double	Frekvens standardavvik
pw_avg	double	Pulsbredde middel
pw_std	double	Pulsebredde standardavvik
NoEDWs	int	Antall EDW'er brukt for beregning av mode-parametere (frekvens, pulsbredde)
NoPDWs	int	Som ovenfor, men i antall PDW'er
stagger	int	Antall stagger nivåer
pri_type	int	Angir type pri-mønster (konstant PRI, stagger eller switched dwell)
... og for hver stagger har vi ...		
pri_avgs	double	PRI middelvei
pri_stds	double	PRI standardavvik
pri_reps	int	Antall PDW'er brukt for beregning av PRI-middel og -standardavvik

Tabell 9.1 Tabell over parametere i typedatabasen.

Parameter	Datatype	Beskrivelse
name	String	Navn på individ
description	String	Beskrivelse av individ (valgfritt)
NoEDWs	int	Antall EDW'er brukt for beregning av pulsformparametere
NoPDWs	int	Som ovenfor, men i antall PDW'er
env_size	int	Antall omhyllingskurve-koeffisienter
env_avgs	double[]	Array med middelveier for omhyllingskurve-koeffisienter
env_stds	double[]	Array med standardavvik for omhyllingskurve-koeffisienter
ifv_size	int	Antall fase-koeffisienter
ifv_avgs	double[]	Array med middelveier for fase-koeffisienter
ifv_stds	double[]	Array med standardavvik for fase-koeffisienter

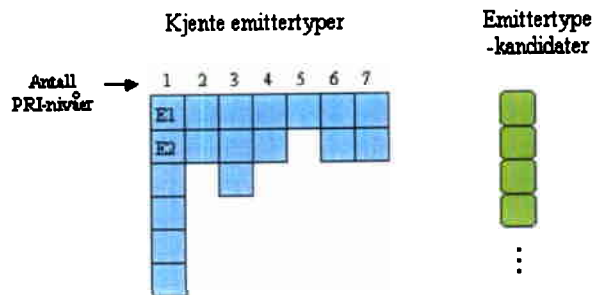
Tabell 9.2 Tabell over parametere i individdatabasen.

9.2 Intern databasestruktur

Felles for både type- og individdatabasen er at de skiller mellom kjente elementer (emittertyper eller emitterindivider) og nye kandidater. De kjente databaseelementene leses inn fra fil i starten av programmet og lagres i egnede datastrukturer som ligger fast under kjøring. Når ukjente emisjoner mottas, opprettes nye emittertype- eller emitterindividkandidater, og disse lagres i datastrukturer som kan endres dynamisk.

En forskjell mellom de to datastrukturene har med korrelasjonsprosedyren å gjøre. Når EDWen korreleres med databasen sammenlignes de først med kjente databaseelementer. Hvis vi finner et kjent databaseelement som matcher EDWen stopper prosedyren opp. Bare dersom EDWen ikke matcher kjente elementer vil vi gå videre og sammenligne med kandidatene. Hvis heller ikke det gir match, vil en ny kandidat opprettes. Mer om dette i avsnitt 9.4.

De kjente emittertypene lagres i en egen datastruktur som er sortert etter antall PRI-nivåer, se Figur 9.3. Dette gjøres for å effektivisere korrelasjonsmetoden. En slipper da å sjekke en EDW mot alle emittertypene når en kun er interessert i emittertyper med samme antall PRI-nivåer. For individdatabasen finnes ikke noe tilsvarende sorterings-kriterium, alle emitterindividene har samme faste antall koeffisienter for omhyllingskurve og fase.



Figur 9.3 Intern datastruktur for lagring av typedatabase omfatter en fast struktur for de kjente emittertypene, og en dynamisk struktur for de nye emittertypekandidatene som opprettes under kjøring.

Informasjonen som lagres om kjente databaseelementer og nye kandidater er den samme, bortsett fra at kandidatene mangler navn og beskrivelse. Nye kandidater godkjennes ved å oppgi navn og en (valgfri) beskrivelse. Ved neste lagring av databasen til fil vil også de godkjente kandidatene tas med.

9.3 Oppdatering

Selv om en radar sender ut pulstog med både konstant bærefrekvens, pulsbredde og PRI, vil det likevel være små variasjoner i de verdier sensoren måler og legger inn i EDWen. Målte verdier for pulsbredde vil f eks være påvirket av signal-til-støy-nivået. I EDWen lagres både middelerdi og standard-avvik for å si noe om *fordelingen* av de målte verdiene gjennom et pulstog. Det samme gjelder for koeffisientene som representerer omhyllingskurve og fase og som brukes for identifikasjon.

Ved oppdatering av og korrelering med databasen antas de underliggende pulsparameterne å være gaussfordelte. Tetthetsfordelingen for en enkelt parameter er da entydig bestemt ut fra middelerdi μ og standardavvik σ ved formelen

$$p(x) = \frac{1}{(2\pi)^{1/2} \sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Med tilleggsantagelsen om ukorrelerte parametere vil den generelle formelen for en n -

dimensjonal gaussfordeling være gitt ved

$$p(x_1, x_2, \dots, x_n) = \frac{1}{(2\pi)^{n/2} \sigma_1 \sigma_2 \dots \sigma_n} \exp\left(-\sum_{i=1}^n \frac{(x_i - \mu_i)^2}{2\sigma_i^2}\right), \quad (9.1)$$

der μ_i og σ_i er de tilhørende middelerverdier og standardavvik. For typedatabasen avhenger dimensjonen n av antall forskjellige PRI-nivåer for den enkelte emittertype - og om vi i tillegg bruker frekvens og/eller pulsbredde. For individdatabasene er dimensjonen n konstant.

Når en emittertypekandidat opprettes vil den få parametere fra kun en EDW. Middelerverdiene og standardavvikene i denne EDWen er ikke eksakt lik parametere for den underliggende emittertypen, men de representerer et estimat for disse (ukjente) parametere. Ettersom kandidaten blir godkjent, lagt inn i databasen, og nye EDWer blir klassifisert som denne emittertypen ønsker vi å forbedre databasens estimater. Jo flere EDWer som brukes, desto sikrere blir estimatene av de ukjente fordelingene.

Gitt at en av parametere til en emittertype E i databasen er representert ved middelerverdi μ_1 og standardavvik σ_1 , og at disse verdiene er basert på totalt n_1 pulser (fra en eller flere EDWer). Anta at en ny EDW klassifiseres som emittertype E , og at de tilsvarende EDW-parametre er henholdsvis μ_2, σ_2 og n_2 . Vi ønsker å finne nye databaseparametere basert på nå til sammen $n = n_1 + n_2$ pulser. Ny middelerverdi μ finnes enkelt ved formelen

$$\mu = \left(\frac{n_1}{n}\right)\mu_1 + \left(\frac{n_2}{n}\right)\mu_2 \quad (9.2)$$

Tilsvarende formel for nytt standardavvik σ kan vises å være

$$\sigma^2 = \left(\frac{n_1}{n}\right)(\sigma_1^2 + d_1^2) + \left(\frac{n_2}{n}\right)(\sigma_2^2 + d_2^2), \quad (9.3)$$

der $d_1 = \mu_1 - \mu$ og $d_2 = \mu_2 - \mu$, dvs d_i er differansen mellom gammel middelerverdi og ny middelerverdi. Ved oppdatering erstattes databasens gamle verdier μ_1, σ_1 og n_1 med de nye verdiene μ, σ og n . For emittertypedatabasen oppdateres PRI (og eventuelt frekvens, pulsbredde) ihht (9.2) og (9.3). Ved oppdatering av emitterindivider skjer tilsvarende oppdatering av koeffisienter for omhyllingskurve og fase.

Vi kan velge å oppdatere databasen automatisk hver eneste gang en EDW klassifiseres eller identifiseres, eller vi kan samle opp EDWer og gjøre oppdatering etter en viss tid eller et visst antall EDWer. Det kan også være aktuelt å sette en begrensning på hvor lenge oppdatering skal foregå. Etter 10000 pulser er det kanskje ikke mer å tjene på videre oppdatering. Et siste alternativ, som er det vi foreløpig har brukt mest, er å gjøre oppdatering kun når brukeren manuelt ber om det.

9.4 Korrelasjon

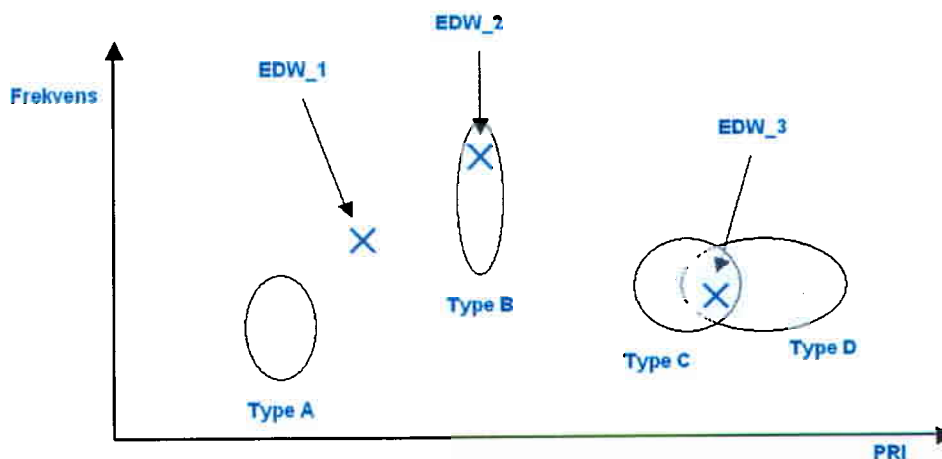
Dette avsnittet beskriver korrelasjonsmetoden for typedatabasen. Individdatabasene bruker foreløpig en enklere korrelasjonsmetode, men i prinsippet skal den samme korrelasjonsmetoden også kunne brukes for individdatabasene.

For å avgjøre hvilken emittertype en gitt EDW kommer fra, sammenlignes parametere i EDWen med parametere i databasen. Vi kan dele opp i tre tilfeller:

- 1) EDWen "matcher" ingen databaseelementer
- 2) EDWen "matcher" ett databaseelement
- 3) EDWen "matcher" flere databaseelementer

I tilfelle 1) vil det opprettes en ny kandidat med parametere fra den gitte EDWen, og EDWen klassifiseres som denne nye kandidaten. I tilfelle 2) er det heller ingen tvil om resultatet. I tilfelle 3) derimot må en se nærmere på hvilket databaseelement som mest *sannsynlig* har gitt opphav til EDWen.

Hvor god "match" det er mellom en EDW og et element i databasen bestemmes av et avstandsmål - som beskrives i neste avsnitt. Deretter beskrives hvordan vi finner "mest sannsynlige" databaseelement.



Figur 9.4 Illustrasjon av de tre mulige alternativene når en EDW korreleres med databasen – ingen match (ny kandidat), entydig match eller flertydig match (overlappende klasser).

9.4.1 Avstandsmål

Vi ønsker å måle avstanden mellom en EDW og en emittertype E , hvor EDWen er representert ved parametere $x = (x_1, x_2, \dots, x_n)$ og emittertype E er representert ved tilsvarende middelveier $\mu_E = (\mu_1, \mu_2, \dots, \mu_n)$ og standardavvik $\sigma_E = (\sigma_1, \sigma_2, \dots, \sigma_n)$. Som avstandsmål bruker vi den såkalte *Mahalanobis-avstanden*. Dette avstandsmålet tar hensyn til at variasjonen eller

usikkerheten i måleverdiene ikke er like stor i alle retninger eller dimensjoner. Vi antar at våre parametere er ukorrelerte, og da reduseres det generelle avstandsmålet til formelen

$$d_E(x) = \sqrt{\sum_{i=1}^n \frac{(x_i - \mu_i)^2}{\sigma_i^2}}. \quad (9.4)$$

For hver komponent (parameter) divideres avviket fra middelveiden med standardavviket, så det vi essensielt måler er totalt antall standardavvik fra middelveidi. Merk at hvis $\sigma_i = 1$ for alle i så er Mahalanobis-avstand identisk med vanlig Euklidsk avstandsmål, dvs $d_E(x) = \|x - u_E\|$.

EDWer tilhørende emittertype E antas å ha gaussfordelte parametere $x = (x_1, x_2, \dots, x_n)$ gitt ved sannsynlighetstettheten i (9.1). Sannsynlighetstettheten $p(y)$ til den stokastiske variabelen $y = d_E^2(x)$ er da kji-kvadratfordelt med n frihetsgrader, dvs

$$p(y) = \frac{1}{2^{n/2} \Gamma(n/2)} y^{(n/2)-1} \exp(-y/2), \quad y > 0 \quad (9.5)$$

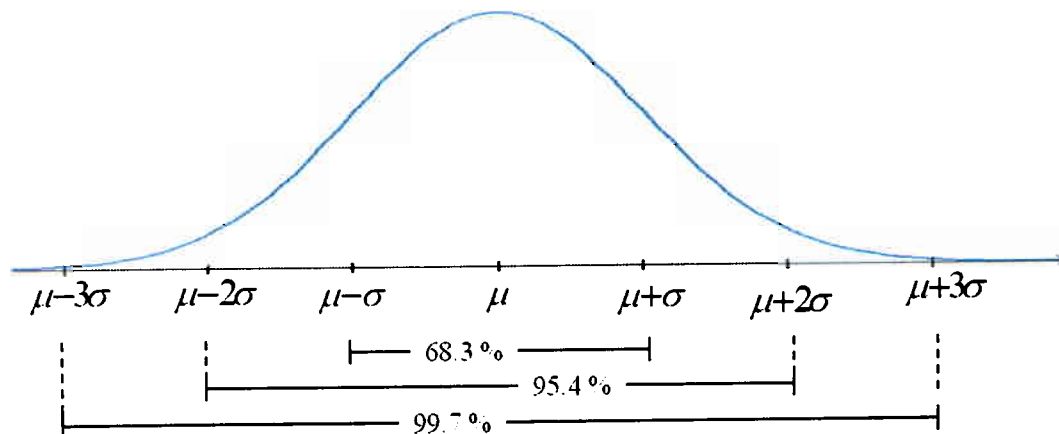
der Γ er gammafunksjonen, og n er antall frihetsgrader. Sannsynligheten for at $d_E(x) < T$ er gitt ved integralet av fordelingsfunksjonen $p(x_1, x_2, \dots, x_n)$ i (9.1) over alle $x = (x_1, x_2, \dots, x_n)$ som tilfredsstillir $d_E(x) < T$. Dette integralet er igjen lik integralet fra 0 til T^2 av fordelingsfunksjonen $p(y)$. Det finnes ingen analytisk formel for disse integralene, men for det siste integralet kan vi gjøre oppslag i tabeller som er beregnet numerisk

#frihetsgrader	90%	95%	99%	99.9%
1	2.71	3.84	6.64	10.83
2	4.61	5.99	9.21	13.82
3	6.25	7.82	11.35	16.27
4	7.78	9.49	13.28	18.47
5	9.24	11.07	15.09	20.52
6	10.65	12.59	16.81	22.46
7	12.02	14.07	18.48	24.32
8	13.36	15.51	20.09	26.13
9	14.68	16.92	21.67	27.88
10	15.99	18.31	23.21	29.59

Tabell 9.3 Utdrag av tabell med kji-kvadrat statistikk

For å illustrere bruken av denne tabellen for å sette terskelen T la oss først se på en 1-dimensjonale gaussfordeling. Figur (9.5) viser velkjente sannsynligheter for sampler innenfor henholdvis 1, 2 og 3 standardavvik fra middelveidi. Nå inneholder ikke Tabell 9.3 kolonner for verken 68.3%, 95.4% eller 99.7%, men hvis den hadde gjort det ville verdiene vært henholdvis 1^2 , 2^2 og 3^2 . For å få 99.9% av samplene fra fordelingen innenfor T standardavvik ser vi fra tabellen at vi må velge $T = (10.83)^{1/2} \approx 3.3$. Jo større verdi vi velger for terskelen T , desto

større andel av samplene fanges opp.



Figur 9.5 1-dimensjonal gaussfordeling. Intervallene angir sannsynlighet for at sampler fra fordelingen vil ligge innenfor henholdsvis 1, 2 og 3 standardavvik fra middelerdi.

Vi har satt kravet til sannsynlighet til 99.9% og valgt terskelen T deretter. For en emittertype med n type-parametere (dvs antall PRI-parametere + eventuelt frekvens og/eller pulsbredde) finner vi verdien for T^2 i rad n og kolonne 99.9%. Risikoen ved å velge for stor terskel er at da øker faren for at nye ukjente emittertyper feilaktig blir klassifisert som en av de kjente emittertypene i databasen.

Normalfordelingene i databasene estimerer de (ukjente) fordelingene til pulsegenskaper for en gitt emittertype eller individ. Som et eksempel kan vi se på parameteren frekvens til en vilkårlig emittertype E i typedatabasen. Estimater av frekvensfordelingen er representert ved normalfordelingen $N(\mu, \sigma)$. Gitt at vi har målt frekvensen x til en enkeltpuls fra emittertype E . Avstanden fra x til μ kan beregnes fra (9.4), og vi kan finne en terskel T for denne avstanden slik at f.eks. 99.9% av de målte frekvensverdiene for pulser fra emittertype E ligger nærmere μ enn terskelen. En EDW derimot inneholder ikke frekvensverdier for enkeltpulser, men frekvensmiddel over alle m pulserne i et pulstog. Hvis EDWen kommer fra emittertype E vil fordelingen til denne middelfrekvensen være gitt ved $N(\mu, \sigma/\sqrt{m})$. Når vi finner Mahalanobis-avstander ved formel (9.4) bruker vi altså ikke databasens standardavvik σ som gjelder enkeltpulser, men standardavviket σ/\sqrt{m} , der m er antall pulser i pulstog. Intuitivt skyldes dette at jo flere pulser det er i pulstog, jo nærmere vil pulstogets middelerdi ligge den riktige middelerdien, og derfor innsnevres avstandskravet tilsvarende.

9.4.2 Overlappende klasser

I tilfellet at en EDW kan komme fra flere databaseelementer (pga overlappende klasser) må vi finne hvilket databaseelement som er det mest sannsynlige. For dette vil vi bruke Bayes beslutningsregel fra statistisk desisjonsteori.

Igen vil vi bruke typedatabasen som eksempel, men samme metode brukes også for individ-databasen. La $\omega_1, \omega_2, \dots, \omega_M$ være de forskjellige klassene (dvs emittertypene) i databasen. For

hver klasse ω_i har vi klassebetingede sannsynlighetstettheter $p(x | \omega_i)$ som angir fordelingen til sample fra klasse ω_i . I tillegg har vi for hver klasse en a priori sannsynlighet $P(\omega_i)$ som angir sannsynligheten for denne klassen. Sannsynligheten for at et gitt sample x tilhører klasse ω_i , $P(\omega_i | x)$, er gitt ved formelen

$$P(\omega_i | x) = \frac{p(x | \omega_i)P(\omega_i)}{p(x)} \quad (9.6)$$

hvor $p(x) = p(x | \omega_1)P(\omega_1) + \dots + p(x | \omega_M)P(\omega_M)$. Bayes beslutningsregel kan nå formuleres som

$$\text{Bayes}_1: \text{ Hvis } P(\omega_i | x) > P(\omega_j | x) \text{ for alle } j \neq i \text{ så hører } x \text{ til klasse } \omega_i \quad (9.7)$$

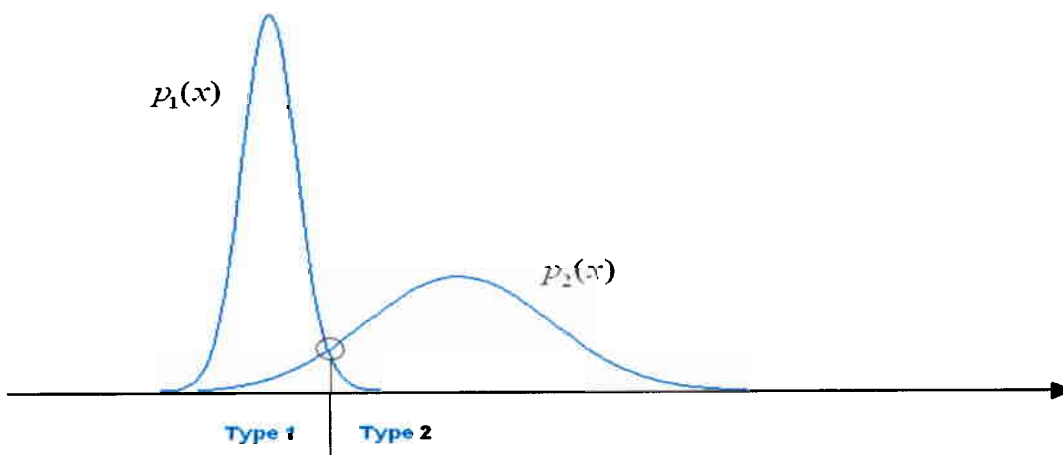
mao, vi tilordner sample x til klasse ω_i hvis denne klassen har størst a posteriori sannsynlighet. Siden leddet under brøkstreken i (9.6) er likt for alle klassene ω_i er (9.7) ekvivalent med regelen

$$\text{Bayes}_2: \text{ Hvis } p(x | \omega_i)P(\omega_i) > p(x | \omega_j)P(\omega_j) \text{ for alle } j \neq i \text{ så hører } x \text{ til klasse } \omega_i \quad (9.8)$$

I mangel av kjennskap til de a priori sannsynlighetene $P(\omega_i)$ antas alle radartypene å forekomme like ofte. Da reduseres (9.8) til

$$\text{Bayes}_3: \text{ Hvis } p(x | \omega_i) > p(x | \omega_j) \text{ for alle } j \neq i \text{ så hører } x \text{ til klasse } \omega_i \quad (9.9)$$

Dette er måten vi avgjør hvilken emittertype som mest sannsynlig er opphav til en EDW.



Figur 9.6 Skillet mellom to klasser (emittertyper eller individer) går der hvor de respektive tetthetsfordelingene har samme verdi.

9.4.3 Modifisert match

Dette avsnittet omhandler problemer spesielt knyttet til PRIer og gjelder kun for typedatabasen. Problemet er at selv om en emitter sender ut samme type pulstog hver gang er det ikke sikkert at EDWene som produseres i sensoren er like. Observerte variasjoner i EDWene omfatter

- 1) Multiple PRIer (pga manglende pulser i pulstoget).
- 2) Oppsplitting av PRIer med stort standardavvik i flere PRIer.
- 3) Manglende PRIer i "switched dwell".

Manglende pulser kan skyldes dårlig signal-til-støy forhold og vil variere med avstand til emitteren. Noen pulstog inneholder PRIer med svært stort standardavvik, og ved få pulser vil clustringsmetoden lett kunne splitte dette opp i flere PRI-nivåer. Ved "switched dwell"-pulstog kan det hende at sensoren ikke mottar alle PRI-nivåene.

Vi ønsker ikke å legge alle disse EDW-variantene inn i databasen. Det vil bli en unødig stor og komplisert database. Vi har valgt å modifisere korrelasjonsmetoden for å håndtere disse tilfellene.

Multiple PRIer.

Der er bare når vi ikke finner matchende databaseelementer til en EDW ved vanlig korrelasjonsprosedyre at vi sjekker om EDWen inneholder multiple PRIer. Hvis vi finner multiple PRIer blir disse (midlertidig) fjernet fra EDWen, og vi gjennomgår matchrutinen på nytt med modifisert EDW. Figur 9.7 viser eksempel på EDW som blir riktig klassifisert til tross for dobbel-PRI.

No	Time	MHz	Angle	NoP	#st	Type	Match	
283	08:11:10.81	9346	183.6	104	1	1714.5	0.137	1714621(75)
284	08:11:11.36	9346	221.1	229	1	1714.5	0.258	1714645(169)
285	08:11:12.63	9346	248.4	465	1	1714.5	0.243	1714642(373)
286	08:11:14.49	9347	268.5	91	2	--NN--		1714584(61) 3810095(9)
287	08:11:15.90	9346	303.7	126	1	1714.5	0.285	1714649(101)
288	08:11:21.06	9346	78.8	294	1	1714.5	0.176	1714629(241)
289	08:11:25.72	9347	187.8	91	2	1714.5	0.003	1714567(50) 3429237(8)

Figur 9.7 Utdrag fra klassifikasjonslogg. Kolonnen "Type" viser navn på emittertype. EDW nummer 289 blir riktig klassifisert til tross for ekstra cluster med dobbel-PRI. (EDW nummer 286 blir ikke klassifisert pga ekstra-PRI som antagelig er sammenslåing av dobbel- og trippel-PRI til ett cluster).

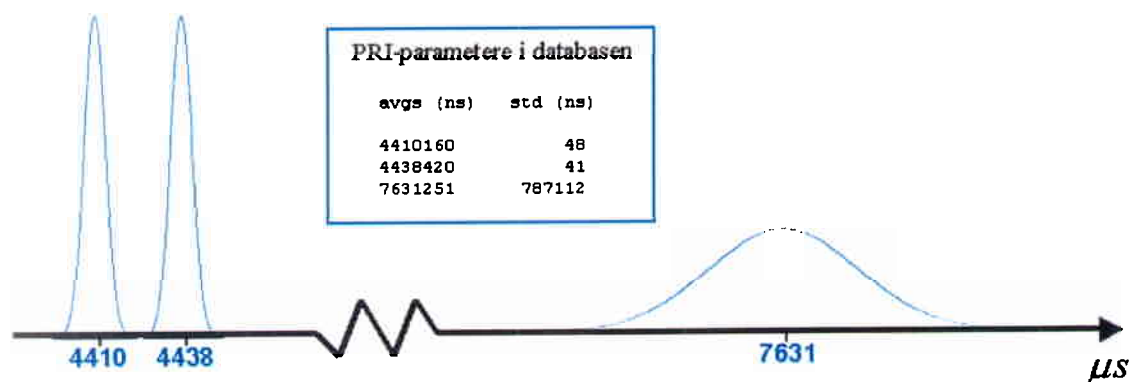
Oppsplitting av PRIer med stort standardavvik.

Det er PRI-nivåene med stort standardavvik som clustringsmetoden har lettest for å splitte opp i flere PRIer. Vi begrenser oss derfor til å kontrollere disse PRI-nivåene. Vi kan velge å kontrollere alle PRI-nivåer med standardavvik større enn en viss terskel – eller vi kan manuelt markere i databasen de PRI-nivåene vi vil kontrollere. Når en EDW korreleres med et slikt database-element vil vi (midlertidig) slå sammen flere av PRI-nivåene i EDWen til ett PRI-nivå før vi korrelerer med databasen på nytt på vanlig måte. Hvilke PRI-nivåer i EDWen som slås sammen avhenger av hvilken av PRIene i databaseelementet som har stort standardavvik, og

dennes plassering i forhold til de andre PRI-verdiene. Figur 9.8 viser eksempel på at EDW'er blir riktig klassifisert til tross for oppsplitting av PRI-nivå i flere PRI-clustere. De "riktige" PRI-nivåene i databasen er vist i Figur 9.9.

No	#st	Type	Match					
106	4	4410++	0.039	4410176 (19)	4438441 (18)	7277743 (9)	8812737 (10)	
107	5	4410++	0.048	4410178 (41)	4438448 (39)	7225761 (8)	7423430 (9)	8598711 (21)
108	4	4410++	0.067	4410179 (21)	4438338 (21)	7426223 (13)	9203914 (11)	
109	4	4410++	0.042	4410175 (31)	4438448 (30)	7284732 (15)	8614352 (17)	
110	3	4410++	0.042	4410182 (23)	4438445 (23)	7302702 (12)		
111	4	4410++	0.055	4410177 (22)	4438447 (22)	7229874 (11)	9014623 (12)	

Figur 9.8 Utdrag fra klassifikasjonslogg (noe redigert for å få plass til alle PRI-nivåene). Alle EDW'ene klassifiseres riktig til tross for at øverste PRI-nivå ofte deles opp i flere PRI-clustere.



Figur 9.9 Eksempel på emittertype "4410++" med tre PRI-nivåer. Standardavviket for det høyeste PRI-nivået er mer enn 10000 ganger større enn for de to andre PRI'ene (grafene viser ikke dette forholdet riktig). Med få pulser i pulstøget vil ofte det øverste nivået bli splittet opp i flere PRI-clustere.

Manglende PRI'er i "switched dwell".

Dette tilfellet er ikke ferdig uttestet, men her inneholder databaseelementene informasjon om at match-kriteriet skal være mindre strengt. Et "switched dwell"-databaseelement med f.eks. 6 PRI-nivåer kan angi at en EDW bare trenger 4 av de 6 PRI-nivåer for å matche databaseelementet.

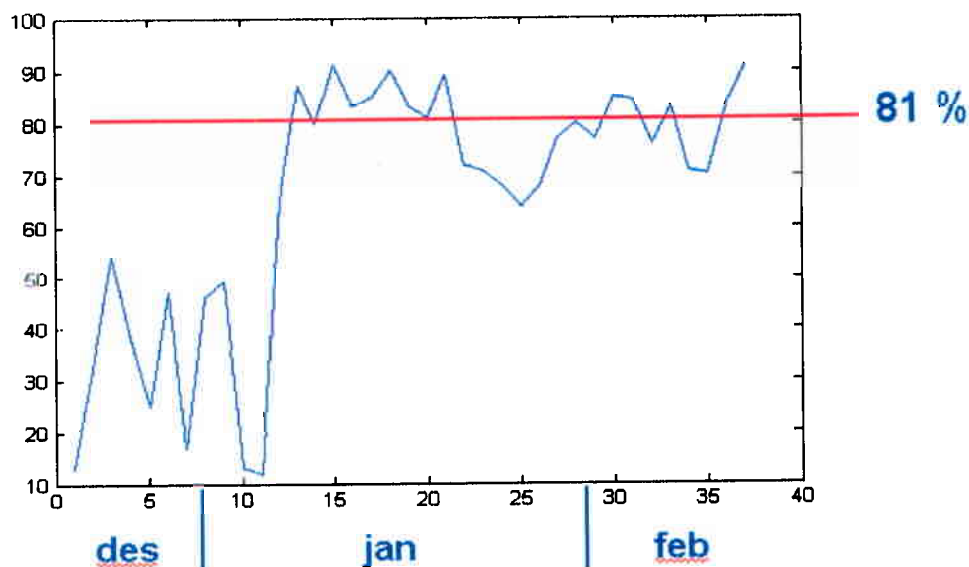
9.5 Oppsummering og resultater

Enkel databasefunksjonalitet er implementert og fungerer tilfredsstillende:

- *Databaseutvidelser.* Omfatter automatisk oppretting av emittertype- og individkandidater, samt rutiner for å godkjenne kandidater med påfølgende lagring til fil.
- *Databaseoppdatering.* Omfatter rutiner for oppdatering av middelvei og standardavvik for de enkelte databaseparametere, basert på tilsvarende parametere i mottatte EDWer.
- *Korrelasjonsmetoder.* Vi har metoder som fungerer bra for de emittere vi så langt har dataopptak av.

På noen områder har vi implementert alternative metoder som vi må velge mellom basert på hva som fungerer best i praksis. F eks kan vi velge å oppdatere databaseparametere automatisk for hver eneste EDW som bli gjenkjent, eller vi kan samle opp EDWer før vi gjør en felles oppdatering. Det siste alternativet er minst regnekrevende, og er det som er blitt brukt mest hittil. Når det gjelder valg av hvilke parametere som skal brukes til typegjenkjenning er databasene forberedt for å bruke både frekvens, pulsbredde og PRI. Vi fant imidlertid fort ut at pulsbredde ikke var noen god parameter for typegjenkjenning. Målt pulsbredde kan variere stort avhengig av flere årsaker, f eks variasjon i signal-til-støy forholdet. Vi valgte derfor å ikke bruke pulsbredde i typegjenkjenning. I testresultatene nedenfor har vi heller ikke brukt frekvens, selv om det er en parametere vi på sikt ønsker å ta i bruk.

I perioden fra desember 2002 til februar 2003 har vi gjort daglige opptak av værradarene til sivile passasjerfly i nærheten av FFI. Underveis i denne perioden er typedatabasen blitt utvidet med nye emittertyper, og både metoden for å beregne EDWer og metoden for å korrelere EDWer med databasen er forbedret. Figur 9.10 viser hvordan andelen av klassifiserte EDWer har variert i denne perioden. Typedatabasen inneholder nå i overkant av 20 emittertyper.



Figur 9.10 Oversikt over andelen av mottatte EDWer som ble klassifisert i perioden desember 2002 til februar 2003. Den røde streken viser andelen klassifiserte EDWer som ble oppnådd når alle EDWer i ettetid ble klassifisert på nytt med siste versjon av typedatabasen.

Med den begrensede typedatabasen vi hittil har bygd opp, og med kravet om at 99.9% av samplene fra en fordeling skal gjenkjennes (som brukes for å sette terskelen for Mahalanobis-avstanden), så vet vi at de aller fleste emittertyper i databasen ikke overlapper. Vi er sikre på at EDW'er fra disse klassene ikke blir feilklassifisert. Databasen inneholder imidlertid også noen få emittertyper med konstant PRI som ligger så nær hverandre at feilklassifisering er en reell mulighet. Et omtrentlig anslag ut fra inspeksjon av klassifikasjonsloggen er at omtrent en promille av EDW'ene kan være feilklassifisert. For om mulig å redusere denne andelen vil vi se på om vi ved å introdusere frekvens i typegjenkjenningen klarer å skille klassene bedre fra hverandre.

Noen hovedområder for videre arbeid med databaser for emittergjenkjenning :

- Fortsatt databaseutvidelse. Blant de sivile værradarene vi har flest opptak av er det fremdeles enkelte "sjeldne" emittertyper (moder) som ikke er lagt inn i typedatabasen. Det er også mulig å forbedre både korrelasjonsmetoden og metoden for PRI-clustering.
- Foreløpig inneholder emittertyper i typedatabasen kun en mode. Vi bør finne ut hvilke moder som hører til samme emittertype og slå disse sammen. Databaserepresentasjonen er forberedt for emittertyper med flere moder. Det vil gjøre påfølgende emisjonstracking mer robust mot modeskift.
- Utvide typegjenkjenning med frekvens i tillegg til PRI.

APPENDIKS

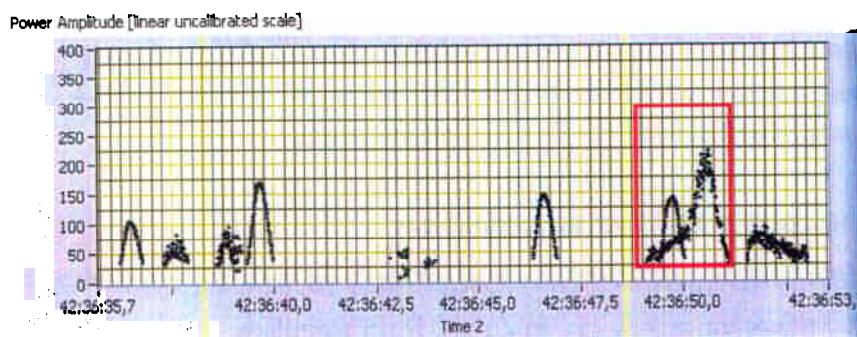
A.1 Parametersett i LABVIEW-koden

Parametervektor	LabView-rutine	Setting	Hensikt
[Rank, threshold, number_of_bins]	Fclusters.vi	5.0, 0.4, 200	Generere frøpunkter, for frekvensclustering
[Max PRI]	PDW_to_EDW.vi	3.0	Ekskludere effekter fra manglende pulser
[Min PRI]	PDW_to_EDW.vi	10 mikrosek.	Eliminere effekter av pulsdrops
[Rank, threshold, number_of_bins]	Cluster.vi	Level1: 5.0, 0.4, 100 Level2: 5.0, 0.5, 50	Generere frøpunkter for hvert av de to nivåene i hierarkisk clustering for PRI.
[Nsigma]	Cluster.vi	2	Beregne middelerverdi innenfor +/- N x sigma i fordelingen (per cluster). Eliminere effekt av "outliers"-merkbart når antall datapunkter er mindre enn 100.
[Min_Npri]	Filter_PRI.vi	0.1	Om Npri < fact x Max(Npri) slettes tilhørende cluster.
[alpha]	PRI.vi	5	Melder switched dwell når sentercluster har <i>alpha</i> ganger mer tellinger enn i sidecluster

Tabell A.1 Uttestede parametere for puls-sortering. Disse parameterne bør "tunes" ved endring av emitterscenario.

A.2 Sortering på amplitude

I Figur A.1 vises et eksempel på muligheten for deinterleaving med hensyn på amplitude. Den røde rammen angir pulsamplitude for funksjon av tid for to overlappende signaler. Ved å følge kurvene (med støy) vil man i prinsippet kunne skille de to sveipene fra hverandre. En vil søke etter å trekke ut kontinuitetsegenskapene til en glatt kurve, ved for eksempel å beregne avstanden mellom predikert og virkelig amplitude. Den predikerte amplituden kan baseres på et vektet middel av foregående amplituder, for eksempel via en minste kvadraters tilpasning til et lavere ordens polynom.



Figur A.1 Den røde rammen angir pulsamplitude for funksjon av tid for to overlappende signaler. Ved å følge kurvene (med støy) vil man i prinsippet kunne skille de to sveipene fra hverandre.

A.3 Legendre-polynomier

Legendre-polynomene P_k er definert som løsningene av differensialligningen

$$\frac{d}{dt}[(1-t^2)\frac{dP_k(t)}{dt}] + k(k+1)P_k(t) = 0, \quad -1 \leq t \leq 1$$

som tilfredsstillers $P_k(1) = 1$. Vi har relasjonen

$$(k+1)P_{k+1}(t) = (2k+1)tP_k(t) - kP_{k-1}(t)$$

som kan brukes rekursivt til å bestemme polynomene av høyere orden. Legendre-polynomene utgjør en ortogonal basis for $L_2[-1,1]$ med vektingsfunksjon $w(t) \equiv 1$. De ortonormaliserte Legendre-polynomene er gitt som

$$p_k(t) = \sqrt{\frac{2k+1}{2}}P_k(t), \quad k = 0, 1, 2, \dots$$

De første fire ortonormaliserte Legendre-polynomene er gitt i Tabell A.2

$p_0(t)$	$\frac{1}{\sqrt{2}}$	$p_2(t)$	$\frac{1}{2}\sqrt{\frac{5}{2}}(3t^2 - 1)$
$p_1(t)$	$\frac{\sqrt{3}}{\sqrt{2}}t$	$p_3(t)$	$\frac{1}{2}\sqrt{\frac{7}{2}}(5t^3 - 3t)$

Tabell A.2 De ortonormaliserte Legendre-polynomene opp til grad 3.

Legendre-polynomene har pene approksimasjonsegenskaper. En artikkel av Yin et al. (18) behandler avstanden mellom funksjonene f og \tilde{f}_N , hvor \tilde{f}_N er den trunkerte Legendre-approksimasjonen (med $N+1$ ledd) til f pålagt støy med middel 0 og varians σ^2 og digitalisert med samplingsrate Δ . Da gjelder følgende ulikhet for den forventede avstanden i L_2 -norm mellom f og \tilde{f}_N :

$$E \|\tilde{f}_N - f\|_2^2 \leq \left[\frac{A}{N^p} + B\Delta^2 \right]^2 + C\sigma^2\Delta(N+1)$$

Her er A , B og C konstanter som avhenger av f og parameteren p som representerer glattheten til f ; jo glattere f , desto større p . Vi legger merke til at høyresiden i ulikheten har et veldefinert minimum som funksjon av N . Dette betyr at det i prinsippet bør være mulig å bestemme det optimale antall koeffisienter $N+1$ i pulsrepresentasjonen.

A.4 Sentrale og ikke-sentrale kji-kvadrat-fordelinger

La $a_k, k = 1, 2, \dots, n$, betegne normalfordelte variable med middel 0 og varians 1 (dersom a_k har syandardavvik $\rho_k \neq 1$ kan variabelen $\tilde{a}_k = a_k / \rho_k$ innføres). Da er størrelsen

$$D = \sum_{k=1}^n a_k^2 \sim \chi_n^2,$$

hvor χ_n^2 betegner en kji-kvadrat-fordeling med frihetsgrad n . Kji-kvadrat-fordelingen har sannsynlighetstettheten

$$f_n(y) = \frac{1}{2^{n/2} \Gamma(n/2)} y^{n/2-1} \exp(-y/2)$$

hvor Γ er gammafunksjonen og n er antall frihetsgrader. Forventningen og variansen til denne fordelingen er henholdsvis n og $2n$.

La nå $b_k, k = 1, 2, \dots, n$, betegne normalfordelte variable med middel λ_k og varians 1 (igjen, dersom b_k har standardavvik $\rho_k \neq 1$ kan variabelen $\tilde{b}_k = b_k / \rho_k$ innføres). Da er størrelsen

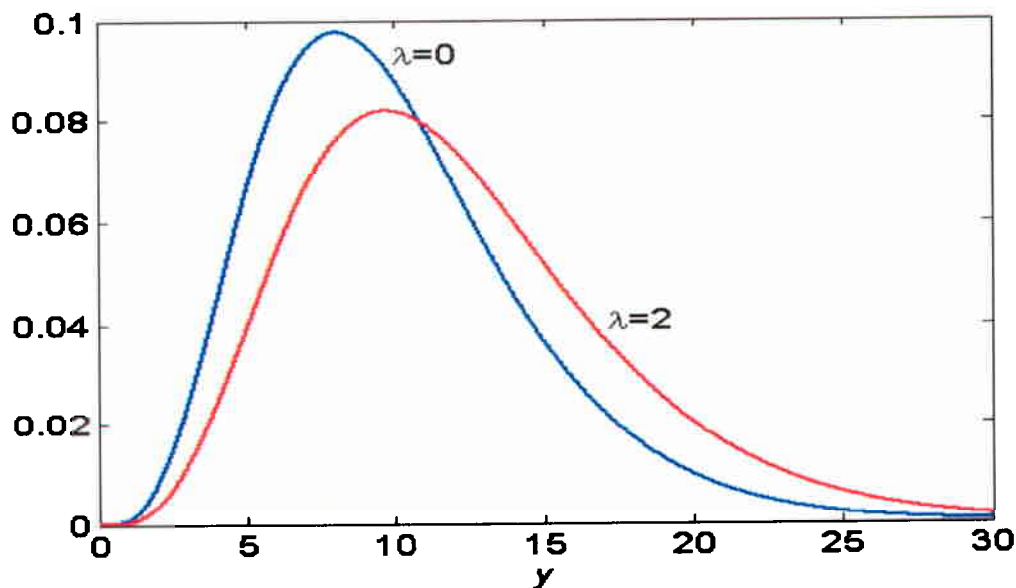
$$D = \sum_{k=1}^n b_k^2 \sim \chi_{n,\lambda}^2,$$

hvor $\chi_{n,\lambda}^2$ betegner en *ikke-sentral* kji-kvadrat-fordeling med frihetsgrad n og ikke-sentral faktor

$\lambda = \sum_{k=1}^n \lambda_k^2$. Denne fordelingen har sannsynlighetstetthet

$$g_{n,\lambda}(y) = f_n(y) \exp(-\lambda/2) \left[1 + \sum_{k=1}^{\infty} \frac{(\lambda y / (2n))^k}{k! \prod_{j=0}^{k-1} (1 + 2j/n)} \right].$$

Forventningen og variansen til denne fordelingen er henholdsvis $n + \lambda$ og $2(n + 2\lambda)$. Legg merke til at den sentrale kji-kvadrat-fordelingen svarer til spesialtilfellet av den ikke-sentrale kji-kvadrat-fordelingen for $\lambda = 0$. Figur A.2 viser en sammenligning av vanlig kji-kvadrat- og ikke-sentral kji-kvadrat-fordelingen.



Figur A.2 Grafen til de ikke-sentrale kji-kvadrat-fordelingene med frihetsgrad 10 for $\lambda = 0$ (blå kurve, tilsvarende den vanlige kji-kvadrat-fordelingen) og $\lambda = 2$ (rød kurve).

A.5 Emittorgjenkjenning

Dersom f_N , representert ved koeffisientene a_k , er en databasepuls (amplitude eller fase) og $\tilde{f}_N^{(p)}, p = 1, 2, \dots, M$, er en mengde innkommende pulser (amplitude eller fase) som hver er representert ved koeffisientene $\tilde{a}_k^{(p)}$, kan λ (definert i Appendiks A.4) estimeres ved å beregne

middel og varians av størrelsene

$$d^{(p)} = \frac{1}{\sigma^2 \Delta} \sum_{k=1}^N (\tilde{a}_k^{(p)} - a_k)^2, p = 1, 2, \dots, M.$$

Siden hver av størrelsene $d^{(p)}$ har forventning $N + \lambda$ og varians $2[N + 2\lambda]$ kan vi beregne to estimater $\hat{\lambda}_1$ og $\hat{\lambda}_2$ for λ ;

$$\hat{\lambda}_1 = -N + \frac{1}{M} \sum_{p=1}^M d^{(p)},$$

$$\hat{\lambda}_2 = \frac{1}{4} \left\{ -2N + \frac{1}{M-1} \sum_{p=1}^M (d^{(p)} - \overline{d^{(p)}})^2 \right\}, \overline{d^{(p)}} = \frac{1}{M} \sum_{p=1}^M d^{(p)}.$$

Dersom estimatene er signifikant større enn 0 kan vi konkludere med at innkommende puls og databasepuls ikke er like og derfor ikke kommer fra samme emitter. En annen og mer robust, men også mer regnekrevende, metode er å plote histogrammet for $d^{(p)}$, $p = 1, 2, \dots, M$, og så, ved bruk av minste kvadraters metode, bestemme verdien av λ som gjør at sannsynlighetstettheten til den ikke-sentrale kji-kvadrat-fordelingen passer best til histogrammet.

A.6 Følsomhet for multipath

Som illustrert i Figur 8.5 er fasen mindre påvirket av multipath enn amplituden. Vi vil forsøke å begrunne dette med en overfladisk matematisk analyse. Vi tenker oss at vi har en puls beskrevet ved modellen

$$S_{dir}(t) = A(t) \exp(i\varphi(t)) = A(t) \exp(i(\omega_0 t + \varphi_0(t)))$$

hvor $A(t)$ er pulsamplituden, ω_0 er bærefrekvensen og $\varphi_0(t)$ er fasens ikkelineære komponent, som tidligere. S_{dir} representerer et direkte signal som ikke er påvirket av multipath. La oss anta at vi i tillegg har en komponent S_{ref} som er reflektert og som derfor er forsinket med en tidsforsinkelse τ i forhold til det direkte signalet. Det mottatte signalet S_{rec} blir da

$$S_{rec}(t) = S_{dir}(t) + S_{ref}(t) = A(t) \exp(i(\omega_0 t + \varphi_0(t))) + \alpha A(t + \tau) \exp(i(\omega_0(t + \tau) + \varphi_0(t + \tau))).$$

Amplituden og fasen til S_{rec} er

$$\text{Amplitude} = \{A(t)^2 + 2\alpha A(t)A(t + \tau) \cos(\omega_0 \tau + \varphi_0(t + \tau) - \varphi_0(t)) + \alpha^2 A(t + \tau)^2\}^{1/2},$$

$$\text{Fase} = \arctan \left\{ \frac{A(t) \sin(\omega_0 t + \varphi_0(t)) + A(t + \tau) \sin(\omega_0(t + \tau) + \varphi_0(t + \tau))}{A(t) \cos(\omega_0 t + \varphi_0(t)) + A(t + \tau) \cos(\omega_0(t + \tau) + \varphi_0(t + \tau))} \right\}.$$

Ved å anta at τ er et lite tall og at amplituden og fasen i det opprinnelige signalet endrer seg lite over små tidsintervaller får vi tilnærmingene $A(t + \tau) \approx A(t)$ og $\varphi_0(t + \tau) \approx \varphi_0(t)$. Dermed blir et tilnærmet uttrykk for amplituden

$$\text{Amplitude} \approx A(t) \{1 + 2\alpha \cos(\omega_0 \tau) + \alpha^2\}^{1/2}$$

og for fasen

$$\text{Fase} \approx \arctan \left\{ \frac{\tan f + \alpha(c \tan f + s)}{1 + \alpha(c - s \tan f)} \right\},$$

$$f = f(t) = \omega_0 t + \varphi_0(t), c = \cos(\omega_0 \tau), s = \sin(\omega_0 \tau).$$

La nå $p(f)$ betegne fasen som funksjon av f . Derivasjon av p gir

$$p'(f) = \frac{(1 + \tan^2 f)[(1 + \alpha c - \alpha s \tan f)(1 + \alpha c) + \alpha s((1 + \alpha c) \tan f + \alpha s)]}{[1 + (\frac{\tan f + \alpha(c \tan f + s)}{1 + \alpha(c - s \tan f)})^2](1 + \alpha(c - s \tan f))^2} = 1.$$

Mao er $p(f) = f + C$, hvor $C = p(0) = \arctan(\alpha s / (1 + \alpha c))$. Skrevet med de opprinnelige variablene får vi

$$\text{Fase} \approx \omega_0 t + \varphi_0(t) + \arctan\left\{\frac{\alpha \sin(\omega_0 \tau)}{1 + \alpha \cos(\omega_0 \tau)}\right\}.$$

Litteratur

- (1) Malnes E, Eilivstjønn J, Kulbotten G, Kvernsveen K, Øhra H (2000): Spesifikasjon av ESM-demonstrator, FFI/RAPPORT-2000/06388, Ugradert
- (2) Hanssen L, Danielsen R (2003): Karakterisering av DIPP-kort, FFI/NOTAT-2003/00095, Begrenset
- (3) Sundgot R, Øyehaug L (2002): Metoder for emitteridentifisering i en ESM-demonstrator, FFI/RAPPORT-2003/00555, Ugradert
- (4) Hämmerlin G, Hoffmann, K-H (1990): Numerical Mathematics, Springer-Verlag, New York.
- (5) Strang G, Nguyen T (1997): Wavelets and Filter Banks, Wellesley-Cambridge Press, Wellesley.
- (6) Everitt, B S (1993): Cluster Analysis, Halsted Press, New York.
- (7) Mardia, H K (1989): New techniques for the deinterleaving of repetitive sequences, *IEE proceedings* **136**, 4, 149.
- (8) Davies, C L and Hollands, B A (1982): Automatic processing for ESM, *IEE proceedings* **129**, 3, 164.
- (9) Malnes E, Berg, Smestad T (1998): Metoder for sortering av pulstog (deinterleaving), FFI/RAPPORT-98/01476, Ugradert
- (10) Malnes E (2000): Pulssortering og emitteridentifikasjon, FFI/RAPPORT-2000/02428, Ugradert
- (11) Mjanger, M (2002): prosjekt KREST, privat kommunikasjon.
- (12) Øhra, H (2003): prosjekt KREST, privat kommunikasjon.
- (13) Danielsen, R (2003): prosjekt PROFIL, privat kommunikasjon.
- (14) Hammerstad, J (2002): prosjekt KREST, privat kommunikasjon.
- (15) Hamran S-E, Danielsen R, Hanssen L, Haugen R, Knapskog A, Skartlien R, Smestad T, Sundgot R, Øyehaug L (2003): PROFIL - Sluttrapport, FFI/RAPPORT-2003/00886, Begrenset
- (16) Online (2003): <http://www.mathworld.com>
- (17) Eilevstjønn J (2001): FIA - programvare for klassifisering og identifikasjon av radarpulser, FFI/NOTAT-2002/02261, Ugradert
- (18) Yin J, De Pierro A R, Wei M (2002): Analysis for the reconstruction of a noisy signal based on orthogonal moments, *Appl Math Comp* **132**, 249-263.

FORDELINGSLISTE

FFIE

Dato: 7 mai 2003

RAPPORTTYPE (KRYSS AV)		RAPPORT NR.	REFERANSE	RAPPORTENS DATO	
<input checked="" type="checkbox"/> RAPP	<input type="checkbox"/> NOTAT	<input type="checkbox"/> RR	2003/00881	FFIE/805/113	7 mai 2003
RAPPORTENS BESKYTTELSESGRAD			ANTALL TRYKTE UTSTEDT	ANTALL SIDER	
UGRADERT			35	49	
RAPPORTENS TITTEL METODER FOR PULSSORTERING OG EMITTERGJENKJENNING I ESMEX.			FORFATTER(E) SKARTLIEN Roar, SUNDGOT Rune, ØYEHAUG Leiv		
FORDELING GODKJENT AV FORSKNINGSSJEF Torleiv Maseng <i>T. Maseng</i>			FORDELING GODKJENT AV AVDELINGSSJEF: Johnny Bardal <i>Johnny Bardal</i>		

EKSTERN FORDELING

INTERN FORDELING

ANTALL	EKS NR	TIL	ANTALL	EKS NR	TIL
			5		FFI-Bibl
			1		FFI-ledelse
1		LUKS/LOI	1		FFIE
1		v/Ob Jon R Krogstad	1		FFISYS
1		v/Oblt Erik Pande-Rolfsen	1		FFIBM
			1		FFIN
1		FO/FST	3		Forfattereksemplarer
1		v/Oblt Gaute Dyrdal	1		Svein-Erik Hamran, FFIE
1		v/Oblt Geir Stadsvoll	1		Atle Knapskog, FFIE
			1		Tore Smestad, FFIE
1		FO/E	1		Reidar Haugen, FFIE
1		v/Rådgiver Tore Belsnes	1		Leif Hanssen, FFIE
			1		Roar Danielsen, FFIE
1		FLVUKS	1		Hans Øhra, FFIE
1		v/Oblt Willy Borg-Lundquist	1		Morten Mjanger, FFIE
			1		Jan Hammerstad, FFIE
1		FLO/LUFT-DEE	1		Svein Hauge, FFIE
1		v/Maj Nils Aa. Østby			
					Elektronisk fordeling:
					FFI-veven

Benytt ny side om nødvendig.