

Assosiasjon mellom posisjonsinformasjon og peilinger fra en satellitt over nordområdene

Tor Inge Birkenes

Lektorutdanning med master i realfag

Innlevert: Juni 2012

Hovedveileder: Håkon Tjelmeland, MATH

Medveileder: Oddvar Hallingstad, UNIK
Tore Smestad, Forsvarets forskningsinstitutt

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Oppgavebeskrivelse

Forsvarets forskningsinstitutt (FFI) har et satellittsystem for å posisjonere skip basert på “Automatisk Identifikasjonssystem” (AIS) og ønsker å komplettere dette med peilinger basert på skipenes navigasjonsradar. For at dette skal være mulig kreves det god kunnskap om satellittens stilling. I masteroppgaven skal kandidaten undersøke i hvilken grad det er mulig å forbedre kunnskapen om satellittens stilling ved å assosiere AIS-posisjoner og radarpeilinger.

Sammendrag

Forsvarets forskningsinstitutt (FFI) har ønske om å utvide eksisterende satellittovervåking av nordområdene fra kun posisjonsinformasjon mottatt via “Automatisk indentifikasjonssystem” (AIS) til å også inkludere radarpeilinger. Dette krever nøyaktig kunnskap om satellittens stilling og denne masteroppgaven undersøker om det er mulig å oppnå dette ved å assosiere AIS-posisjoner med radarpeilingene. Det er utarbeidet en matematisk modell for problemet og estimatorer basert på nærmeste nabo (NN) og global nærmeste nabo (GNN) assosiasjon er definert. Estimatorenes statistiske egenskaper er undersøkt ved simuleringer som inkluderer realistiske data av skipsposisjoner og AIS-data. Når usikkerheten i peilevinklene blir satt til 0.10 grader oppnår GNN-estimatoren nøyaktighet på 0.012–0.041 grader for to av rotasjonsaksene til satellitten og 0.25–0.40 grader for den siste aksen. Det er beregnet tilnærmede Cramér Rao nedre grenser, som i dette tilfellet er 0.0012–0.041 grader for de to første aksene og 0.065–0.46 for den siste. Likheten mellom empiriske og teoretiske nøyaktigheter viser at estimatoren er optimal eller nær optimal. Størrelsen på usikkerheten i peilevinklene er funnet proporsjonal med de empiriske standardavvikene for GNN-estimatoren. Under estimeringen ble 50–90% av AIS-posisjonene assosiert korrekt med radarpeilingene. Det er funnet ut at GNN estimerer med høyere presisjon enn NN, men NN er raskere å beregne og kan med fordel brukes til å finne et grovestimat som GNN kan bruke som utgangspunkt. Nøyaktigheten i assosiasjon og stillingsestimat er bedre i åpent farvann enn nær kysten, noe som ser ut til å ha sammenheng med større tetthet av skip og at det mottas AIS-posisjoner fra en lavere andel skip. Resultatene viser at det ved de undersøkte forholdene er mulig å oppnå god kunnskap om satellittens stilling ved å assosiere AIS-posisjoner og radarpeilinger.

Abstract

The Norwegian Defence Research Establishment (FFI) wishes to expand existing satellite surveillance from only position information received by “Automatic Identification System” (AIS) to also include radar bearings. This requires precise knowledge of the satellites orientation. This master thesis explores the possibility of improving the knowledge about the orientation by associating AIS positions with radar bearings. To achieve this it is developed a mathematical model of the problem and estimators based on the nearest neighbor (NN) and global nearest neighbor (GNN) methods is defined. The estimators statistical properties are estimated by simulations that incorporate realistic ship positions and AIS data. When the magnitude of the noise in the bearings is 0.1 degrees the estimator based on GNN achieved a precision of 0.012–0.041 degrees for two rotation axes and 0.25–0.40 degrees for the last axis. Estimated Cramér Rao lower bounds are calculated and is in this case 0.0012–0.041 degrees for the first two axis and 0.065–0.46 degrees for the last axes. The similarity of empirical and theoretical precisions show that the estimators are optimal or near optimal. The precision of the estimation of the orientation has been shown to be proportional to the magnitude of the noise in the bearings. During the estimation 50–90% of the AIS positions were correctly associated with the corresponding radar bearings. It has been discovered that GNN estimates has higher precision than NN estimates, but the NN estimates are faster to compute. Therefore the speed of the estimation process can be improved by finding a rough estimate by NN and using this as a starting point for GNN. The precision of association and estimation of orientation is better in the open sea than near the coast. This seems to be connected to higher density of ships and reception of AIS messages from a lower portion of the ships. The results show that it is possible to improve the knowledge of the orientation of the satellite by associating AIS positions with radar bearings in the examined situations.

Forord

Denne masteroppgaven er skrevet som avslutning av studieprogrammet “Lektorutdanning med mastergrad i realfag” ved NTNU våren 2012. Min spesialisering har vært i matematikk og fysikk, med hovedvekt på og masteroppgave i matematikk.

Opgaven er skrevet i samarbeid med Forsvarets forskningsinstitutt (FFI) på Kjeller og bygger delvis på resultater av eksperimenter jeg har vært med på gjennom to sommerjobber hos FFI i 2010 og 2011. Hensikten har vært å undersøke om det kan være mulig å få bedre kunnskap om stillingen til en overvåkingssatellitt ved å samordne radarpeilinger med data mottatt gjennom AIS-systemet. Undersøkelsen er utført ved å bygge en matematisk modell av problemet og simulere det på en datamaskin.

Under arbeidet har det vært ett besøk til FFI på Kjeller, ellers har arbeidet foregått ved NTNU i Trondheim. Hovedveileder har vært professor Håkon Tjelmeland fra Institutt for matematiske fag, NTNU. Gjennom ukentlige møter har han vært til stor hjelp ved utvikling av metoder, beskrivelse av dem og generelt alle andre deler av oppgaveskrivingen. Biveileder og forsker Tore Smestad fra FFI har vært min kontaktperson hos dem. Gjennom samtaler på telefon og e-post har han bidratt med gode idéer til fremgangsmåter og realistiske rammer for problemet. Under besøket på FFI fikk jeg god informasjon om viktige egenskaper til satellitten gjennom et møte med forsker Øystein Olsen, som også utførte en meget nyttig simulering slik at jeg kunne få dataene fra den. Jeg vil også takke forsker Brita Hafskjold Gade som stilte opp på meget kort varsel og bidro til utviklingen av global nærmeste nabo algoritmen som er meget sentral i oppgaven.

Min andre biveileder Oddvar Hallingstad var til god hjelp i etableringen av prosjektet.

Til slutt vil jeg gjerne takke min kjæreste Nina for korrekturlesing og mye annen støtte.



Tor Inge Birkenes

Innhold

Sammendrag	i
Abstract	ii
Forord	iii
Innhold	iv
Figurer	vi
Tabeller	viii
Liste over symboler	x
1 Innledning	1
2 Matematisk bakgrunn	3
2.1 Notasjon og definisjoner	3
2.2 Derivasjon av matriser	3
2.3 Geometri	4
2.4 Minste kvadraters metode	8
2.5 Nærmeste nabo- og global nærmeste nabo-metoden	9
2.6 Transformasjon av sannsynlighetstetthetsfunksjoner	11
2.7 Normalfordelingen for vektorer	11
2.8 Cramér Rao nedre grense ved normalfordelt støy	13
2.9 Bayesestimering og Metropolis-Hastings algoritmen	13
3 Modellering	17
3.1 Koordinatsystemer	17
3.2 Peilinger	19
4 Estimeringsmetoder	25
4.1 Minste kvadraters estimering ved kjente koblinger	25
4.2 Minste kvadraters estimering ved ukjente koblinger	29
4.3 Metropolis-Hasting simulering ved ukjent assosiasjon	39
5 Resultater	49
5.1 Simuleringsmodell	49
5.2 Sammenligning av MKK og LMK	58
5.3 Undersøkelse av $J^{GNN}(\phi)$ og $J^{NN}(\phi)$	60
5.4 Sammenligning av NN og GNN	64
5.5 Resultater fra MH-simulering	66

5.6	Simuleringer basert på realistiske data	70
5.7	Effekt av enkeltparametere	76
6	Oppsummering	81
7	Anbefalinger	83
8	Referanser	84
A	Spesifikasjoner for maskinvare	86
B	En auksjonsalgoritme for å finne GNN-assosiasjon	86
B.1	Tilpasning av algoritmen om ikke alle personer må tilordnes	89
B.2	Implementering av GNN assosiasjonen fra kapittel 4.2.3	89
C	Deteksjonsansynligheter for alle situasjoner	90
D	Taylorutvikling av $r(\nu; \phi)$	92
E	Korrelasjonskoeffisienter for situasjonene	94

Figurer

1	Illustrasjon av transformert koordinatsystem	5
2	Eksempler på rotasjoner. $\psi = (\psi_x, \psi_y, \psi_z)$	6
3	Eksempelbilde for å finne punkt inni rektangel	8
4	Eksempeldata for NN og GNN	10
5	NN mot GNN assosiasjon	10
6	Illustrasjon av LH i forhold til SJ. Merk at z_{LH} er tegnet motsatt vei. Figuren er basert på Andersen (2005).	18
7	Illustrasjon av SI. Figuren er basert på Andersen (2005).	19
8	Illustrasjon av aktuelle vinkler.	20
9	Illustrasjon av assosiasjonsproblemet.	26
10	Eksempel på hendelse som det ikke er tatt hensyn til ved beregning av sannsynligheten for type 2 feil i kapittel 4.2.4.	34
11	Feiltyper for NN.	34
12	Plott av l avhengig del av sannsynligheten for type 3 feil for gitt skip og radarpeiling fra bestemt, annet skip.	37
13	Plott av feilsannsynlighet som funksjon av vindustørrelsen i antall standardavvik, l	39
14	Oversikt over skipsposisjoner mellom Finnmark og Spitsbergen.	51
15	Automatisk utvelging av skip i satellittens åpningsområde.	52
16	Oversikt over de tre situasjonene.	54
17	Oversikt over skipsfordeling i situasjonene. Bilder eksportert fra Google Earth.	55
18	Empirisk deteksjonssannsynlighet etter passering.	56
19	Plott over projiserte hjørner av åpningsvinkelen, korresponderende område for å plassere skip og eksempel på skipsposisjoner.	57
20	Sammenligning av estimert forventningsskjevhet for MKK og LMK etter målestøy.	59
21	Sammenligning av estimert standardavvik for MKK og LMK etter målestøy.	60
22	Plott over oppførselen til J^{GNN} og J^{NN}	62
23	Empirisk bestemte standardavvik for NN og GNN.	64
24	Empirisk bestemte forventningsskjevheter for NN og GNN.	65
25	Andel riktige assosiasjoner for NN og GNN, målt i prosent.	66
26	Utvikling ved start i fasit	67
27	Utvikling ved start annet sted enn fasit	68
28	Plott over logaritmen til aposteriorifordelingen opp til en konstant.	69
29	Empirisk bestemte standardavvik for GNN i situasjonene beskrevet i kapittel 5.1.1.	70
30	Boksplott for estimeringsfeilen for de tre komponentene.	72
31	Gjennomsnittlig andel beregnet minimum av feilmålet utgjør av maksimal verdi ($N_v L^2$) for situasjonene beskrevet i kapittel 5.1.1.	73

32	Empirisk bestemte forventningsskjevheter for GNN i situasjonene beskrevet i kapittel 5.1.1.	74
33	Empirisk bestemt treffprosent for GNN i situasjonene.	75
34	Klassifisering av feilassosiasjonene for GNN i situasjonene.	75
35	Oversikt over hvor mye av åpningsvinkelen som er dekket av vinduene rundt AIS-peilingene.	77
36	Empirisk bestemte standardavvik ved varierende σ_ϕ	78

Tabeller

1	Forventet antall type 1 feil og sannsynligheten (sanns.) for ingen type 1 feil ved $L = l\sigma_w$	36
2	Forventet antall type 2 feil og sannsynligheten (sanns.) for ingen type 2 feil ved forskjellige l	37
3	Forventet antall type 3 feil og sannsynligheten (sanns.) for ingen type 3 feil ved forskjellige l	38
4	Empirisk deteksjonssannsynlighet (P_{AIS}) og antall skip (N_S) for alle situasjonene vist i figur 16. Verdiene er beregnet som gjennomsnitt av verdiene fra passering 5 til 8.	53
5	Oversikt over $\partial a/\partial b$ innsatt $\phi = 0$ og $v = 0$	94
6	Oversikt over andrederiverte som ikke er null.	94
7	Korrelasjonskoeffisienter for situasjonene. De to koordinatene angir hvilken korrelasjonskoeffisient det gjelder. For eksempel betyr xy SR at $\text{Kov}(\hat{\phi}_x^{GNN}, \hat{\phi}_x^{GNN}) / ((\text{Std } \hat{\phi}_x^{GNN} \text{ Std } \hat{\phi}_y^{GNN}))$ hvor kovariansen og standardavvikene er beregnet ut fra resultatene fra SR-metoden.	94

Liste over symboler og forkortelser

0	Tallet null, eller en nullvektor av passende dimensjon.
ϵ	Støy på vinkelpeilingen w .
ν	Peiling. Brukes steder hvor peilingen kan være både AIS- og radarpeiling.
ϕ^f, ϕ	Fasitverdi for stillingen til SI i forhold til FSI, og et forslag på hva den kan være. $d^{FSI} = R(\phi)d^{SI}$.
Ψ	Stillingen til FSI i forhold til LH. $d^{LH} = R(\Psi)d^{FSI}$.
ψ	Stillingen til LH i forhold til SJ. $d^{SJ} = R(\Psi)d^{LH} + r_s$
σ_ϕ	Komponentvis standardavvik i stillingen til ϕ .
σ_w	Komponentvis standardavvik i støyen ϵ på radarpeilingene w .
v	Åpningsvinkelen til antennen. $v = (v_1, v_2)$.
A^f, A	Koblingene mellom AIS-peilingene i V og radarpeilingene i A , og assosiasjoner som er gjetninger på koblingene.
d^X, D^X	Retningsvektor uttrykt i gitt koordinatsystem X , og matrisen som samler alle slike retningsvektorer.
I	Identitesmatrisen av passende dimensjon.
L	Radius til vinduet som blir brukt til å avgrense aktuelle radarpeilinger for NN og GNN.
N_s	Antall skip innenfor åpningsvinkelen.
N_v	Totalt antall AIS-peilinger. Kolonner i V .
N_w	Totalt antall radarpeilinger. Kolonner i W .
P_{AIS}	Sannsynlighet for å motta AIS-melding (eller peiling) fra et bestemt skip.
P_{rad}	Sannsynligheten for å motta en radarpeiling fra et gitt skip.
$r(\nu, \phi)$	Rotasjonsfunksjon for peilinger. Ved å finne rotene retningen ν representerer med ϕ vil peilingen endres til $r(\nu, \phi)$.
$R(\psi)$	Rotasjonsmatrise som representerer rotasjonen beskrevet ved ψ .
r_s	Satellittens posisjon uttrykt i SI.
v, V	AIS-peiling, og matrisen sammensatt av alle AIS-peilingene.
w, W	Radarpeiling, og matrisen sammensatt av alle radarpeilingene.
AIS	Automatisk identifikasjonssystem.

- CR Cramér Rao - som i Cramér Rao nedre grense.
- FFI Forsvarets forskningsinstitutt.
- FS Fasitstart - metode for å finne minimum for GNN.
- FSI Forventet versjon av satellittens interne koordinatsystem.
- GNN Global nærmeste nabo metoden.
- LH Lokal horisont koordinatsystemet.
- LMK Linearisert minste kvadraters metode ved kjente koblinger.
- MH Metropolis-Hastings.
- MKK Minste kvadraters metode ved kjente koblinger.
- NN Nærmeste nabo metoden.
- SI Satellittens interne koordinatsystem.
- SJ Statisk jord koordinatsystemet.
- SR Smart rutenettsøk - metode for å finne minimum for GNN.

1 Innledning

Forsvarets forskningsinstitutt (FFI) har utviklet og siden 12. juni 2010 sørget for driften av en overvåkingssatellitt (AISSAT-1) basert på mottak av meldinger fra "Automatisk identifikasjonssystem" (AIS). Skip som bærer AIS-sendere sender med jevne mellomrom meldinger som inneholder identitet og posisjon, samt andre detaljer om skip og reise. For den neste versjonen av satellitten ønsker FFI å komplettere informasjonen som mottas fra AIS med annen, uavhengig, informasjon. En mulighet for å oppnå dette er peilinger av skipene basert på utstråling fra deres navigasjonsradarer. Slike peilinger må være svært nøyaktige, noe som krever meget god kjennskap til satellittens stilling i rommet. En nærliggende mulighet for å oppnå dette er å assosiere AIS-posisjoner med radarpeilingene fra de samme skipene, og denne oppgaven gransker dette nærmere. Det er flere vanskeligheter forbundet med å utføre disse assosiasjonene. Det er en usikkerhet i radarpeilingene og flere AIS-posisjoner kan ligge innenfor denne. Noen radarpeilinger eller AIS-posisjoner kan også ha store feil. Hvor godt assosiasjonen kan utføres vil variere med fordeling av skip og usikkerheter i mottatt informasjon. Denne oppgaven formulerer dette assosiasjonsproblemet matematisk og undersøker hvor godt det er mulig å løse det. Den prøver også å bruke assosiasjonene til å estimere satellittens stilling. Det vil undersøkes hvor høy nøyaktighet som kreves i peilinger og satellittens stilling for at høy andel korrekte assosiasjoner og estimering av stilling er mulig. Sammenhengen mellom nøyaktighet i peilinger og estimat vil bli undersøkt og sammenlignet med teoretisk optimale verdier når det er tilgjengelig.

Fremgangsmåten som er valgt i denne oppgaven er å lage en matematisk modell av et forenklet system og simulere alle relevante data. Forenklingene er gjort for å begrense problemet og for å kunne fokusere på de mest sentrale delene av det. Det er også mottatt en del data fra FFI som er brukt til å lage realistiske situasjoner. For å sammenligne AIS-posisjoner og radarpeilinger er det valgt å konvertere AIS-posisjonene til de peilingene de ville gitt opphav til uten støy. Dette gjør at forholdet mellom radarpeilingene og de beregnede AIS-peilingene kun er avhengig av satellittens stilling, noe som gjør at samme metoder kan anvendes i et stort utvalg av situasjoner uten store endringer. I behandlingen av problemet utnyttet dette ved å se bort fra atmosfæriske påvirkninger av radarpeilingene og utstrakt bruk av en flat jord. I tillegg er det valgt å se på et statisk system hvor hverken skip eller satellitt beveger på seg.

Løsningsmetoder for problemet blir undersøkt først ved å finne metoder for å estimere stillingen hvis koblingene mellom AIS-peilinger og radarpeilinger er kjent. Den beste metoden blir så satt sammen med assosiasjonsmetoder og tilpasset tilfellet med ukjente koblinger. I tillegg undersøkes en bayesisk metode basert på en fullstendig statistisk beskrivelse av problemet. Metodene blir sammenlignet ut fra prestasjoner på simulerte data, og prestasjonen til de beste blir også undersøkt

i mer realistiske simuleringer basert på data fra FFI. I en praktisk anvendelse er det viktig at algoritmene som metodene gir opphav til kan kjøres i satellitten, noe som gir begrensinger på kjøretid og minnekrav. Denne oppgaven fokuserer på hvor godt det er mulig å få til estimering og assosiasjon og tar dermed ikke hensyn til dette i arbeidet. Erfaringer fra arbeidet gir likevel noen indikasjoner på hva som kan være fruktbare fremgangsmåter, det vil bli beskrevet i avslutningen av oppgaven.

Første delen av oppgaven beskriver det matematiske grunnlaget for estimeringen, assosiasjonen og modelleringen som gjøres senere. Etter grunnlaget er klart beskrives den matematiske modelleringen, og bakgrunnen for den, før de utviklede estimeringsmetodene presenteres. Så beskrives og kommenteres resultater fra simuleringene, og konklusjoner trekkes på bakgrunn av disse. Oppgaven avsluttes med kommentarer om hvordan oppgaven best kan videreføres med tanke på anvendelse i reelle situasjoner.

2 Matematisk bakgrunn

I arbeidet med oppgaven er det blitt bruk for mange forskjellige matematiske metoder og resultater. De vil bli beskrevet i dette kapitlet slik at det i senere kapitler er lettere å fokusere på de aktuelle problemene.

2.1 Notasjon og definisjoner

Det brukes ingen notasjon for å skille mellom skalarer og vektorer, sammenheng og definisjoner vil vise hva som er aktuelt. Det brukes alltid stor bokstav for matriser. For en matrise M benevnes elementet i rad i og kolonne j med m_{ij} , og kolonne j med m_j . Hvis M er kvadratisk betyr $|M|$ determinanten til matrisen. Den transponerte av en matrise oppnås ved å bytte om på rader og kolonner. Altså er $n_{ji} = m_{ij}$ når $N = M^T$. Om $M^T = M$ kalles matrisen symmetrisk. En n -dimensjonal kvadratisk matrise M kalles positivt semidefinit om $x^T M x \geq 0$ for alle $x \in \mathbb{R}^n$ untatt $x = 0$, 0 representerer her null-vektoren av passende dimensjon. Hvis ulikheten er streng kalles matrisen positivt definit. Om M er symmetrisk og positivt definit, finnes det en unik øvre triangulær matrise L med positive diagonalelementer som oppfyller $M = L^T L$ (Trefethen og Bau, 1997, s. 174). Dette kalles Choleskyfaktoriseringen av matrisen M . Med normen $\|\cdot\|$ til en vektor menes den euklidske normen $\|x\| = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$.

Hatt over en variabel angir en estimator for den variabelen, for eksempel er $\hat{\theta}$ en estimator for θ . Forventningsverdien angis ved $E\hat{\theta}$, varians med $\text{Var}\hat{\theta} = E(\hat{\theta} - E\hat{\theta})^2$ og standardavvik med $\text{Std}\hat{\theta} = \sqrt{\text{Var}\hat{\theta}}$. Forventningsskjevheten til estimatoren skrives $\text{Bias}\hat{\theta} = E\hat{\theta} - \theta$, om $\text{Bias}\hat{\theta} = 0$ kalles estimatoren forventningsrett.

2.2 Derivasjon av matriser

Veldig mange av størrelsene som brukes i denne oppgaven er vektorer, og det vil i noen tilfeller være behov for å derivere med hensyn på alle komponentene av en vektor. Ved å innføre den følgende notasjonene vil dette kunne skrives mye enklere.

La x er en kolonnevektor med n komponenter og f være en funksjon som kan ta x som argument og gir vektor med m komponenter som resultat. Da defineres den

deriverte $\partial f(x)/\partial x^T$ av Duda et al. (2001) til å være

$$\frac{\partial f(x)}{\partial x^T} = \begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1} & \frac{\partial f_1(x)}{\partial x_2} & \cdots & \frac{\partial f_1(x)}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m(x)}{\partial x_1} & \frac{\partial f_m(x)}{\partial x_2} & \cdots & \frac{\partial f_m(x)}{\partial x_n} \end{bmatrix}. \quad (1)$$

Denne definisjonen gjør at Jacobimatrisen $J(x)$ brukt for transformasjoner innen flerdimensjonale integraler kan skrives $\partial f(x)/\partial x^T$. Andre ordens taylorutvikling av en reell funksjon av vektorargumenter,

$$f(x) \approx f(x_0) + \sum_i \frac{\partial f}{\partial x_i} \Big|_{x=x_0} (x_i - x_{i0}) + \frac{1}{2!} \sum_i \sum_j \frac{\partial^2 f}{\partial x_i \partial x_j} \Big|_{x=x_0} (x_i - x_{i0})(x_j - x_{j0}), \quad (2)$$

kan da skrives som

$$f(x) \approx f(x_0) + \frac{\partial f}{\partial x^T} \Big|_{x=x_0} (x - x_0) + \frac{1}{2!} (x - x_0)^T \frac{\partial^2 f}{\partial x \partial x^T} \Big|_{x=x_0} (x - x_0). \quad (4)$$

Når grensene til en sum ikke spesifiseres, som i $\sum_i \frac{\partial f}{\partial x_i} \Big|_{x=x_0} (x_i - x_{i0})$, skal det summeres over alle verdier av i som gir mening. I det nevnte eksempelet skal summen løpe over alle komponentene til x , altså fra $i = 1$ til $i = n$ om x er en vektor med n komponenter. Når M er en konstant matrise vil

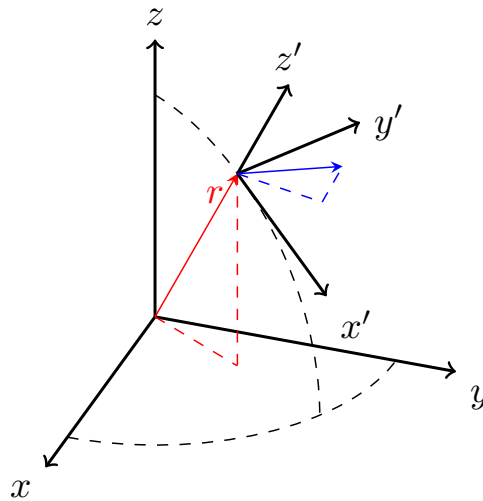
$$\frac{\partial}{\partial x^T} (Mx) = M, \quad (5)$$

$$\frac{\partial}{\partial x^T} (x^T Mx) = (M + M^T)x. \quad (6)$$

Dette kan verifiseres ved å skrive ut komponentene.

2.3 Geometri

Problemet denne oppgaven behandler har mange geometriske aspekter som vil bli behandlet i dette kapittelet. Det viktigste er hvordan rotasjon i tre dimensjoner beskrives, siden resultatene her brukes gjennom hele oppgaven. Dette er også det vanskeligste temaet, siden det er utfordrende å se for seg. Oppgaven vil være vesentlig lettere å følge med en god forståelse av hva rotasjonene representerer, og det er dermed vel verdt å legge inn en ekstra innsats her. Resten av temaene som diskuteres er enkle å se for seg slik at overfladisk lesning er tilstrekkelig for å følge bruken av dem.



Figur 1: *Illustrasjon av transformert koordinatsystem*

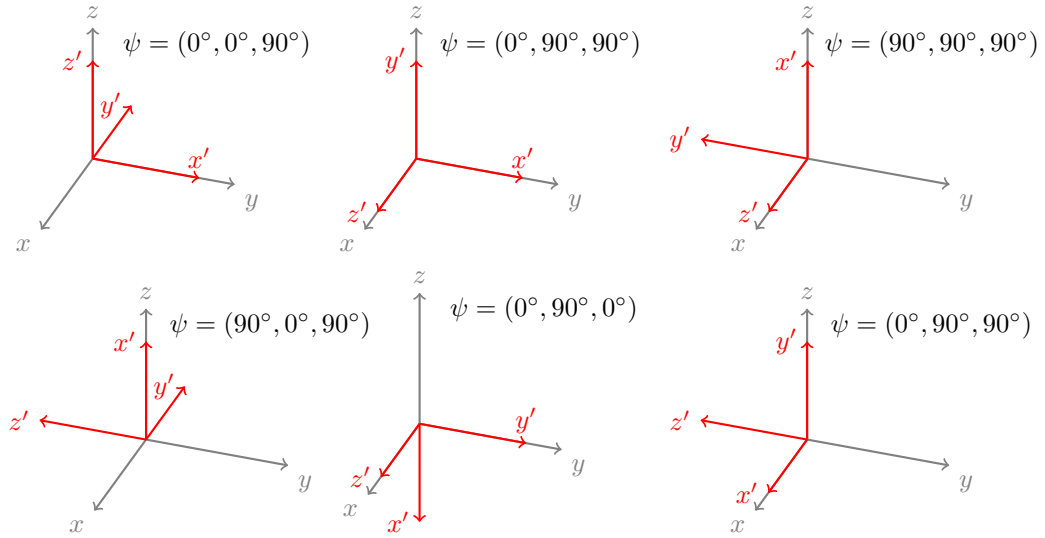
2.3.1 Rotasjon av koordinatsystemer i tre dimensjoner

Figur 1 illustrerer hvordan koordinatsystemer kan være plassert i forhold til hverandre. Fullstendig beskrivelse må inneholde både forflytning og relativ orientering, og det er nødvendig å kunne begge for å forstå metodene og modelleringen som brukes senere. At forflytning uten rotasjon kan utføres ved å legge til eller trekke fra vektoren mellom de to origoene regnes som kjent, derfor vil kun rotasjon behandles i dette kapitlet.

Relativ orientering mellom koordinatsystemer kan beskrives og uttrykkes på en mengde måter, felles for dem er at det kreves minst tre tall. I denne oppgaven er det valgt å bruke rotasjoner om tre forskjellige akser og å uttrykke dem ved rotasjonsmatriser. En rotasjon er fullstendig beskrevet ved tre vinkler ψ_x , ψ_y og ψ_z og utføres ved først å rotere med vinkelen ψ_z om z -aksen, så ψ_y om y -aksen og til slutt ψ_x om x -aksen. Alle rotasjoner skjer mot klokken etter vanlig høyrehåndsregel. Figur 2 viser en del enkle eksempelrotasjoner som kan brukes til å teste forståelsen. De merkede aksene er resultatet etter å ha utført rotasjonen beskrevet ved $\psi = (\psi_x, \psi_y, \psi_z)$ på det opprinnelige umerkede koordinatsystemet. Det er viktig å merke seg at rekkefølgen på rotasjonene har betydning og at det hele tiden roteres om de opprinnelige aksene, ikke de nye.

For å utlede rotasjonsmatriser for dette er det enklest å gå veien om todimensjonal rotasjon. I planet kan rotasjon med en vinkel θ mot klokken beskrives med matrisen

$$\begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}. \quad (7)$$



Figur 2: Eksempler på rotasjoner. $\psi = (\psi_x, \psi_y, \psi_z)$

Rotasjon mot klokken om en av koordinataksene i tre dimensjoner oppnås ved å holde koordinaten til rotasjonsaksen fiksert, og anvende plan rotasjon på de to andre. For rotasjon om henholdsvis x , y og z aksen gir dette matrisene

$$R_x(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{bmatrix}, \quad (8a)$$

$$R_y(\theta) = \begin{bmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{bmatrix} \text{ og} \quad (8b)$$

$$R_z(\theta) = \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (8c)$$

Ved å kombinere tre slike rotasjoner i en bestemt rekkefølge kan en beskrive enhver rotasjon i tre dimensjoner. Den totale rotasjonsmatrisen $R(\psi)$ kan regnes ut ved å gange sammen matrisene for de enkelte rotasjonene. Den nevnte rotasjonsrekkefølgen gir $R(\psi) = R_x(\psi_x)R_y(\psi_y)R_z(\psi_z)$, som utregnet blir

$$R(\psi) = \begin{bmatrix} \cos \psi_y \cos \psi_z & -\cos \psi_y \sin \psi_z & \sin \psi_y \\ \cos \psi_x \sin \psi_z & \cos \psi_x \cos \psi_z & -\cos \psi_y \sin \psi_x \\ +\sin \psi_y \cos \psi_z \sin \psi_x & -\sin \psi_y \sin \psi_z \sin \psi_x & \\ \sin \psi_z \sin \psi_x & \cos \psi_z \sin \psi_x & \cos \psi_x \cos \psi_y \\ -\cos \psi_x \cos \psi_z \sin \psi_y & +\cos \psi_x \sin \psi_z \sin \psi_y & \end{bmatrix}. \quad (9)$$

Fra dette kan det utledes at om en har et koordinatsystem S som kan transformeres til koordinatsystemet S' ved rotasjonen beskrevet i ψ , vil en vektor $v_{S'}$ uttrykt i S' kunne transformeres til S ved $v_S = R(\psi)v_{S'}$. Dette betyr også at transformasjon andre veien blir $v_{S'} = R^{-1}(\psi)v_S$, som også er lik $R^T(\psi)v_S$ siden $R^{-1} = R^T$ for alle rotasjonsmatriser. Det er lett å se at dette holder for R_x , R_y og R_z , siden inversen til en rotasjon må være å rotere motsatt vei og $\sin -\theta = -\sin \theta$. Fra dette fås det generelle resultatet ved å bruke at $R(\psi) = R_x(\psi_x)R_y(\psi_y)R_z(\psi_z)$ og reglene for invers, determinant og transponert av produkter.

2.3.2 Skjæringspunkt mellom linje og kule

En kule med sentrum i s og radius r består av de punktene x som oppfyller $(x - s)^T(x - s) = r^2$. En linje som går gjennom punktet p og er parallell med vektoren v består av punktene $x = p + tv$ der $t \in \mathbb{R}$. Eventuelle skjæringspunkt finnes der begge likningene er oppfylt. Altså må det finnes t slik at

$$r^2 = (x - s)^T(x - s) = (p + tv - s)^T(p + tv - s), \quad (10)$$

$$0 = (p - s)^T(p - s) - r^2 + 2tv^T(p - s) + t^2v^Tv. \quad (11)$$

Dette er en andregradslikning i t hvor løsningene er gitt ved

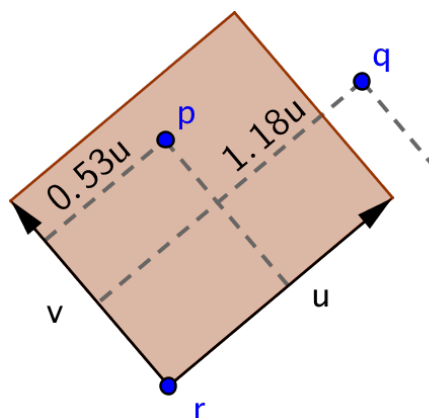
$$t_{\pm} = \frac{-2v^T(p - s) \pm \sqrt{(2v^T(p - s))^2 - 4v^Tv((p - s)^T(p - s) - r^2)}}{2v^Tv}, \quad (12)$$

Skjæringspunktene er gitt ved $x_{\pm} = p + t_{\pm}v$ så lenge t_{\pm} er reell, hvis t_{\pm} er kompleks finnes ingen skjæringspunkt. Dette vil bli brukt til å finne hvilket punkt på jordoverflaten en retningen v fra satellitten i punktet p tilsvarer, og det er derfor viktig å vite hvilket av de to punktene som er nærmest. Hvis en antar at p er utenfor kulen må t_+ og t_- ha samme fortegn og den nærmeste vil være den med lavest absoluttverdi. Siden roten alltid er positiv avgjøres dette av fortegnet til b . Resultatet er at x_+ må være nærmest p om b er positiv, og x_- er nærmest om b er negativ.

2.3.3 Skjæringspunkt mellom linje og plan

La linjen være beskrevet ved $x = p + tv$ som over. Et plan som inneholder punktet q med normalvektor n består av de punktene som oppfyller $(x - q)^Tn = 0$. Innsetting av likningen for den rette linjen gir $(tv + p - q)^Tn = 0$ som ved løsning med hensyn på t gir

$$t = -\frac{(p - q)^Tn}{v^Tn}, \quad (13)$$



Figur 3: Eksempelbilde for å finne punkt inni rektangel

når $v^T n \neq 0$. Det vil si når linjen ikke er parallell med planet. Altså er et eventuelt skjæringspunkt x gitt ved

$$x = p - \frac{(p - q)^T n}{v^T n} v. \quad (14)$$

2.3.4 Hvordan avgjøre at et punkt er inni et rektangel?

La punktet r være et av hjørnene i rektangelet og la u, v være vektorer som er like de to kantene i rektangelet som er koblet til r . Spørsmålet er da om et punkt p ligger innenfor rektangelet, gitt at p ligger i det samme planet som rektangelet.

Indreproduktet kan uttrykkes som $x^T y = \|x\| \|y\| \cos \theta$, hvor θ er vinkelen mellom de to vektorene x og y . I tillegg er $\|y\| \cos \theta$ lengden av komponenten til y i x -retning. Dette gir at

$$\frac{x^T y}{\|x\|^2} = \frac{\|y\| \cos \theta}{\|x\|} \quad (15)$$

er lengden av komponenten til y i x retning målt i antall x . Ved å se på figur 3 er det lett å innse at for at et punkt p skal ligge inni rektangelet, må ikke komponenten til $p - r$ langs kantene være større enn kantene, eller mindre enn null. Dette kan oversettes til at

$$\frac{x^T (p - r)}{\|x\|^2} \in [0, 1], \quad x = u, v. \quad (16)$$

2.4 Minste kvadraters metode

I problemer hvor det ønskes å finne det elementet som “passer best” med måledata er minste kvadraters metode (MKM) en mulighet. Det mest kjente eksempelet

er lineær regresjon hvor det søkes etter den rette linjen som passer best til punktene (x_i, y_i) , $i = 1 \dots n$. Med å passe best menes det å minimere den vertikale kvadratavstanden mellom modellen $\hat{y} = ax + b$ og punktene. Dette tilsvarer å finne verdiene for a og b slik at

$$J(a, b) = \sum_i (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - ax_i - b)^2 \quad (17)$$

er minimal. (Walpole et al., 2007) Ved å se nærmere på uttrykket for J kan det finnes idéer som også kan brukes i andre tilfeller. Leddet $(y_i - \hat{y}_i)^2$ er en form for modellfeil siden det er avhengig av differansen mellom målt verdi og verdien modellen \hat{y} forutsier. Hvis differansen mellom forutsagt verdi og målt verdi vokser, vil også størrelsen på feilen vokse. Dette gjør at hvis $J(a_1, b_1) < J(a_2, b_2)$ så har en grunn til å tro at linjen $y = a_1x + b_1$ passer bedre med dataene enn linjen $y = a_2x + b_2$. Denne grunnleggende idéen vil nå bli anvendt for mer generelle problemer.

Anta at du har et sett med måledata (x_i, y_i) hvor sammenhengen mellom x og y avhenger av en parameter θ . Sammenhengen mellom x og y modelleres som $y = f(x; \theta)$, og det ønskes å finne θ som passer best med dataene. En måte å finne den beste er da å velge den¹ verdien som minimerer

$$J(\theta) = \sum_i \|y_i - f(x_i; \theta)\|^2. \quad (18)$$

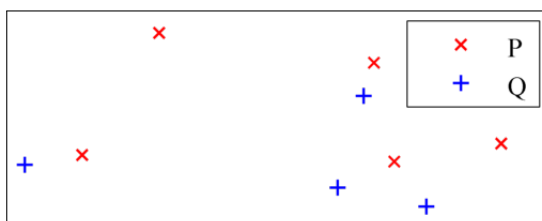
En enda mer generell metode er å konstruere et feilmål $J(\theta)$ basert på måledataene og igjen velge den beste θ som den som minimerer J . For at en skal kunne forvente gode resultater fra disse metodene må J eller f velges slik at det som regnes som en god θ vil ventes å gi lavere verdi for J eller $y_i - f(x_i; \theta)$. Et eksempel på dette er om y_i er lik $f(x_i; \theta_0)$ i en ideell situasjon (uten støy) når θ_0 er den virkelige verdien for θ .

For å finne minimum kan man i noen tilfeller sette de deriverte lik null og løse for θ , men i de generelle tilfellene blir ofte dette meget komplisert eller umulig. I slike tilfeller vil numeriske fremgangsmåter være nødvendig. Det er viktig å være oppmerksom på at det er mulig å ha flere lokale maksimum i de generelle tilfellene, noe som kan gjøre søket etter minimum vanskeligere.

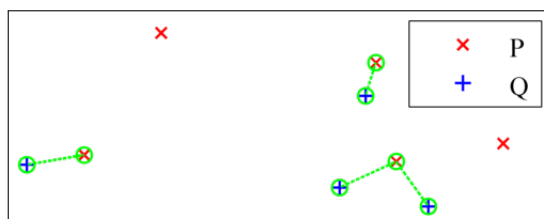
2.5 Nærmeste nabo- og global nærmeste nabo-metoden

Et sentralt problem i denne oppgaven er å assosiere to sett data med hverandre på en fornuftig måte. Nærmeste nabo-metoden (NN) og global nærmeste nabo-metoden (GNN) er to beslektede og intuitive måter å gjøre dette på. For å bli kjent med metodene er det nyttig å se på et eksempel.

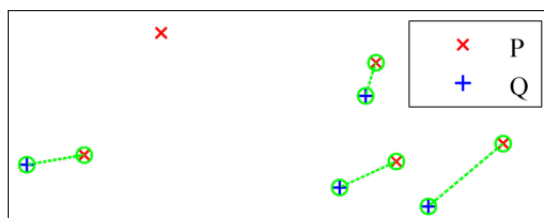
¹Minimum kan oppnås for flere θ ved (u)passende f . Ser bort fra det i denne beskrivelsen.



Figur 4: Eksempeldata for NN og GNN



(a) NN



(b) GNN

Figur 5: NN mot GNN assosiasjon

Gitt to punktmengder i planet P og Q hvor en ønsker å assosiere alle punktene i Q til et punkt i P . Et eksempel på hvordan dette kan se ut vist i figur 4. Som navnet tilsier løser NN problemet ved å assosiere hvert punkt q_i det punktet fra P som er nærmest. For eksempelet gir dette assosiasjonene vist i figur 5(a). GNN metoden gjør omtrent det samme, men krever at ingen punkt i P skal ha flere enn ett punkt fra Q assosiert med seg. Hvis to punkter fra Q “ønsker” seg samme punktet løses konflikten ved å assosiere de punktene med to fra P slik at summen av kvadratet av avstandene minimeres. Figur 5(b) viser resultatet fra GNN i eksempelet.

Begge disse metodene kan formuleres matematisk ved hjelp av minste kvadrater fremgangsmåten i kapittel 2.4. Nummerer punktene i P og Q , og la assosiasjonene være beskrevet ved vektoren A ved at $a_i = j$ om punkt i fra Q er assosiert med punkt j fra P . Det ønskes da å finne en A som er slik at $J(A) = \sum_i \|q_i - p_{a_i}\|^2$ minimeres. For GNN gjelder samme summen men det kreves at ingen elementer i A skal være like. Å finne assosiasjonene definert av NN er lett siden det resultatet for et element fra Q er uavhengig av alle andre elementer. Assosiasjonene kan derfor finnes ved å la $a_i = \arg \min_j \|q_i - p_j\|$. Denne uavhengigheten holder ikke

for GNN, som derfor krever det en mer komplisert optimeringsalgoritme. I denne oppgaven er det brukt en såkalt auksjonsalgoritme hentet fra Castanon (1992), som er basert på en alternativ matematisk formulering av problemet. Algoritmen og den nødvendige formuleringen er beskrevet nærmere i vedlegg B.

2.6 Transformasjon av sannsynlighetstetthetsfunksjoner

En stokastisk vektor X har sannsynlighetstettheten $f_X(x)$ om det for alle integrerbare mengder $D \in \mathbb{R}^n$ er slik at

$$P(X \in D) = \int_D f_X(x) dx, \quad (19)$$

der dx er et n -dimensjonalt volumelement. Sannsynlighetstetthetsfunksjonen f_X kalles også fordelingen til X . Hvis $x = g(u)$, og g er en-til-en på D kan integralet skrives om ved jacobideterminanten $\left| \frac{\partial x}{\partial u^T} \right|$

$$\int_D f_X(x) dx = \int_S f_X(x(u)) \left| \frac{\partial x}{\partial u^T} \right| du = \int_S f_X(x(u)) \left| \frac{\partial g(u)}{\partial u^T} \right| du, \quad (20)$$

når S er slik at $g(S) = \{g(u) | u \in S\} = D$. (Adams, 2006, kapittel 14.6). Dette betyr at hvis $U = g^{-1}(X)$ så er

$$P(X \in S) = P(U \in S) = \int_S f_X(x(u)) \left| \frac{\partial x}{\partial u^T} \right| du, \quad (21)$$

og fordelingen til U må være

$$f_U(u) = f_X(x(u)) \left| \frac{\partial x}{\partial u^T} \right|. \quad (22)$$

I dette kapittelet og neste skrives navnet på den tilfeldige variabelen med senket skrift for å vise hvilken fordeling som brukes. Ellers i oppgaven sløyfes den senkede skriften og argumentet vil vise hvilken fordeling som brukes, det vil si at $f_X(x)$ skrives $f(x)$. Denne forskjellen skyldes spesielle behov i disse to kapitlene, som ikke er aktuelle i resten av oppgaven.

2.7 Normalfordelingen for vektorer

En stokastisk vektor X er normalfordelt om fordelingen har formen

$$f_X(x) = \frac{1}{\sqrt{2\pi^d |\Sigma|^{1/2}}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right), \quad (23)$$

og Σ er positivt definit. (Duda et al., 2001, s.33). Her er $\mu = E(X)$ forventningsvektoren og $\Sigma = \text{Kov}(X) = E(X - \mu)(X - \mu)^T$ er kovariansmatrisen.

Det vil være av interesse å finne sannsynligheten for å havne nær forventningsvektoren. I en dimensjon formuleres dette som $|x - \mu| < r$. Ved å sammenligne avstanden med standardavviket er det mulighet å få et generelt uttrykk. La $r = l\sigma$ og $z = (x - \mu)/\sigma$, da er

$$P(|x - \mu| < r) = P(|z| < l) = \int_0^l \sqrt{\frac{2}{\pi}} \exp\left(-\frac{z^2}{2}\right). \quad (24)$$

I Duda et al. (2001, s.36) generaliseres dette til høyere dimensjoner ved å bruke Mahalanobis-avstanden

$$r^2 = (x - \mu)^T \Sigma^{-1} (x - \mu). \quad (25)$$

Området som begrenses av $r < \text{konstant}$ gir ellipser i 2 dimensjoner og ellipsoider i høyere dimensjoner. Ved å se på Cholesky-faktoriseringen L av Σ , og definere den tilfeldige vektoren $Z = L(X - \mu)$ kan uttrykket for avstanden forenkles til

$$r^2 = (L(x - \mu))^T (L(x - \mu)) = z^T z. \quad (26)$$

Ved egenskapene til determinanten følger det at $|\Sigma|^{1/2} = |\Sigma^{-1}|^{-1/2} = (|L|^2)^{-1/2} = |L|^{-1}$, og likning (5) gir at $\frac{\partial x}{\partial z^T} = L^{-1}$. Da må fordelingen til Z i følge likning (22) være

$$f_Z(z) = f_X(L^{-1}z + \mu) \left| \frac{\partial x}{\partial z^T} \right| = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^T z\right). \quad (27)$$

Sannsynligheten for å være innenfor l standardavvik $P(r < l)$ kan da uttrykkes som

$$P(r < l) = P(\|Z\| < l) = \frac{1}{\sqrt{2\pi}^d} \int_0^l \exp\left(-\frac{r^2}{2}\right) v(r) dr, \quad (28)$$

hvor $v(r)$ er overflaten på en d -dimensjonal kule med radius r . I to dimensjoner er dette omkretsen av en sirkel, $v(r) = 2\pi r$. Da er

$$P(R < l) = \frac{1}{2\pi} \int_0^l 2\pi r \exp\left(-\frac{r^2}{2}\right) dr = \left[-\exp\left(-\frac{r^2}{2}\right) \right]_0^l = 1 - e^{-l^2/2}. \quad (29)$$

Hvis R er en tilfeldig variabel definert ved $R = \sqrt{(x - \mu)^T \Sigma^{-1} (x - \mu)}$, må likning (29) gi den kumulative fordelingen for R . Siden den kumulative fordelingen er deriverbar kan fordelingen til Mahalanobis avstanden finnes ved å derivere den. Det gir fordelingen $f_R(r) = r \exp(-r^2/2)$. Det er viktig å merke seg at fordelingen til R er uavhengig av kovariansen og forventningsverdien til X .

2.8 Cramér Rao nedre grense ved normalfordelt støy

Cramér Rao (CR) nedre grense er en grense for hvor god en forventningsrett estimator kan være, med hensyn på kovarians. Om den kan beregnes er det nyttig å sammenligne estimatorer med denne grensen for å undersøke om det er mulig å lage noe som er bedre. Det tilfellet som blir brukt senere er estimat av en parameter $\theta = (\theta_1, \dots, \theta_m)^T$ som påvirker målinger som inneholder additiv, normalfordelt støy. Hvis $z_1, z_2 \dots$ er målingene kan dette skrives som:

$$Z = h(\theta) + W, \quad Z = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}, \quad h(\theta) = \begin{bmatrix} h_1(\theta) \\ \vdots \\ h_n(\theta) \end{bmatrix}, \quad W = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}. \quad (30)$$

Her er Z målingene, $h(\theta)$ er en kjent funksjon og W er den normalfordelt støyen. Det antas at $EW = 0$ og $\text{Kov } W = K$. Hvis $\hat{\theta}$ er en forventningsrett estimator for θ og de deriverte av h er gitt ved $J = \partial h(\theta)/\partial \theta$, garanterer CR nedre grense at

$$\text{Kov } \hat{\theta} - (J^T K^{-1} J)^{-1} \quad (31)$$

er positiv semidefinit. Bevis for dette kan finnes i Lehmann og Casella (1998, kap 6). Likning (31) innebærer at alle diagonalelementene i $\text{Kov } \hat{\theta}$ er større eller lik korresponderende diagonalelement i $(J^T K^{-1} J)^{-1}$.

Ved behov er den deriverte estimert numerisk ved at

$$\left. \frac{\partial f}{\partial x} \right|_{x=x_0} \approx \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}, \quad (32)$$

når Δx er liten nok.

2.9 Bayesestimering og Metropolis-Hastings algoritmen

I tillegg til MKM forsøkes det også å estimere interessante parametere ved såkalt bayesestimering. Dette er fundamentalt forskjellige fremgangsmåter siden parameteren som skal estimeres sees på som en stokastisk variabel med en fordeling istedet for et tall. Resultatet av estimeringen blir derfor en fordeling, kalt aposteriorifordelingen, for parameteren i lys av målingen istedet for et tall-estimat. En fordel med bayesestimering er at det er mulig å dra nytte av a priori kunnskap om parameterverdiene, og at aposteriorifordelingen i prinsippet gir tilgang til all informasjon om parameteren som estimeres. Matematisk betyr det at ved bayesestimeringen finner aposteriorifordelingen $f(\theta|x)$ til θ gitt målingen x . Ved å beskrive eksisterende kunnskap om θ ved apriorifordelingen $f(\theta)$ og sannsynligheten $f(x|\theta)$ for observere målingen x gitt verdien av θ sier Bayes regel at

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{\int f(x|\theta)f(\theta) d\theta} = \frac{f(x|\theta)f(\theta)}{f(x)}. \quad (33)$$

I mange situasjoner er nevneren $f(x)$ vanskelig å beregne, så i praksis er det kun kjent at $f(\theta|x)$ er proporsjonal med $f(x|\theta)p(\theta)$. Det eksisterer imidlertid algoritmer som kan beregne $f(\theta|x)$ bare på grunn av dette, for eksempel Metropolis-Hastings (MH) algoritmene. MH er basert på såkalte markovkjeder, et tema som er for omfattende til å dekke grundig her. En grunnleggende innføring finnes i Ross (2010), den følgende beskrivelsen er basert på fremstillingen i Gamerman og Lopes (2006).

2.9.1 Markovkjeder og MH

En markovkjede er en serie med tilfeldige variable som er egnet til å beskrive prosesser hvor den fremtidige tilstanden kun er avhengig av hva den er nå og ikke hva den har vært. For den presise definisjonen vil det brukes diskrete markovkjeder hvor de mulige tilstandene er tellbare og nummerert ved $\mathbb{N} = \{1, 2, \dots\}$. Mer presist formulert kan en da si at en markovkjede er en serie med tilfeldige variable $\theta^{(0)}, \theta^{(1)}, \dots$ hvor det er en fiksert sannsynlighet $p(\theta, \tilde{\theta})$ for å gå fra tilstand $\theta \in \mathbb{N}$ til tilstand $\tilde{\theta} \in \mathbb{N}$, og den fordelingen til den neste tilstanden i rekken kun avhenger av de foregående tilstandene gjennom den rett før, det vil si at

$$P(\theta^{(n+1)} = \tilde{\theta} | \theta^{(n)} = \theta, \theta^{(n-1)} = i_{n-1}, \dots, \theta^{(0)} = i_0) \quad (34)$$

$$= P(\theta^{(n+1)} = j | \theta^{(n)} = i_n) = p(\theta, \tilde{\theta}). \quad (35)$$

Funksjonen $p(\theta, \tilde{\theta})$ kalles overgangskjernen, og $p(\theta, \cdot)$ er en sannsynlighetsfordeling for enhver tilstand θ . Hvis overgangskjernen oppfyller bestemte betingelser finnes det en såkalt grensefordeling. Det vil si at fordelingen til $\theta^{(n)}$ konvergerer til den samme fordelingen når $n \rightarrow \infty$, uansett hva startverdien $\theta^{(0)}$ er. MH baserer seg på å konstruere en markovkjede hvor grensefordelingen er lik den fordelingen som det ønskes å finne ut mer om. Ved å simulere kjeden på en datamaskin og la den gå lenge nok vil det være mulig å finne ut alt som ønskes om grensefordelingen.

La $f(\theta)$ være fordelingen det ønskes å trekkes fra, det kan for eksempel være $f(\theta|x)$ beskrevet over. Hvis en kan finne en overgangskjerne $p(\theta, \tilde{\theta})$ som oppfyller

$$f(\theta)p(\theta, \tilde{\theta}) = f(\tilde{\theta})p(\tilde{\theta}, \theta), \quad (36)$$

kalles kjeden $p(\theta, \tilde{\theta})$ gir opphav til reversibel, og det finnes teoremer som garanterer at grensefordelingen finnes og er lik $f(\theta)$. I MH settes denne overgangskjernen $p(\theta, \tilde{\theta})$ sammen av to elementer: En vilkårlig overgangskjerne $q(\theta, \tilde{\theta})$ og en akseptanssannsynlighet $\alpha(\theta, \tilde{\theta})$ slik at

$$p(\theta, \tilde{\theta}) = q(\theta, \tilde{\theta})\alpha(\theta, \tilde{\theta}), \quad \tilde{\theta} \neq \theta. \quad (37)$$

Akseptanssynligheten α er definert til

$$\alpha(\theta, \tilde{\theta}) = \min \left(1, \frac{f(\theta)q(\tilde{\theta}, \theta)}{f(\theta)q(\theta, \tilde{\theta})} \right). \quad (38)$$

For at $p(\theta, \cdot)$ skal være en sannsynlighetsfordeling må total summen bli 1, som betyr at

$$p(\theta, \tilde{\theta}) = 1 - \sum_{\tilde{\theta} \neq \theta} q(\theta, \tilde{\theta})\alpha(\theta, \tilde{\theta}). \quad (39)$$

Ved innsetting og undersøkelse av mulige α -verdier følger det at likning (36) er oppfylt. I praksis realiseres denne kjeden med denne overgangskjernen på følgende vis:

1. Start med en vilkårlig initialverdi $\theta^{(0)}$. Sett rundetelleren $j = 1$.
2. Trekk et forslag på ny verdi $\tilde{\theta}$ basert på fordelingen $q(\theta^{(j-1)}, \cdot)$.
3. Trekk $u \sim U(0, 1)$ og aksepter $\tilde{\theta}$ om $\alpha(\theta^{(j-1)}, \tilde{\theta}) \geq u$. Ved aksept sett $\theta^{(j)} = \tilde{\theta}$, ved forkastning sett $\theta^{(j)} = \theta^{(j-1)}$.
4. Øk rundetelleren fra j til $j + 1$. Gå til steg 2. Repeter til konvergens er oppnådd.

Beskrivelsen her er basert på at tilstandene $\theta^{(j)}$ er heltall, men algoritmen er fleksibel og fungerer også for vektorer med diskret, kontinuerlig eller blandet fordeling. Overgangen er heldigvis enkel, siden de naturlige endringene fungerer. For eksempel vil beskrivelsen over passe for kontinuerlige tilstander om overgangskjernen tolkes som sannsynlighetstettheten og summene erstattes med integraler. En beskrivelse av det generelle tilfellet krever en målteoretisk beskrivelse, noe som er for omfattende til å gi her.

2.9.2 Nyttig triks for overgangssannsynligheten

Det kan i mange tilfeller være betydelig enklere å bruke en overgangskjerne $q_\gamma(\tilde{\theta}, \theta)$ som er avhengig av en tilfeldig parameter γ , istedet for den fullstendige $q(\tilde{\theta}, \theta)$. Parameteren γ kan for eksempel beskrive rekkefølgen endringsforslagene til θ -komponentene beregnes.

For at algoritmen skal gi riktig grensefordeling må $f(\theta)q(\tilde{\theta}, \theta) = f(\tilde{\theta})q(\theta, \tilde{\theta})$. Det tilsvarende uttrykket ved $q_\gamma(\tilde{\theta}, \theta)$ er

$$f(\theta)q_\gamma(\tilde{\theta}, \theta) = f(\tilde{\theta})q_\gamma(\theta, \tilde{\theta}). \quad (40)$$

Hvis fordelingen til γ er $f(\gamma)$ vil simultanfordelingen til θ og $\tilde{\theta}$ være $f(\gamma)q_\gamma(\tilde{\theta}, \theta)$ som gir

$$q(\theta, \tilde{\theta}) = \sum_{\gamma} f(\gamma)q_\gamma(\theta, \tilde{\theta}). \quad (41)$$

Siden grensefordelingen $f(\theta)$ ikke påvirkes av γ kan den flyttes utenfor summen. Ved hjelp av likning (40) følger det at

$$f(\theta)q(\tilde{\theta}, \theta) = f(\theta) \sum f(\gamma)q_{\gamma}(\tilde{\theta}, \theta) \quad (42)$$

$$= f(\tilde{\theta}) \sum f(\gamma)q_{\gamma}(\theta, \tilde{\theta})d\gamma \quad (43)$$

$$= f(\tilde{\theta})q(\theta, \tilde{\theta}), \quad (44)$$

altså er likning (36) oppfylt om likning (40) holder.

3 Modelling

Som nevnt i innledningen brukes det forenklete matematiske modeller av systemet for å undersøke problemet. I dette kapitlet vil forenklingene som er gjort beskrives og resulterende matematiske modeller utledes. Kapitlet starter med beskrivelse av koordinatsystemene som vil brukes, sammenhengen mellom dem og hvordan konvertering mellom dem blir utført. Så blir vinklene brukt til peiling gjort rede for, og det utledes en sammenheng mellom AIS-peilingene og radarpeilingene som kun er avhengig av feil i satellittens stilling. Til slutt beskrives måten dette er implementert og noen ekstra forenklinger benyttes for å gjøre implementasjonen av modellen enklere.

3.1 Koordinatsystemer

Det blir brukt fire høyrehendte koordinatsystemer denne oppgaven. Ett koordinatsystem har origo i jordens sentrum, mens de tre andre har origo i satellittens sentrum. Valg av koordinatsystemer og beskrivelse av dem er inspirert av Andersen (2005). Den følgende listen introduserer koordinatsystemene, og de beskrives nærmere under.

SJ Statisk jord. SJ har origo i jordens sentrum, z -aksen mot nordpolen og x -aksen mot null-meridianen.

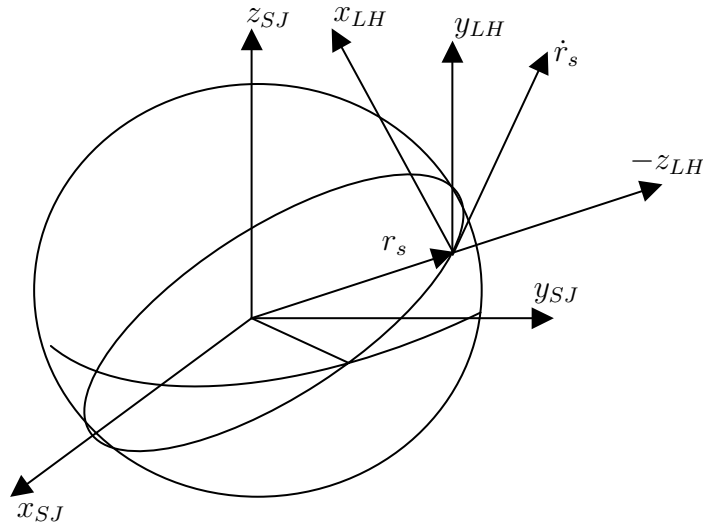
LH Lokal horisont. Dette koordinatsystemet er basert på satellittens posisjon og bevegelse. Origo er plassert i satellittens sentrum, z -akse er rettet mot jordens sentrum og y -aksen er (omtrent) i fartsretningen.

SI Satellittens interne koordinatsystem. SI har origo i fasesenteret til satellittens antenne og aksene er rettet mot bestemte punkter på satellitten. Radarpeilingene utføres fra dette systemet.

FSI Forventet versjon av satellittens interne koordinatsystem. Satellitten vil inneholde utstyr som skal holde orden på stillingen, dette er slik SI ville vært om utstyret fungerte perfekt. Se nærmere beskrivelse under.

3.1.1 Statisk jord (SJ) koordinatsystemet

SJ koordinatsystemet koordinatsystem er fiksert i forhold til jorden og brukes til å oppgi posisjoner for skip og satellitt. Origo er som nevnt i jordens sentrum og systemet har z_{SJ} -aksen pekende mot nordpolen, x_{SJ} -aksen mot skjæringen mellom null meridianen og ekvator og y_{SJ} -aksen slik at det blir et høyrehåndssystem. Det er noen ganger behov for å bruke posisjoner fra lengde/breddegrad, og de konverteres da til SJ under antakelsen at jorden er en kule med radius på 6371km.



Figur 6: Illustrasjon av LH i forhold til SJ. Merk at z_{LH} er tegnet motsatt vei. Figuren er basert på Andersen (2005).

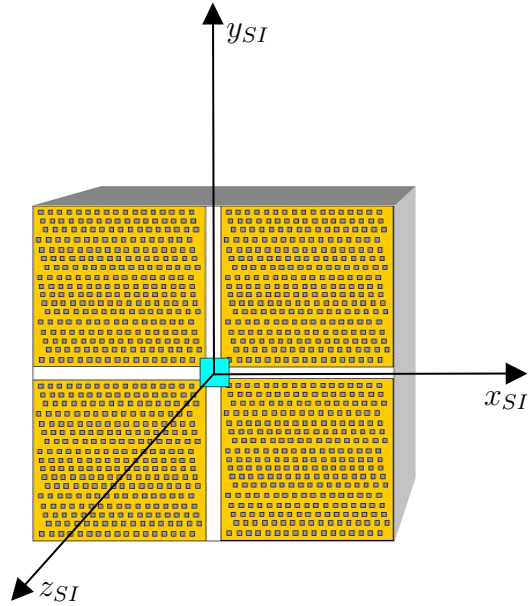
3.1.2 Lokal horisont (LH) koordinatsysteme

LH er definert ut fra satellittens posisjon og bevegelse. Det sier ingenting om satellittens orientering, men det er nyttig å definere orienteringen relativt til dette koordinatsystemet. Siden denne oppgaven ser på et statisk tilfelle må det velges en tenkt bevegelsesretning. Origo til systemet er i satellittens sentrum (r_s uttrykt i SJ), z_{LH} -aksen peker rett ned mot jordens sentrum. Om fartsretningen \dot{r}_s er vinkelrett på z_{LH} akse vil y_{LH} -aksen være i samme retning, men siden dette ikke alltid er tilfelle defineres de resterende aksene via kryssproduktet: x_{LH} -aksen peker i retning $\dot{r}_s \times z_{LH}$, og y_{LH} akse peker i retningen til $z_{LH} \times x_{LH}$. Figur 6 viser hvordan dette kan se ut.

Vinklene $\psi = (\psi_x, \psi_y, \psi_z)$ beskriver orienteringen til LH i forhold til SJ ved metodene fra kapittel 2.3.1. En vektor v^{LH} uttrykt i LH kan dermed uttrykkes i SJ ved å først rotere ($R(\psi)v^{LH}$) og så legge til r_s : $v^{SJ} = r_s + R(\psi)v^{LH}$. Transformasjon fra LH til SJ blir $v^{LH} = R^T(\psi)(v^{SJ} - r_s)$.

3.1.3 Satellittens interne koordinatsystem (SI) og forventet SI (FSI)

SI er fiksert i forhold til satellittkroppen, og beskriver hvordan satellitten er rettet. Systemet er illustrert i figur 7, z_{SI} -aksen peker i samme retning som antennen og de to andre aksene mot fikserte punkter på satellitten. Satellitten inneholder utstyr som måler denne stillingen, og rapporterer hvordan satellitten står i forhold til LH gjennom vinklene $\Psi = (\Psi_x, \Psi_y, \Psi_z)$. Hvis utstyret hadde fungert perfekt ville dette vært SI, men siden dette desverre ikke er tilfellet blir dette slik SI



Figur 7: Illustrasjon av SI. Figuren er basert på Andersen (2005).

forventes å være. Derfor får systemet som har stillingen Ψ i forhold til LH navnet FSI, det vil si at $v^{LH} = R(\Psi)v^{FSI}$. Den virkelige stillingen til SI beskriver i forhold til FSI ved vinklene $\phi = (\phi_x, \phi_y, \phi_z)$. Konvertering gjøres ved

$$v^{FSI} = R(\phi)v^{SI} \text{ og } v^{SI} = R^T(\phi)v^{FSI}. \quad (45)$$

Denne stillingen ϕ blir også kalt skjevstillingen siden den uttrykker feilen i målt stilling. Skjevstillingen ϕ regnes som ukjent og estimeres senere i oppgaven.

3.2 Peilinger

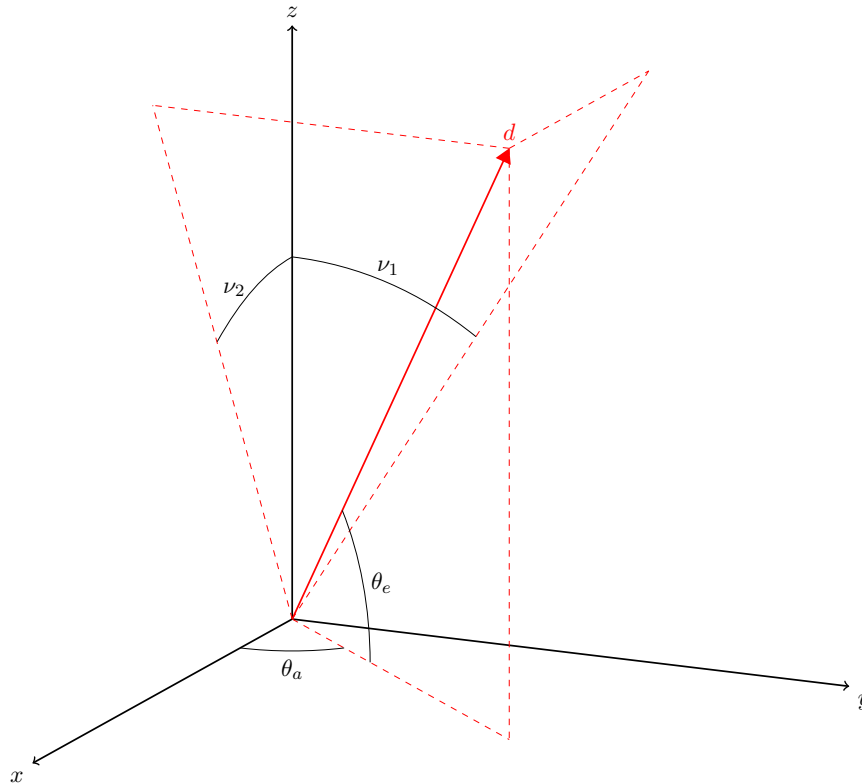
En peiling er en beskrivelse av ankomstretningen til et signal. Ankomstretningen har to frihetsgrader og vil bli beskrevet ved to vinkler. Hvordan disse to vinklene kan beregnes, hva de representerer og hvorfor de er valgt vil beskrives i dette kapittelet.

Om $d = (d_x, d_y, d_z)^T$ beskriver ankomstretningen er den korresponderende peilingen gitt ved

$$\nu = \begin{bmatrix} \arctan\left(\frac{d_y}{d_z}\right) \\ \arctan\left(\frac{d_x}{d_z}\right) \end{bmatrix}. \quad (46)$$

Peilevinklene kan tolkes grafisk som vinklene mellom z-aksen og projeksjonene av ankomstretningen i xz - og yz -planene, som vist i figur 8. Ved å rotere SI vist

i figur 7 slik at aksene faller sammen med koordinatsystemet vist i figur 8 er det enklere å se hva dette betyr for satellitten. Antennen på satellitten “ser” i z retning og d blir ankomstretningen til et signal. Peilingen ν blir en alternativ måte å representere ankomstvinkelen d på.



Figur 8: Illustrasjon av aktuelle vinkler. Vinklene ν_1 og ν_2 definert i likning (46) blir brukt videre, θ_a og θ_e er asimut- og elevasjonsvinklene som er forkastet. Figuren er basert på et eksempel laget av Hein (2009).

3.2.1 Om valg av vinkler

Det finnes flere måter å beskrive ankomstvinklene på, og dette kapitlet vil gi begrunnelsen for valget av vinklene beskrevet over. De viktigste estimatorene brukt i denne oppgaven er basert på minste kvadraters metode, som beskrevet i kapittel 2.4. Med denne fremgangsmåten er det viktig at peilinger som representerer retninger som er nær hverandre er like og retninger som er langt fra hverandre gir mest mulig ulike peilinger. Vinklene som er valgt oppfyller dette kravet så lenge peilevinklene ikke er nær eller over 90 grader, noe som ikke er aktuelt i denne oppgaven. Det mest aktuelle alternativet er asimut- og elevasjonsvinklene illustrert ved ϕ_a og ϕ_e i figur 8, og brukt av Andersen (2005). Disse vinklene ble forkastet

på grunn av at de feiler begge kravene nevnt over. Ved elevasjonsvinkel nær 90 grader er retningene meget like uansett asimutvinkel, og uansett elevasjonsvinkel er retninger med asimutvinkel 179 og -179 mye nærmere enn avstanden i verdi tilsier.

3.2.2 AIS- og radarpeilinger

Det som vil bli brukt som grunnlag for å estimere stillingen vil være radarpeilingene og såkalte AIS-peilinger. I det reelle tilfellet vil satellitten i løpet av en tidsperiode motta radarsignaler fra en mengde skip og AIS-meldinger fra en annen, delvis overlappende, mengde skip. Radarpeilingene vil bli beregnet på bakgrunn av radarsignalene, AIS-peilingene på grunnlag av posisjonene mottatt i AIS-meldingene og kunnskap om satellittens stilling. Idéen bak dette er at en AIS-peiling fra et skip ville vært identisk med radarpeilingen fra samme skip om stillingen til satellitten er kjent og det ikke er støy noe sted. Hvordan AIS- og radarpeilingene er modellert vil beskrives i dette kapittelet.

En radarpeiling skal angi ankomstretningen til radarsignalene. Siden beregningen av dem baseres på radarsignalene antennene mottar vil ankomstretningen måtte oppgis i forhold til satellittens orientering, altså i forhold til SI. Målingene antenne foretar er forbundet med feil, noe som gjør at radarpeilingene ikke vil være eksakte. Hvis d^{SI} angir den eksakte ankomstretningen til radarsignalet gir likning (46) at den tilhørende radarpeilingen w må bli

$$w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} \arctan \left(\frac{d_y^{SI}}{d_z^{SI}} \right) \\ \arctan \left(\frac{d_x^{SI}}{d_z^{SI}} \right) \end{bmatrix} + \epsilon. \quad (47)$$

Her representerer ϵ feilen i peilingen. Fordelingen til ϵ vil avhenge av antennene til satellitten og algoritmene som brukes til å estimere peilevinkelen ut fra dataene fra antennene. I den grad det er behov for en eksplisitt fordeling videre er den antatt å være en bivariat normalfordeling med uavhengige, likt fordelte, komponenter sentrert i 0. Det vil si at $E \epsilon = 0$ og $\text{Kov } \epsilon = \sigma_w^2 I$.

Som sagt skal en AIS-peilingen ideelt sett være identisk med radarpeilingen fra samme skip. Den beste muligheten for dette ville vært å beregne d^{SI} ut fra kunnskap om forplantningen av radarsignalene og satellittens posisjon og stilling. Men siden stillingen til SI er ukjent er dette umulig, og det nest beste er å bruke ankomstretningen i FSI, d^{FSI} . I AIS-meldingene mottas kun posisjonen, og for å kunne beregne d^{FSI} må det gjøres antakelser om bevegelsen til radarsignalene. I denne oppgaven antas det at radarsignalene beveger seg rettlinjet, slik at retningsvektoren fra satellitt til skip også angir ankomstretningen. Hvis u er posisjonen til et skip vil ankomstretningen i SJ være $d^{SJ} = u - r_s$. Ved konverteringsformlene

fra kapittel 3.1 blir da ankomstretningen i LH $d^{LH} = R^T(\psi)d^{SJ}$ og videre til FSI ved $d^{FSI} = R^T(\Psi)d^{LH}$. Fra ankomstretningen d^{FSI} i FSI beregnes AIS-peilingen v ved likning (46). Det gir

$$v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} \arctan\left(\frac{d_y^{FSI}}{d_x^{FSI}}\right) \\ \arctan\left(\frac{d_z^{FSI}}{d_x^{FSI}}\right) \end{bmatrix}. \quad (48)$$

Hvis utstyret på satellitten som følger stillingen ikke har feil vil SI og FSI falle sammen, og ved $\epsilon = 0$ følger det fra likningene (47) og (48) $v = w$ forutsatt at modelleringen av signalgangen er korrekt. Altså oppfyller denne definisjonen ønsket om at AIS- og radarpeilingene fra samme skip vil være identiske i en ideell situasjon.

Det er ikke korrekt at radarsignalene går i en rett linje, siden de vil avbøyes i atmosfæren. Denne forenklingen er gjort fordi en reell modellering ville vært betydelig mer komplisert å jobbe med og assosiasjonsproblemet er likt ved begge modelleringene. På å grunn av $d^{FSI} = R(\phi)d^{SI}$ er forholdet mellom AIS- og radarpeilingene ved korrekt signalgang kun avhengig av stillingen, den eksakte sammenhengen utledes i kapittel 3.2.3. Løsningsmetodene for assosiasjonsproblemet i denne oppgaven fokuserer kun på forholdet mellom AIS- og radarpeilingene, som er helt likt uansett signalgang.

Når det er snakk om flere radarpeilinger vil de indekseres med nummer. Det vil si at w_1, \dots, w_{N_w} betyr radarpeiling 1,2 og så videre opp til totalt antall radarpeilinger w_{N_w} . Alle radarpeilingene samles i en matrise $W = [w_1, \dots, w_{N_s}]$ med en radarpeiling i hver kolonne, og det brukes dobbelindeks for å beskrive komponentene i w_i , w_{1i} og w_{2i} . Med denne notasjonen kan det være vanskelig å avgjøre om w_1 angir den første radarpeilingen eller første komponenten til en enkel radarpeiling som i likning (47). Dette blir ikke noe problem siden det vil være klart fra kontekst hvilken tolkning som gjelder. Tilsvarende notasjon brukes for AIS-peilingene, v_i angir AIS-peiling nr i med komponenter v_{1i} og v_{2i} , og alle AIS-peilingene samles i V . Totalt antall AIS-peilinger angis med N_v .

3.2.3 Grunnlag for estimering

Siden AIS-peilingene baserer seg på d^{FSI} , radarpeilingene på d^{SI} og $d^{SI} = R^T(\phi)d^{FSI}$ bør det være mulig å finne en sammenheng mellom radar og AIS-peilinger fra samme skip som kun er avhengig av ϕ . Denne sammenhengen eksisterer og utledes i dette kapittelet. Grunnlaget for utledningen er sammenhengen mellom d^{SI} og d^{FSI} fra likning (45) og likningene (47) og (48) som kobler ankomstretningene med peilingene.

La w være en radarpeiling og v en AIS-peiling som stammer fra samme skip . Ved å løse likning (48) for komponentene i d^{FSI} følger det at

$$d_y^{FSI} = d_z^{FSI} \tan v_1 \text{ og } d_x^{FSI} = d_z^{FSI} \tan v_2. \quad (49)$$

Siden det kun er retningen som er av interesse kan d^{FSI} velges til å være en enhetsvektor. Det gir at retningsvektoren i FSI som korresponderer med AIS-peilingen v er

$$d^{FSI} = C \begin{bmatrix} \tan v_2 \\ \tan v_1 \\ 1 \end{bmatrix}, \quad C = \frac{1}{1 + \tan^2 v_2 + \tan^2 v_1}. \quad (50)$$

Denne retningsvektoren kan konverteres til SI ved likning (45). Sammen med likningene (9) og (50) gir den at retningsvektoren i SI blir

$$d^{SI} = C \begin{bmatrix} \cos \phi_y \cos \phi_z \tan v_2 \\ + (\cos \phi_x \sin \phi_z + \sin \phi_y \cos \phi_z \sin \phi_x) \tan v_1 \\ + \sin \phi_z \sin \phi_x - \cos \phi_x \cos \phi_z \sin \phi_y \\ - \cos \phi_y \sin \phi_z \tan v_2 \\ + (\cos \phi_x \cos \phi_z - \sin \phi_y \sin \phi_z \sin \phi_x) \tan v_1 \\ + \cos \phi_z \sin \phi_x + \cos \phi_x \sin \phi_z \sin \phi_y \\ \sin \phi_y \tan v_2 - \cos \phi_y \sin \phi_x \tan v_1 + \cos \phi_x \cos \phi_y \end{bmatrix}. \quad (51)$$

Siden w og v stammer fra samme skip vil d^{SI} være ankomstretningen som peilingen w tilsvare. Fra likning (47) følger det da at radarpeilingen w kan skrives som

$$w = \begin{bmatrix} \arctan \left(\frac{-\cos \phi_y \sin \phi_z \tan v_2 + (\cos \phi_x \cos \phi_z - \sin \phi_y \sin \phi_z \sin \phi_x) \tan v_1 + \cos \phi_z \sin \phi_x + \cos \phi_x \sin \phi_z \sin \phi_y}{\sin \phi_y \tan v_2} \right) \\ - \cos \phi_y \sin \phi_x \tan v_1 + \cos \phi_x \cos \phi_y \\ \arctan \left(\frac{\cos \phi_y \cos \phi_z \tan v_2 + (\cos \phi_x \sin \phi_z + \sin \phi_y \cos \phi_z \sin \phi_x) \tan v_1 + \sin \phi_z \sin \phi_x - \cos \phi_x \cos \phi_z \sin \phi_y}{\sin \phi_y \tan v_2} \right) \\ - \cos \phi_y \sin \phi_x \tan v_1 + \cos \phi_x \cos \phi_y \end{bmatrix} + \epsilon. \quad (52)$$

Denne transformasjonen vil være sentral gjennom hele oppgaven, og den ikke-stokastiske delen av den vil bli referert til som en funksjon r , det betyr at r er

slik at likning (52) kan skrives

$$w = r(v; \phi) + \epsilon. \quad (53)$$

Denne sammenhengen gjør at algoritmer som skal estimere ϕ kan baseres på kun AIS-peilingene V og radarpeilingene W .

Funksjonen $r(v; \phi)$ sies å rotere AIS-peilingen v med ϕ fordi $r(v, \phi)$ tilsvarer samme retningen som v om den tolkes i et koordinatsystem som har stillingen ϕ i forhold til systemet ν er målt i. For å få en følelse av hvordan rotasjonen påvirker peilingene er det nyttig å se på en forenklet versjon. For små verdier av x er $\arctan x \approx \tan x \approx \sin x \approx x$ og $\cos x \approx 1$. Nevneren i likning (52) kan da skrives som $1 +$ andre ordens ledd, som omtrent er 1 om både peile- og stillingsvinklene er små. Ved også å se bort fra ledd over andre orden i tellerene i likning (52) fås tilnærmingen

$$r(v; \phi) \approx \begin{bmatrix} v_1 + \phi_x - \phi_z v_2 + \phi_z \phi_y \\ v_2 - \phi_y + \phi_z v_1 + \phi_z \phi_x \end{bmatrix}. \quad (54)$$

Ved å se på v som et punkt i planet og kun ha en ϕ -komponent ulik null kan den grunnleggende effekten til r sees. Da tilsvarer $\phi_x \neq 0$ og $\phi_y \neq 0$ forflytninger langs de to koordinataksene, siden verdien legges til en v -komponent. Endring av ϕ_z tilsvarer rotasjon mot klokken, fordi likning (54) med $\phi_x = \phi_y = 0$ er det samme som tilsvarer v ganget med den en forenklete versjonen av rotasjonsmatrisen fra likning (7).

Når det finnes flere AIS-peilingen som er samlet i matrisen V angir $r(V; \phi)$ matrisen som oppnås ved å erstatte kolonnene v_i med $r(v_i; \phi)$, det vil si å bruke r på alle enkeltpeilingene i V . Sammenhengen i likning (54) kan også utledes mer presist ved å finne andre ordens taylorutvikling av $r(v, \phi)$ om $\phi = 0$ og $v = 0$, noe som er gjort i vedlegg D.

4 Estimeringsmetoder

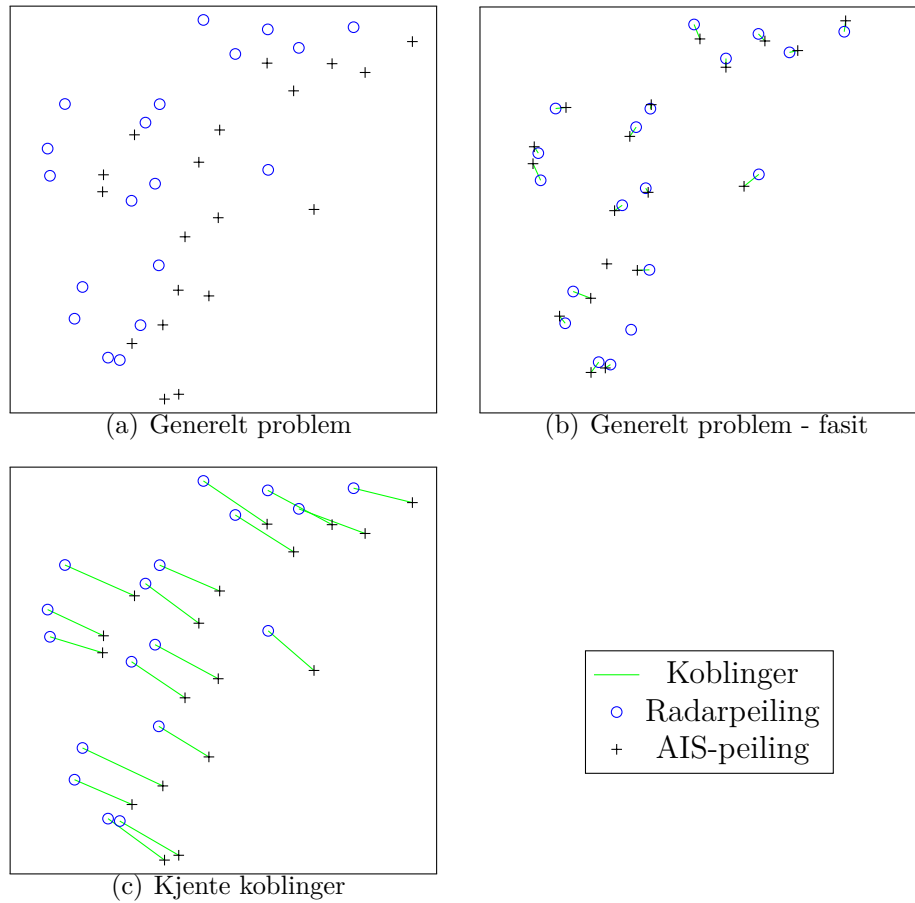
I dette kapitlet vil forskjellige metoder for å estimere skjevstillingen diskuteres. I den følgende diskusjonen vil det bli nødvendig å skille mellom fasit-verdien av stillingen, som vil skrives ϕ^f , og en mer eller mindre vilkårlig verdi for stillingen, ϕ . På grunn av sammenhengen r utledet i forrige kapittel baserer alle metodene seg kun på bakgrunn av peilingene V og W . Dette er fordi at assosiasjonsproblemet nå kan sees på som å få to punktmengder i planet V og W til å passe best mulig sammen ved å transformere V ved hjelp av r . Med å passe sammen menes å få punktene til å ligge nærmest mulig hverandre, samtidig som en tar høyde for at det finnes punkt uten “partner” i den andre mengden. Dette illustreres i figur 9(a) hvor et mulig eksempel av W og V er plottet. Ut fra tolkningen tilnærmingen av r i likning (54) blir problemet å finne en horisontal forflytning ϕ_x , en vertikal forflytning ϕ_y og en rotasjon ϕ_z som gjør at de svarte kryssene havner tettest mulig på de blå. På øyemål er det mulig å se at punktene bør flyttes oppover til venstre, og resultatet ved den korrekte transformasjon vises i figur 9(b). De grønne linjene viser hvilke punkt som er koblede, altså kommer fra samme skip.

Dette er et rimelig komplisert problem, så for å lettere komme i gang vil dette kapitlet først behandle en enklere versjon hvor koblingene er kjent. Da blir problemet å finne transformasjonen som gjør at punktene som hører sammen havner tettest mulig. I denne versjonen har ensomme punkt ingen betydning, og kan derfor bare fjernes. Ved å gjøre dette blir eksempelet fra figur 9(a) seende ut som figur 9(c). Her er det betydelig enklere å se at de svarte punktene må flyttes opp og til venstre for å havne nær de blå. Ved korrekt transformasjon er resultatet som i figur 9(b) etter at ensomme punkt er fjernet.

Starten av dette kapitlet behandler problemet med kjente koblinger og definerer et mål på hvor godt W og $r(V, \phi)$ passer sammen ut fra minste kvadraters tenkemåten fra kapittel 2.4. En kan da estimere ϕ^f ved å velge den verdien for ϕ som gjør at W og $r(V, \phi)$ passer best sammen. Metoden herfra generaliseres til det generelle problemet i kapittel 4.2 ved å assosiere AIS-peilinger i V med radarpeilinger i W som ser ut til å passe. Til slutt utledes det i kapittel 4.3 en metode basert på bayesestimering fra kapittel 2.9

4.1 Minste kvadraters estimering ved kjente koblinger

Å løse det fullstendige estimeringsproblemet direkte er utfordrende, derfor startes arbeidet med å finne en estimator ved å se på et enklere problem. Forenklingen som er brukt her er at koblingene mellom AIS- og radarpeilingene er kjent, det vil si at det er kjent hvilke peilinger som stammer fra samme skip. Det er forholdet mellom koblede peilinger som inneholder informasjon om ϕ^f , derfor kan ukoblede



Figur 9: Illustrasjon av assosiasjonsproblemet. De svarte kryssene representerer W , mens de blå punktene representerer V i a) og c) og $r(V, \phi^f)$ i b). Grønne linjer viser hvilke punkter som hører sammen, altså hvilke peilinger som stammer fra samme skip.

peilinger forkastes uten tap av informasjon. De kjente koblingene gjør at dette er mulig, og gjør samtidig at peilingene i W og V kan sorteres slik at AIS-peilingen i kolonne i fra V , v_i kommer fra samme skip som radarpeilingen i kolonne i fra W , w_i . Altså følger det fra kapittel 3.2.3 at sammenhengen mellom w_i og v_i vil være gitt ved likning (52). Det vil si at $w_i = r(v_i; \phi^f) + \epsilon_i$, der ϵ_i er feilen i radarpeilingen w_i . Hvis $E\epsilon = 0$ kan det forventes at w_i er rimelig nær $r(v_i; \phi^f)$. Altså kan metoden fra kapittel 2.4 som gir feilmål av typen i likning (17) brukes. Ved å la r spille rollen til f i likning (17) fås feilmålet

$$J^{MKK}(\phi) = \sum_{i=1}^{N_W} \|w_i - r(v_i, \phi)\|^2. \quad (55)$$

Forkortelsen MKK står for “Minste Kvadraters metode ved Kjente koblinger”. Funksjonen $r(v_i; \phi)$ er som vist i likning (52) meget komplisert, og det er dermed uaktuelt å minimere J^{MKK} analytisk. Det er forsøkt to andre minimeringsmetoder: en som erstatter r med en lineær tilnærming for så å minimere analytisk, og en numerisk metode.

4.1.1 Minimering ved tilnærming

Grunnlaget for tilnærmingen er at det i stor grad er snakk om rimelig små vinkler. Stillingsvinklene vil typisk være under 5 grader og nesten alltid under 10 grader. Peilevinklene er enten begrenset til enten 20 grader eller 5 grader. Det gjør at tilnærmingene $\sin x \approx \tan x \approx \arctan x \approx x$ og $\cos x \approx 1$ har under 6% og 1% feil. Som utgangspunkt for tilnærmingen tas likning (54), andre ordens taylorutviklingen av r . For å få et feilmål som det er mulig å minimere analytisk er det en stor fordel om tilnærmingen er lineær i ϕ , derfor sløyfes også leddene $\phi_x \phi_z$ og $\phi_y \phi_z$. Dette gir tilnærmingen

$$r(v_i; \phi) \approx \begin{bmatrix} v_{1i} + \phi_x - \phi_z v_{2i} \\ v_{2i} - \phi_y + \phi_z v_{1i} \end{bmatrix}. \quad (56)$$

Ved innsetting av denne tilnærmingen blir feilmålet

$$J^{LMK}(\phi) = \frac{1}{2} \sum_{i=1}^{N_W} (w_{1i} + \phi_x v_{2i} - v_{1i} - \phi_x)^2 + (w_{2i} - \phi_z v_{1i} - v_{2i} + \phi_y)^2. \quad (57)$$

Der LMK står for “Linearisert Minste kvadraters metode ved Kjente koblinger”. Dette uttrykket er såpass enkelt at det kan minimeres analytisk, ved å kreve de deriverte lik null. Ved å kreve de partiellderiverte med hensyn på ϕ_x og ϕ_y lik null

fås likningene

$$\frac{\partial J^{LMK}}{\partial \phi_x} = - \sum_i (w_{1i} - v_{1i} + \phi_z v_{2i} - \phi_x) = 0, \quad (58)$$

$$n\phi_x = \phi_z \sum_i v_{2i} + \sum_i w_{1i} - \sum_i v_{1i}, \quad (59)$$

$$\frac{\partial J^{LMK}}{\partial \phi_y} = \sum_i (w_{2i} - v_{2i} - \phi_z v_{1i} + \phi_y) = 0, \quad (60)$$

$$n\phi_y = \phi_z \sum_i v_{1i} - \sum_i w_{2i} + \sum_i v_{2i}, \quad (61)$$

Siden både likningene (59) og (61) inneholder ϕ_z lønner det seg å løse for ϕ_z først. Kravet om at den partiellderiverte av J^{LMK} med hensyn på ϕ_z skal være lik null gir likningen.

$$\begin{aligned} \frac{\partial J^{LMK}}{\partial \phi_z} &= \sum_i v_{2i} (w_{1i} - v_{1i} + \phi_z v_{2i} - \phi_x), \\ &\quad - \sum_i v_{1i} (w_{2i} - v_{2i} - \phi_z v_{1i} + \phi_y) = 0, \end{aligned} \quad (62)$$

$$\begin{aligned} 0 &= \sum_i v_{2i} w_{1i} - \sum_i v_{2i} v_{1i} + \phi_z \sum_i v_{2i}^2 - \phi_x \sum_i v_{2i} \\ &\quad - \sum_i v_{1i} w_{2i} + \sum_i v_{1i} v_{2i} + \phi_z \sum_i v_{1i}^2 - \phi_y \sum_i v_{1i}. \end{aligned} \quad (63)$$

Ved å forkorte like ledd og sette inn for ϕ_x og ϕ_y fra likning (59) og (61) endres likning (63) til

$$\begin{aligned} 0 &= - \sum_i v_{1i} w_{2i} + \sum_i v_{2i} w_{1i} + \phi_z \left(\sum_i v_{1i}^2 + \sum_i v_{2i}^2 \right) \\ &\quad - \frac{1}{n} \left(- \sum_i w_{2i} + \sum_i v_{2i} + \phi_z \sum_i v_{1i} \right) \sum_i v_{1i} \\ &\quad - \frac{1}{n} \left(\phi_z \sum_i v_{2i} + \sum_i w_{1i} - \sum_i v_{1i} \right) \sum_i v_{2i}. \end{aligned} \quad (64)$$

Etter å ha samlet ledd som inneholder ϕ_z kan likning (64) skrives

$$\begin{aligned} 0 &= \phi_z \left(\sum_i v_{1i}^2 + \sum_i v_{2i}^2 - \frac{1}{n} \left(\sum_i v_{2i} \right)^2 - \frac{1}{n} \left(\sum_i v_{1i} \right)^2 \right) \\ &\quad - \sum_i v_{1i} w_{2i} + \sum_i v_{2i} w_{1i} + \frac{1}{n} \sum_i w_{2i} \sum_i v_{1i} \\ &\quad - \frac{1}{n} \sum_i v_{2i} \sum_i v_{1i} - \frac{1}{n} \sum_i w_{1i} \sum_i v_{2i} + \frac{1}{n} \sum_i v_{1i} \sum_i v_{2i}. \end{aligned} \quad (65)$$

I likning (65) forekommer ϕ_z kun en gang. Etter å ha forkortet like ledd kan likningen løses for ϕ_z og stå igjen med at

$$\phi_z = \frac{\sum_i v_{1i} w_{2i} - \sum_i v_{2i} w_{1i} - \frac{1}{n} \sum_i w_{2i} \sum_i v_{1i} + \frac{1}{n} \sum_i w_{1i} \sum_i v_{2i}}{\sum_i v_{1i}^2 - \sum_i v_{2i}^2 + \frac{1}{n} (\sum_i v_{2i})^2 + \frac{1}{n} (\sum_i v_{1i})^2}. \quad (66)$$

Sammen med likningene (59) og (61) bestemmer likning (66) et kritisk punkt for J^{LMK} . Siden J^{LMK} er deriverbar overalt gir denne verdien for ϕ den eneste muligheten for et ekstremalpunkt. Ved å innse at $J^{LMK} \rightarrow \infty$ når $\|\phi\| \rightarrow \infty$, må det kritiske punktet være et minimum. Dette gir at estimatoren $\hat{\phi}^{LMK}$, verdien som minimerer J^{LMK} , er gitt ved

$$\hat{\phi}_z^{LMK} = \frac{\sum_i v_{1i} w_{2i} - \sum_i v_{2i} w_{1i} - \frac{1}{n} \sum_i w_{2i} \sum_i v_{1i} + \frac{1}{n} \sum_i w_{1i} \sum_i v_{2i}}{\sum_i v_{1i}^2 + \sum_i v_{2i}^2 - \frac{1}{n} (\sum_i v_{2i})^2 - \frac{1}{n} (\sum_i v_{1i})^2}, \quad (67)$$

$$\hat{\phi}_x^{LMK} = \frac{1}{n} \left(\hat{\phi}_z^{LMK} \sum_i v_{2i} + \sum_i w_{1i} - \sum_i v_{1i} \right), \quad (68)$$

$$\hat{\phi}_y^{LMK} = \frac{1}{n} \left(\hat{\phi}_z^{LMK} \sum_i v_{1i} - \sum_i w_{2i} + \sum_i v_{2i} \right). \quad (69)$$

4.1.2 Numerisk minimering

Den numeriske minimeringen er utført ved hjelp av funksjonen `fminsearch` som er innebygd i MATLAB (2012). I følge dokumentasjonen bruker den simplex søkemethoden beskrevet i Lagarias et al. (1998). Den finner garantert minimum for to ganger deriverbare funksjoner hvis det kun finnes ett minimum. For de verdiene som er aktuelle i denne oppgaven vil ikke de diskontinuerlige punktene til tangens eller brøkene berøres, dermed vil J^{MKK} være uendelig mange ganger deriverbar. Resultatet fra den lineariserte versjonen tyder på kun ett minimum, og ut fra konstruksjonen forventes det at dette videreføres til det generelle tilfellet. Ut fra dette vil `fminsearch` prestere godt. Estimatet for ϕ basert på denne metoden gis symbolet $\hat{\phi}^{MKK}$.

4.2 Minste kvadraters estimering ved ukjente koblinger

Etter å ha behandlet det forenklete problemet illustrert i figur 9(c) i kan det generelle problemet fra figur 9(a) angripes. Altså antas ikke koblingene lengre kjent. Dette gjør at ensomme peilinger ikke lenger kan forkastes, selv om de fremdeles ikke inneholder informasjon om ϕ , fordi det ikke er kjent hvilke det gjelder. Det er heller ikke mulig å ordne W og V slik at v_i og w_i stammer fra samme skip. Dette gjør at W og V kan ha forskjellig størrelse og det må innføres

en metode for å beskrive koblingene mellom peilingene. For å beskrive koblingene vil det brukes en vektor A^f som har like mange komponenter som det er peilinger i V , N_v . Komponent nummer i , a_i^f , beskriver hvilken radarpeiling AIS-peilingen v_i er koblet med, eller at det ikke finnes noen. Det gjøres ved å la $a_i^f = j$, $1 \leq j \leq N_w$ om v_i er koblet med w_j og la $a_i^f = 0$ om det ikke finnes noen kobling. Koblingene i A^f er i denne versjonen av problemet ukjent, og siden informasjonen om ϕ^f ligger i forholdet mellom koblede peilinger må de estimeres. Et forslag til en kobling mellom en AIS- og radarpeiling kalles en assosiasjon og alle assosiasjonen beskrives med en vektor A som har samme form som A^f . Altså har A N_v komponenter og $a_i = j > 0$ angir at v_i er assosiert med w_j $a_i = 0$ betyr at det ikke er noen radarpeiling assosiert med v_i .

I dette kapittelet blir det utledet to feilmål som kan brukes til å estimere ϕ^f i problemet uten kjente koblinger. Den numeriske metoden fra kapittel 4.1.2 som gir $\hat{\phi}^{MKK}$ vil bli brukt som grunnlag på grunn av at simuleringer beskrevet i kapittel 5.2 viste at den estimerer ϕ^f mer nøyaktig en $\hat{\phi}^{LMK}$. Siden A^f også må estimeres kreves i prinsippet et feilmål $J(\phi, A)$ som minimeres med hensyn på både ϕ og A . Istedet for å gjøre dette direkte er problemet løst ved å finne assosiasjonsmetoder gitt ϕ , lage et feilmål $J(\phi)$ basert på dette og minimere med hensyn på ϕ . Dette valget deler problemet i to: Hvordan finne rimelige assosiasjoner, A , for en gitt ϕ , og hvordan lage et feilmål gitt A . Hvordan feilmålet er konstruert er det første som beskrives videre i dette kapittelet, etter det vil det forklares hvordan NN og GNN fra kapittel 2.5 brukes for å finne A .

4.2.1 Konstruksjon av feilmål gitt A

I dette kapittelet antas det at det finnes en metode som lager et sett assosiasjoner A for en gitt verdi av ϕ , og det ønskes å finne et feilmål $J(\phi)$ basert på dette. Feilmålet J^{MKK} fra kapittel 4.1 brukes som utgangspunkt.

Den mest direkte måten å anvende J^{MKK} i dette tilfellet vil være å bruke A som om den var A^f . Da kan en se bort fra alle ensomme peilinger, v_i hvor i er slik at $a_i = 0$ og w_j hvor j er slik at det ikke finnes i med $a_i = j$, og bruke kvadratavvik tilsvarende J^{MKK} i likning (55). Det vil gi opphav til feilmålet $\sum_{i|a_i \neq 0} \|w_{a_i} - r(v_i; \phi)\|^2$. Dette har et kritisk problem: Verdien av feilmålet kan aldri forverres ved å fjerne en assosiasjon, og hvis ingen AIS-peilinger er assosiert, $a_i = 0 \forall i$, har det verdien 0. Ut fra dette feilmålet vil altså alltid en vilkårlig ϕ verdi og $A = 0$ være bedre enn fasitverdiene $\phi = \phi^f$ og $A = A^f$, noe som ikke er akseptabelt. For å hindre dette gis en straff L^2 for de AIS-peilingene som ikke er assosiert. Det gir det forbedrede feilmålet

$$J_A(\phi) = \sum_{i|a_i \neq 0} \|w_{a_i} - r(v_i; \phi)\|^2 + N_0 L^2, \quad (70)$$

der N_0 er antall AIS-peilinger uten assosiert radarpeiling, det vil si antall i slik at $a_i = 0$. Størrelsen L vil bli referert til som “vinduet” siden den definerer et vindu, et område, rundt v_i som w_j må ligge i for at $a_i = j$ skal gi et mindre bidrag til summen enn $a_i = 0$. Siden leddene i summen har formen $\|w_j - r(v_i; \phi)\|^2$ er randen til vinduet en sirkel med radius L og sentrum i v_i . Dette forbedrede feilmålet løser de problemene som ble presentert over, og egner seg til videre bruk.

4.2.2 NN assosiasjon

NN ble beskrevet i kapittel 2.5 og er en intuitiv metode for å assosiere punkt med hverandre. Den enkleste anvendelsen er å assosiere en AIS-peiling v_i med den radarpeilingen i W som er nærmest, altså konstrueres A ved å la $a_i = \arg \min_j \|w_j - v_i\|$. Men dette er vil ikke gi et godt resultat hvis $\phi^f \neq 0$, så metoden bør også ta hensyn til forslaget på stillingen, ϕ . Dette kan gjøres ved å la avgjørelsen av hva som er nærmest bestemmes av $r(v_i; \phi)$ istedet for v_i direkte. Dette gir at assosiasjonene A bestemmes ved at $a_i = \arg \min_j \|w_j - r(v_i; \phi)\|$. Siden $w_j - r(v_i; \phi^f) = \epsilon$ når w_j er koblet til v_i bør dette gi rimelige assosiasjoner gitt ut fra forslaget for ϕ . Et problem med denne er at den aldri vil gi $a_i = 0$ så lenge det finnes minst en radarpeiling, men problemet tillater AIS-peilinger uten koblet radarpeiling. For å ta hensyn til dette brukes vinduet L fra forrige kapittel ved å fjerne assosiasjoner når det reduserer totalsummen. Det vil si at $a_i = 0$ blir satt til null om minste avstanden fra $r(v_i; \phi)$ til en radarpeiling i W er større enn L , noe som gir assosiasjonene

$$a_i^{NN} = \begin{cases} \arg \min_j \|w_j - r(v_i; \phi)\|, & \min_j \|w_j - r(v_i; \phi)\| \leq L \\ 0, & \text{ellers.} \end{cases} \quad (71)$$

Sammen gir disse assosiasjonene og feilmålet J_A fra likning (70) et fungerende feilmål J^{NN} . På grunn av at $\|w_j - r(v_i; \phi)\|$ brukes både til assosiasjonene og beregning av J_A kan J^{NN} skrives uten å direkte bruke A ,

$$J^{NN}(\phi) = \sum_{i=1}^{N_v} \min \left(L^2, \min_{1 \leq j \leq N_W} \|w_j - r(v_i; \phi)\|^2 \right). \quad (72)$$

$J^{NN}(\phi)$ kan også sees som den minste mulige verdien $J_A(\phi)$ når ϕ er gitt, altså minimum over A . Dette kan skrives som

$$J^{NN}(\phi) = \min_{A|a_i \in \{0, 1, \dots, N_W\}} J_A(\phi). \quad (73)$$

Begge disse to formene er nyttige for å innse at J^{NN} er kontinuerlig med hensyn på ϕ . Det følger fra at minimum over et endelig antall kontinuerlige funksjoner er kontinuerlig, noe som bevises i for eksempel McDonald og Weiss (1999, s.96). J^{NN}

vil ikke være deriverbar siden det oppstår knekkpunkter i punkt hvor minimumet “skifter”, for eksempel steder hvor ledd i i summen endres fra over $\|w_{a_i} - v_i\|$ til L^2 . Vektoren som minimerer J^{NN} gis navnet $\hat{\phi}^{NN}$ og vil være et estimat for stillingen. At kontinuiteten bevares er viktig for søk etter minimum, selv om dette er betydelig vanskeligere enn i tilfellet med kjent assosiasjon. Vinduet gjør at om et forslag for ϕ er langt unna sannheten vil ingen korrekte assosiasjoner gjøres. Dermed kan det ikke finnes noen informasjon om hvor globalt minimum er fra lokale karakteristikk, og det forventes at det finnes mange lokale minimum. I slike tilfeller vil `fminsearch` i følge dokumentasjonen finne lokale minimum, så det må brukes andre minimeringsmetoder. Flere metoder for å finne $\hat{\phi}^{NN}$ er forsøkt, de er beskrevet i kapittel 5.3.

4.2.3 GNN assosiasjon

GNN-assosiasjon er veldig likt NN-assosiasjon, men gir mulighet for å overholde en begrensning på A som ikke NN tar hensyn til. I problemets natur ligger det at en radarpeiling kommer fra ett og bare ett skip, altså bør ingen radarpeiling assosieres med mer enn ett skip. Det er ingenting i likning (73) som hindrer at for eksempel $a_4 = a_7 = 3$, altså at både AIS-peiling nummer fire og syv kobles til radarpeiling nummer tre. Et eksempel på hvordan dette kan skje er vist i figur 5(a) fra kapittel 2.5, la P være radarpeilinger og Q være AIS-peilinger. GNN kan sees på som NN med dette kravet lagt til. Enkleste måten å demonstrere likheten og forskjellene er ved å justere likning (73). Ved å kreve at alle assosiasjonene skal være forskjellige blir resultatet

$$J^{GNN}(\phi) = \min_{\substack{A | a_i \in \{0, 1, \dots, N_W\}, \\ a_i \neq 0 \Rightarrow a_i \neq a_j \forall j \neq i}} J_A(\phi). \quad (74)$$

På samme måte som for J^{NN} gir denne formuleringen at J^{GNN} må være kontinuerlig, men ikke deriverbar. Assosiasjonene som gir minimum i likning (74) kalles GNN-assosiasjonene og skrives A^{GNN} . Verdien av ϕ som minimerer J^{GNN} , $\hat{\phi}^{GNN}$, er GNN-estimatoren for ϕ . Som for NN er metodene brukt for å minimere $\hat{\phi}^{GNN}$ beskrevet i kapittel 5.3.

Som nevnt i kapittel 2.5 er algoritmen som brukes til å finne A^{GNN} og $J^{GNN}(\phi)$ for en gitt ϕ beskrevet i vedlegg B. Hvordan problemet er tilpasset den matematiske formuleringen som kreves er vist i vedlegg B.2.

4.2.4 Sannsynlighet for feilassosiasjoner

Ved å se på rollen vinduet spiller i kapitlene 4.2.2 og 4.2.3 blir det klart at størrelsen på vinduet vil ha en betydelig effekt på assosiasjonene som gjøres. Hvis

vinduet settes veldig lite vil det være vanskelig å få en radarpeiling innenfor, og det vil bli for mange AIS-peilinger uten assosiert radarpeiling. Hvis vinduet settes for stort vil AIS-peilinger uten koblet radarpeilinger assosieres med radarpeilinger fra andre skip urimelig langt borte. For å finne et fornuftig valg for størrelsen på vinduet vil mulige feiltyper og sannsynligheten for dem analyseres i dette kapittelet. NN brukes som utgangspunkt for analysen siden assosiasjonen for hver AIS-peiling er uavhengig av de andre, men resultatene er også relevante for GNN. NN og GNN vil i de fleste tilfeller gi samme assosiasjon, unntaket er når flere AIS-peilinger assosieres til samme radarpeiling av NN. Et eksempel på dette er vist i figur 10. På denne måten er GNN bedre enn NN, slik at feilsannsynligheten uansett ikke bør være større.

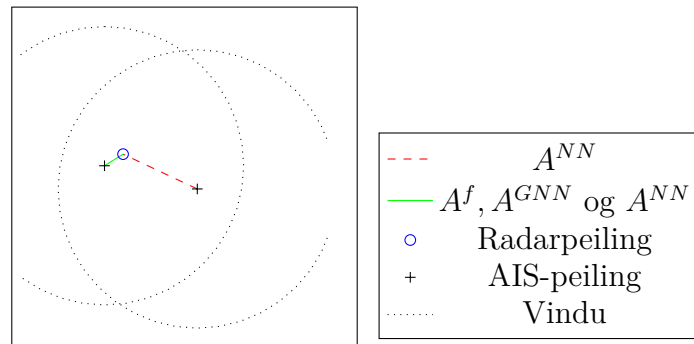
Feiltypene som kan oppstå ved NN-assosiasjon er vist i figur 11 og kan deles inn i følgende kategorier.

1. Radarpeilingene koblet til en AIS-peiling havner utenfor vinduet. Det blir feil uansett om det er en annen radarpeiling innenfor vinduet eller ikke. Tilfellet uten en annen radarpeiling er vist i figur 11(a).
2. Det er mottatt AIS-peiling fra et skip, men ikke radarpeiling. En radarpeiling uten sammenheng er innenfor vinduet. Eksempel vist i figur 11(b) .
3. Både AIS- og radarpeiling fra et skip er mottatt, men det finnes en radarpeiling uten sammenheng nærmere AIS-peilingen enn den riktige. Eksempel vist i figur 11(c).

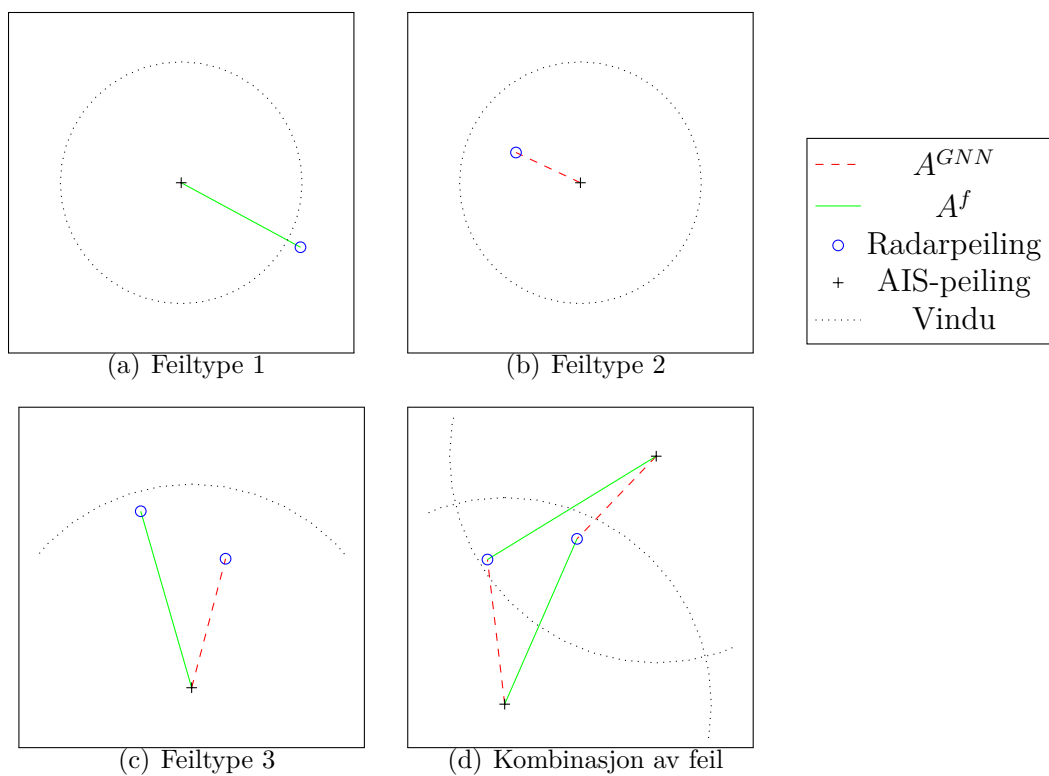
Alle typer feil kan skrives som en kombinasjon av disse feiltypene, for eksempel er feilen i 11(d) en kombinasjon av to type 3 feil. I dette kapittelet vil det finnes tilnærmede uttrykk for sannsynligheten for disse feiltypene. For et skip er de tre feiltypene gjensidig utelukkende, slik at sannsynligheten for en av dem kan beregnes ved å legge dem sammen.

Under disse utledningene antas det at $\phi = 0$, som gir samme resultat som $\phi = \phi^f$ om v_i erstattes $r(v_i; \phi)$. Under denne antakelsen vil $w_{\alpha_i^f} = u_i + \varepsilon_i$. Dette vil forenkle regningen og vil ikke gjøre resultatet mindre gyldig fordi om estimatorene fungerer tilfredsstillende vil ϕ^{GNN} og ϕ^{NN} være nær fasit (ϕ^f), slik at resultatene fra $\phi = \phi^f$ og $\phi = \phi^{NN}$ bør være rimelig like. Feilen ε_i antas å være normalfordelt med null varians og kovariansmatrise $\sigma_w^2 I$.

Siden NN-assosiasjonen for v_i er uavhengig av assosiasjonen for $v_j, j \neq i$ vil det kun være nødvendig å se på en AIS-peiling om gangen, som vil skrives v . Hvis det er flere ukoblede radarpeilinger nær v er det kun den nærmeste som er aktuell for å bli assosiert, altså er det kun nødvendig å se på to forskjellige radarpeilinger: en koblet og en ukoblet. Den koblede peilingen vil skrives ω_k , tilsvarende skrives den ukoblede peilingen ω_u . Det betyr at $\omega_k = v_i + \epsilon$, der ϵ har en bivariat normalfordeling med forventningsverdi 0 og kovariansmatrise $\sigma_w^2 I$.



Figur 10: Eksempel på hendelse som det ikke er tatt hensyn til ved beregning av sannsynligheten for type 2 feil i kapittel 4.2.4.



Figur 11: Feiltyper for NN. De grønne linjene angir koblede AIS og radarpeilinger som lagret i A^f . Røde linjer angir assosiasjonene gjort ved GNN. Sorte sirkler angir vinduet til punktet i sentrum, om det ikke er noen radarpeiling innenfor den vil ingen assosiasjon utføres.

For å lettere kunne beregne sannsynligheten vil det også antas at ω_u er uniformt fordelt over de aktuelle verdiene bestemt av åpningsvinkelen, som er det vinkelområdet som satellitten kan motta peilinger fra. Dette tilsvarer at skipsposisjonene er fordelt slik at AIS-peilingene de gir opphav til er uniformfordelte.² Det antas også at mottak av AIS-meldinger og radarpeilinger fra et skip er uavhengig av mottak av dette fra alle andre skip.

Sannsynlighet for å koblet peiling utenfor vinduet. (Type 1 feil). For å at denne feilen skal skje må satellitten motta både AIS-peiling og radarpeiling for et skip, og feilen ϵ i radarpeilingen må være så stor at den havner utenfor vinduet. Det vil si at $\|v - \omega_k\| = \|\epsilon\| \geq L = l\sigma_w$. Siden peilingene er todimensjonale følger det fra likning (29) at

$$P(\|v - \omega_k\| \leq l\sigma_w | v \text{ og } \omega_k \text{ eksisterer}) = 1 - \exp\left(-\frac{l^2}{2}\right). \quad (75)$$

Hvis vinduet L settes til $l\sigma_w$ er altså sannsynligheten for at w_k havne innenfor vinduet til v $1 - \exp(-l^2/2)$, gitt at begge peilingene mottas. Dermed må sannsynligheten for å havne utenfor vinduet være $\exp(-l^2/2)$. Videre vil det at AIS-peilingen v eksisterer skrives $\exists v$, og tilsvarende for de andre peilingene. Sannsynligheten over kan brukes til å finne sannsynligheten for at det skal oppstå en type 1 feil fra et bestemt skip.

$$P(\text{Type 1 feil fra bestemt skip}) = P(\|v - \omega_k\| > l\sigma_w) \quad (76)$$

$$= P(\|v - \omega_k\| > l\sigma_w | \exists v, \omega_k) P(\exists v) P(\exists \omega_k) \quad (77)$$

$$= \exp\left(-\frac{l^2}{2}\right) P_{AIS} P_{rad} = P_{f1}. \quad (78)$$

Her representere P_{rad} og P_{AIS} sannsynlighetene for å motta henholdsvis en radarpeiling og en AIS-peiling fra et bestemt skip. Siden rapportene fra de forskjellige skipene er antatt uavhengige av hverandre kan de totalt antall type 1 feil ses på som et binomisk eksperiment med feilsannsynligheten P_{f1} for “suksess”. Dette gir at fordelingen for antall type 1 feil må være

$$P(n \text{ type 1 feil}) = \binom{N_s}{n} P_{f1}^n (1 - P_{f1})^{N_s - n}. \quad (79)$$

Hvor N_s er antall skip som kan gi opphav til peilinger. Forventet antall feil er $N_s P_{f1}$. For $N_s = 50$ og $P_{AIS} = P_{rad} = 0.9$ blir resultater for forskjellige l som vist i tabell 1. Sannsynligheten for å unngå type 1 feil helt ($n = 0$) og sannsynligheten per skip P_{f1} er vist i samme tabell. Valg av parameterverdier blir forklart i kapittel 5.1.2.

²Om en ser bort fra kantene, hvor sannsynlighetstettheten for ω_u vil være litt lavere. Antar at dette kantlaget er så tynt at det er uten betydning.

Vindustørrelse - l	1	2	3	4	5
Forventet antall	24.56	5.481	0.4499	0.01359	0.0001509
Sanns. for ingen	2.106e-15	0.003011	0.6364	0.9865	0.9998
Sanns. pr skip - P_{f1}	0.4913	0.1096	0.008998	0.0002717	3.019e-06

Tabell 1: Forventet antall type 1 feil og sannsynligheten (sanns.) for ingen type 1 feil ved $L = l\sigma_w$.

Sannsynligheten for å få feil radarpeiling innenfor vinduet når den riktige ikke eksisterer. (Type 2 feil). Sannsynligheten for type 2 feil avhenger av åpningsvinkel, antall skip og vinduet. Gitt AIS-peilingen v og den ukoblede raderpeilingen ω_u eksisterer. På grunn av uniformfordelingen vil sannsynligheten for å havne innenfor vinduet være lik forholdet mellom arealet på overvåkingsområdet og vinduet. Ved å anta at åpningsvinkelen er slik at $-v_1/2 \leq v_1 \leq v_1/2$ og $-v_2/2 \leq v_2 \leq v_2/2$ vil arealet på overvåkingsområdet være $v_1 v_2$. I det tilfellet vil $v = (v_1, v_2)$ angi størrelsen på åpningsvinkelen, alternativt skrevet at åpningsvinkelen er $v_1 \times v_2$. Arealet på vinduet er $\pi(lr)^2$, og sannsynligheten blir da.

$$P(|v - \omega_u|^2 < l^2 \sigma_w^2 | \exists v, \exists \omega_u) = \frac{\pi l^2 \sigma_w^2}{v_1 v_2}. \quad (80)$$

For GNN må man her ta hensyn til at det ikke vil bli en slik feil om AIS-peilingen fra samme skip som ω_u ligger nærmere ω_u enn v , det gjelder ikke for NN. Men denne feilen kan oppstå med vinkelpeiling fra alle de $N_S - 1$ skipene som ikke har gitt opphav til v . For å finne den totale sannsynligheten må det også tas hensyn til sannsynligheten for å motta en radarpeiling. For å finne sannsynligheten for minst en ukoblet peiling innenfor vinduet beregnes sannsynligheten for ingen innenfor, som gir.

$$P(\text{Type 2 feil} | \exists v, \omega_u \nexists \omega_k) = 1 - \left(1 - \frac{\pi l^2 \sigma_w^2}{v_1 v_2} P_{rad}\right)^{N_S - 1}. \quad (81)$$

For at et bestemt skip i skal gir opphav til en type 2 feil må det ha AIS-peiling, men ikke radarpeiling. Dette gir

$$P(\text{Type 2 feil for skip } i) = \left(1 - \left(1 - \frac{\pi l^2 \sigma_w^2}{v_1 v_2} P_r\right)^{N_S - 1}\right) (1 - P_r) P_{AIS} = P_{f2}. \quad (82)$$

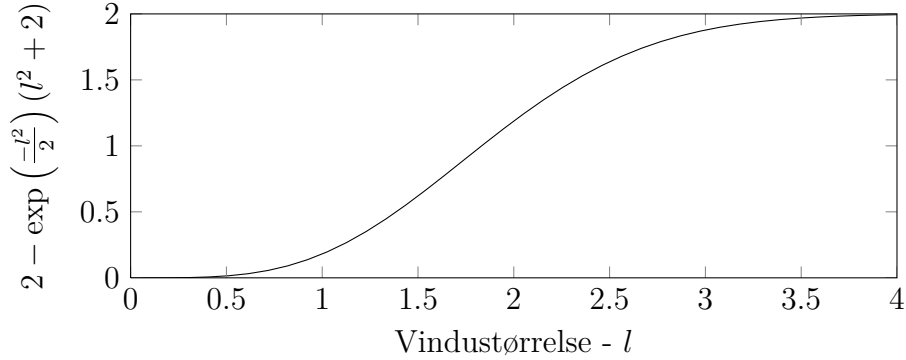
Hvis en antar at dette er uavhengig får en sannsynligheten for n feil på samme måte som likning (79) gir for type 1 feil,

$$P(n \text{ type 2 feil}) = \binom{N_S}{n} P_{f2}^n (1 - P_{f2})^{N_S - n}. \quad (83)$$

Sannsynlighet for ingen, og forventet antall type 2 feil under uavhengighetsantakelsen er vist i tabell 2. For N_S , P_{AIS} og P_{rad} er det brukt samme verdier som

Vindustørrelse - l	1	2	3	4	5
Forventet antall	0.06192	0.2427	0.5282	0.8965	1.321
Sanns. for ingen	0.9399	0.784	0.588	0.4047	0.2621
Sanns. pr skip - P_{f2}	0.001238	0.004855	0.01056	0.01793	0.02643

Tabell 2: Forventet antall type 2 feil og sannsynligheten (sanns.) for ingen type 2 feil ved forskjellige l .



Figur 12: Plott av l avhengig del av sannsynligheten for type 3 feil for gitt skip og radarpeiling fra bestemt, annet skip.

feiltype 1 i tillegg til $v_1 = v_2 = 10$ grader og $\sigma_w = 0.1$ grader. Siden denne typen feil krever at det ikke finnes radarpeiling til skipet, men krever radarpeiling fra annet skip er type 2 feil definitivt ikke uavhengig mellom skip, og likning (83) må brukes med varsomhet.

Både rett og feil radarpeiling er innenfor vinduet, men feil er nærmest. (Type 3 feil). I kapittel 2.7 ble det utledet at Mahalanobis-avstanden r har fordelingen $f(r) = r \exp(-r^2/2)$ i to dimensjoner. Med kovariansmatrise $\sigma_w^2 I$ tar likning (25) formen $r = \|w_i - v_i\|/\sigma_w$, altså er r avstand målt i antall standardavvik. Gitt at avstanden mellom v og ω_k er r standardavvik vil sannsynligheten for at ω_u er innenfor være $\frac{\pi r^2 \sigma_w^2}{v_1 v_2}$ når det er kjent at ω_r finnes. Ved å integrere over mulige r når $L = l\sigma_w$ blir sannsynligheten for å få type 3 feil fra et bestemt skip

$$P(\|v - \omega_u\| < r < l | \exists v, \omega_k, \omega_u) = \int_0^l \frac{\pi r^2 \sigma_w^2}{v_1 v_2} P_{rad} r \exp\left(-\frac{r^2}{2}\right) dr \quad (84)$$

$$= \frac{\pi \sigma_w^2}{v_1 v_2} P_{rad} \left[-\exp\left(\frac{-r^2}{2}\right) (r^2 + 2) \right]_0^l$$

$$P'_{f3} = \frac{\pi \sigma_w^2}{v_1 v_2} P_{rad} \left(2 - \exp\left(\frac{-l^2}{2}\right) (l^2 + 2) \right). \quad (85)$$

Vindustørrelse - l	1	2	3	4	5
Forventet antall	0.1011	0.6612	1.04	1.104	1.107
Sanns. for ingen	0.9037	0.5139	0.3495	0.3275	0.3264
Sanns. pr skip - P_{f1}	0.002022	0.01322	0.02081	0.02208	0.02214

Tabell 3: Forventet antall type 3 feil og sannsynligheten (sanns.) for ingen type 3 feil ved forskjellige l .

Sannsynligheten ved bruk av GNN vil også for denne feiltypen kunne forventes å være litt bedre siden GNN kan hindre slike feil på grunn av globale hensyn. Men dette krever at AIS-peilingen som er koblet til w_u er nærmere den enn w_k .

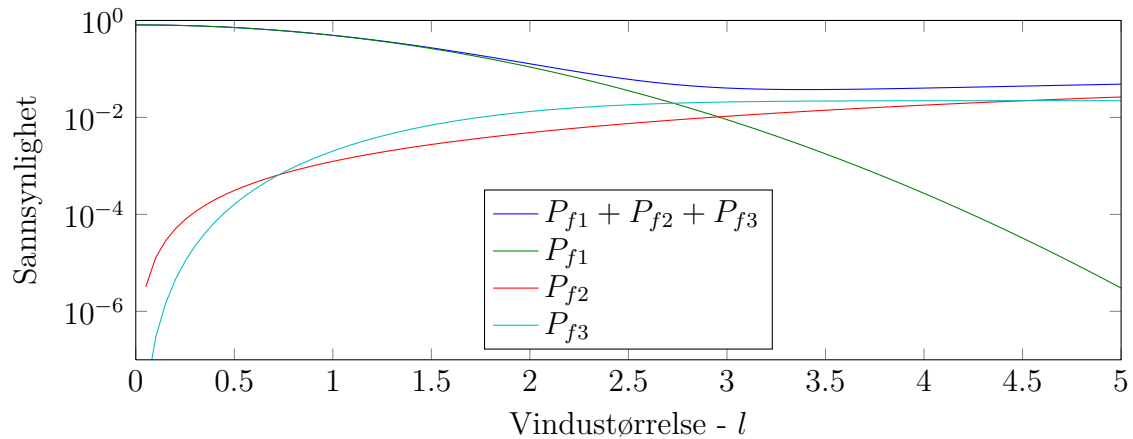
Likning (86) viser at sannsynligheten for type 3 feil er proporsjonal med andelen vinduet utgjør av hele åpningsvingelen når $l = 1$. Den justerende funksjonen $(2 - \exp(-l^2/2)(l^2 + 2))$ starter i null og har grenseverdien 2. Denne er i praksis nådd for $l = 4$ med en verdi på 1.994, for $l = 3$ er verdien 94% av maksimum og resten av formen er vist i figur 12. Dette betyr at sannsynligheten P'_{f3} maksimalt er 2 ganger andelen av totalt vinkelområde som dekkes av en sirkel med radius ett på standardavvik.

For at et bestemt skip skal gi opphav til en slik feil må både AIS- og radarpeilingen eksistere, dette har sannsynlighet $P_{AIS}P_r$. I tillegg må radarpeilingen være innenfor vinduet og minst en av de andre skipenes radarpeiling havne nærmere enn radarpeilingen. For hvert av de andre skipene har dette sannsynligheten P'_{f3} , noe som på grunn av uavhengighet gir

$$P(\text{Type 3 feil for bestemt skip}) = \left(1 - (1 - P'_{f3})^{N_S - 1}\right) P_{AIS}P_r = P_{f3} \quad (86)$$

Det er samme problemet med uavhengighet mellom skip for type 3 som for type 2 feil. Men om en uansett antar uavhengighet blir resultatet som vist i tabell 3. Det er brukt samme verdiene som for feiltype 2.

Valg av beste l ut fra sannsynlighetene. Figur 13 viser sannsynligheten for å få de forskjellige feiltypene for et gitt skip som funksjon av l . Det er tydelig at feiltype 1 dominerer klart frem til omtrent $l = 2.75$. Totalsannsynligheten er svært nær minimum ved $l = 3$, og velges derfor som standardverdi for vindustørrelsen. Ved denne verdien er $P_{f1} = 0.009$, $P_{f2} = 0.011$, $P_{f3} = 0.021$ og summen av dem 0.040. Ved uavhengighet vil det da forventes en 96% riktige assosiasjoner, som tilsvarer 2 feilassosiasjoner ved totalt finnes 50 skip.



Figur 13: Plott av feilsannsynlighet som funksjon av vindustørrelsen i antall standardavvik, l . $L = l\sigma_w$. $P_{f2}(0) = P_{f3}(0) = 0$, noe som ikke vises på grunn av logaritmisk y-akse.

4.3 Metropolis-Hasting simulering ved ukjent assosiasjon

Etter å ha sett grundig på flere minste kvadraters-alternativer er det på tide å se nærmere på bayesestimeringen beskrevet i kapittel 2.9. Det ønskes å finne en aposteriorifordeling for stillingen ϕ og assosiasjonene A samlet, gitt peilingene. Dette vil gjøres ved å bruke likning (33) om gir at $f(\phi, A|V, W) \propto f(V, W|\phi, A)f(\phi, A)$, og så bruke MH for å undersøke $f(\phi, A|V, W)$. I beskrivelsen av MH i kapittel 2.9 er det bare en parameter θ som estimeres mot både ϕ og A her. Ved å la θ være vektoren sammensatt av ϕ og A kan to reduseres til en, men det er hensiktsmessig å holde dem adskilt i dette kapitlet. Ved simulering av kjeden kan ofte verdien på aposteriorifordelingen og relaterte størrelser få enormt lave eller høye verdier. Derfor er det nødvendig å gjøre implementeringen av algoritmen på log-skala, og uttrykk for logaritmen til fordelingene vil bli utledet. Et annet implementasjonstriks er at siden det kun kreves å kjenne en funksjon proporsjonal med $f(\phi, A)$ kan alle faktorer som ikke påvirkes av W, V, A eller ϕ i $f(\phi, A|V, W)$ fjernes.

For å bruke MH er det to sentrale elementer, fordelingen $f(\phi, A|V, W)$ som det ønskes å trekkes fra og forslagsfordelingen $p(\phi, A, \tilde{\phi}, \tilde{A})$. Dette kapitlet starter med å utlede et uttrykk proporsjonalt med $f(\phi, A|V, W)$, så beskrives trekkfordelingen som brukes og til slutt beskrives det hvordan dette er blitt satt sammen.

4.3.1 Aposteriorfordelingen $f(\phi, A|V, W)$

Fra likning (33) følger det at $f(\phi, A|V, W) \propto f(V, W|\phi, A)f(\phi, A)$ når det antas at apriorifordelingene til A og ϕ er uavhengige ($f(\phi, A) = f(\phi)f(A)$). Altså må apriorifordelingene til ϕ og A finnes. Det samme gjelder den betingede fordelingen for peilingene gitt stilling og assosiasjoner. Til sammen gir de et uttrykk proporsjonalt med den ønskede fordelingen.

Modellen tar utgangspunkt i at W og V er mottatt, og konstruerer fordelingene ut fra det. Det innebærer at antall mottatte AIS- og radarpeilinger er kjent. Videre vil N_V representere antall mottatte AIS-peilinger som også vil være lengden på A . For en gitt A angir N_0 antall elementer i A som er lik null. N_W representerer antall radarpeilinger som er mottatt og N_{W0} antall radarpeilinger som ikke er koblet til en AIS-peiling under A .

Apriorifordelingen for ϕ . Skjevstillingen antas normalfordelt om origo med kovariansmatrise $\sigma_\phi^2 I$. Siden ϕ er tredimensjonal gir det fordelingen

$$f(\phi) = \frac{1}{(2\pi)^{3/2} \sqrt{|\sigma_\phi^2 I|}} \exp\left(-\frac{1}{2} \phi^T \frac{1}{\sigma_\phi^2} I \phi\right) \quad (87)$$

$$= \frac{1}{(2\pi)^{3/2} \sigma_\phi^3} \exp\left(-\frac{1}{2\sigma_\phi^2} \phi^T \phi\right) \quad (88)$$

$$\ln f(\phi) = -\frac{3}{2} \ln 2\pi - 3 \ln \sigma_\phi - \frac{1}{2\sigma_\phi^2} \phi^T \phi \quad (89)$$

De to første leddene kan sees bort fra siden de tilsvarer proporsjonalitetskonstanter.

Apriorifordelingen for A . Sannsynligheten for at et skip har radarpeiling P_{rad} er kjent på forhånd. Siden de to datatypene er uavhengige av hverandre må sannsynligheten for skipet som ga opphav til v_i for en vilkårlig i har en koblet radarpeiling være P_{rad} . Det vil si $P(a_i \neq 0) = P_{rad}$. Om en AIS-peiling er er assosiert med en radarpeiling eller ikke er uavhengig av alle andre peilinger, men hvilken den eventuelt er assosiert med kan ikke være det siden AIS-retninger ikke kan være koblet til samme peiling. Det gjør at to assosiasjoner A og A' som har like mange nullelementer bør ha samme sannsynlighet. Altså bør $f(A)$ kun være avhengig av antall nullelementer i A , N_0 , i tillegg til lengden på V og W , N_V og N_W . Sannsynligheten for å få en A med N_0 null-elementer må bli

$$(1 - P_{rad})^{N_0} P_{rad}^{N_V - N_0}. \quad (90)$$

Binominalkoeffisienten er ikke med siden rekkefølgen har betydning. For de elementene som ikke er null har en så N_W valg for den første verdien, $N_W - 1$ for den

andre og så videre for de $N_V - N_0$ elementene frem til det er $N_W - (N_V - N_0 + 1)$ valg for den siste. Det gir apriorifordelingen

$$f(A) = (1 - P_{rad})^{N_0} P_{rad}^{N_V - N_0} \frac{(N_W - N_V + N_0)!}{N_W!}. \quad (91)$$

Som en sjekk for at uttrykket er rett kan det kontrolleres at summen over alle akseptable A er null. For en gitt verdi av N_0 finnes det $\binom{N_V}{N_0}$ måter å velge ut de N_0 elementene som skal være null. For hver av disse mulighetene kan ikke-nullelementene velges på $N_W! / (N_W - N_V + N_0 - 1)!$ måter. Siden hver av disse mulighetene har samme antall null-verdier har de samme sannsynlighet, som totalt blir

$$(1 - P_{rad})^{N_0} P_{rad}^{N_V - N_0} \frac{(N_W - N_V + N_0)!}{N_W!} \binom{N}{N_0} \frac{N_W!}{(N_W - N_V + N_0)!}, \quad (92)$$

hvor fakultetsbrøkene forkortes. Totalt blir summen av sannsynligheten for alle mulige A

$$\sum_A f(A) = \sum_{N_0=0}^{N_V} (1 - P_{rad})^{N_0} P_{rad}^{N_V - N_0} \binom{N}{N_0} = 1. \quad (93)$$

Den siste summen er 1 siden det er det samme som summen av en binomisk fordeling med suksesssannsynlighet P_{rad} og N_V forsøk.

Logaritmen til fordelingen blir

$$\begin{aligned} \ln f(A) &= N_0 \ln(1 - P_{rad}) + (N_V - N_0) \ln P_{rad} \\ &\quad + \ln(N_W - N_V + N_0)! - \ln N_W! \end{aligned} \quad (94)$$

Leddene $-\ln N_W$ kan sløyfes siden det er likt for alle og det kun er nødvendig å kjenne $f(A)$ opp til en konstant. Det samme gjelder $N_V \ln P_{rad}$ som står igjen etter å ha løst opp parentesene.

Betinget fordeling for peilingene gitt ϕ og A . Feilen i vinkelpeilingene antas normalfordelt med kovarians $\sigma_v^2 I$ rundt koblet AIS-peiling. Hvis det ikke finnes en koblet AIS-peiling antas radarpeilingen uniformfordelt over åpningsvinkelen, som i kapittel 4.2.4. I tillegg er radarpeilingene uavhengige av hverandre, altså blir fordelingen

$$f(w_j, V | A, \phi) = \begin{cases} n(w_j; r(v_i; \phi), \sigma_v^2 I) & \exists i | a_i = j \\ 1/O & \text{ellers} \end{cases} \quad (95)$$

$$f(W, V | A, \phi) = \prod_j f(w_j | A, V, \phi) \quad (96)$$

hvor $v_1 v_2$ uttrykker størrelsen på overvåket vinkelområde og $n(x; \mu, \Sigma)$ normalfordelingen i punktet x når forventningsverdien er μ og kovariansmatrisen Σ . Det eksakte uttrykket for fordelingen en gitt ved likning (23) hvor $d = 3$. Ved å ta logaritmen til likning (96) og forenkle den resulterende summen kan logaritmen til fordelingen skrives

$$\ln f(W, V|A, \phi) = \sum_{j=1}^{N_W} \ln f(w_j | A, V, \phi) \quad (97)$$

$$= \sum_{i|a_i \neq 0} \ln n(w_{a_i}, r(v_i; \phi), \sigma_v I) - N_{W0} \ln v_1 v_2. \quad (98)$$

Fra uttrykket for normalfordelingen i likning (23) blir

$$\ln n(w_{a_i}, r(v_i; \phi), \sigma_v I) = -\ln 2\pi\sigma_v^2 - \frac{1}{2\sigma_v^2} \|w_{a_i} - r(v_i; \phi)\|^2 \quad (99)$$

siden $(w_{a_i} - r(v_i; \phi))^T (\sigma^2 I)^{-1} (w_{a_i} - r(v_i; \phi)) = \frac{1}{\sigma^2} \|w_{a_i} - r(v_i; \phi)\|^2$. Til sammen gir dette

$$\begin{aligned} \ln f(W, V|A, \phi) &= -(N_W - N_V + N_0) \ln v_1 v_2 \\ &\quad - \sum_{i|a_i \neq 0} \frac{1}{2\sigma_v^2} \|w_{a_i} - r(v_i; \phi)\|^2 - (N_W - N_0) \ln 2\pi\sigma_v^2. \end{aligned} \quad (100)$$

Etter å ha løst opp parentesene kan leddene som kun inneholder $N_W, N_V, v_1 v_2$ og σ_v fjernes siden de tilsvarer proporsjonalitetskonstanter.

Uttrykk proporsjonalt med $f(\phi, A|V, W)$. Grensefordelingen til MH trekke ønskes å være $f(\phi, A|V, W) \propto f(V, W|\phi, A)f(\phi)f(A)$. Det tilsvarer at $\ln f(\phi, A|V, W) = \ln f(V, W|\phi, A) + \ln f(\phi) + \ln f(A) + \text{konstant}$. Konstanten er konstant med hensyn på ϕ og A . Fra uttrykkene i likningene (89), (94) og (100) følger det da at

$$\begin{aligned} \ln f(\phi, A|V, W) &= -\frac{1}{2\sigma_\phi^2} \phi^T \phi + N_0 \ln(1 - P_{rad}) - N_0 \ln P_{rad} + \ln(N_W - N_V + N_0)! \\ &\quad + N_0 \ln 2\pi\sigma_v - \sum_{i|a_i \neq 0} \frac{1}{2\sigma_v^2} \|w_{a_i} - r(v_i; \phi)\|^2 - N_0 \ln v_1 v_2 + \text{konstant}. \end{aligned} \quad (101)$$

Dette er det nødvendige uttrykket for fordelingen som det skal trekkes fra.

4.3.2 Forslagsfordelingen

Forslagsfordelingen er som nevnt i kapittel 2.9 i prinsippet vilkårlig. Det vil si at markovkjeden i MH konvergerer mot den ønskede fordelingen uansett. Men

valg av forslagsfordeling kan i praksis ha en enorm praktisk betydning, siden en godt konstruert fordeling kan gjøre at “mange nok” iterasjoner blir få nok til at det er realistisk å simulere frem til det punktet. Ved å bruke trikset beskrevet i kapittel 2.9.2 er det brukt tre forskjellige forslagsfordelinger, og hvilken som brukes avhenger av en tilfeldig variabel. Det er en fordeling som bare endrer på ϕ , en som kun endrer på A og en som endrer på begge.

Søk etter symmetrisk forslagsfordeling. Som første forsøk ønskes det en symmetrisk forslagsfordeling for ϕ og A separat. Det vil si å finne forslagsfordeling $q(\phi, \tilde{\phi})$ som oppfyller $q(\phi, \tilde{\phi}) = q(\tilde{\phi}, \phi)$ og tilsvarende for A . Dette gjør at uttrykket for akseptanssynligheten α blir uavhengig av q . Det er lett å få fordelingen symmetrisk med hensyn på ϕ , bare la fordelingen til den nye verdien avhenge av $\phi - \tilde{\phi}$. For det nye forslaget kan for eksempel trekkes fra en normalfordeling som har den gamle verdien som forventningsverdi og den diagonale kovariansmatrisen $\sigma_{MH}^2 I$. For A er dette mer utfordrende.

Det er kravet om at flere AIS-peilinger ikke kan kobles til samme radarpeiling som gir trøbbel, og for å komme til bunns i dette utføres enda en forgrening. Hvis algoritmen velger å endre A vil MH gjøre det på følgende måte:

- a. Med sannsynlighet $P_0^{N_0}$ endre en komponent $a_i > 0$ til 0.
- b. Med sannsynlighet $P_b^{N_0}$ bytt en komponent $a_i > 0$ til en annen, ikke brukt $j > 0$.
- c. Med sannsynlighet $P_{-0}^{N_0}$ endre en komponent $a_i = 0$ til en ubrukt $j > 0$.

Summen av de tre sannsynlighetene må bli en. Komponentene kan anta verdier fra 0 til N_W , $0 \leq a_i, j \leq N_W$. N_0 som vanlig for antall komponenter $a_i = 0$. Matematisk blir det $N_0 = |\{i | a_i = 0, 1 \leq i \leq N_V\}|$, hvor $|\cdot|$ angir antall elementer i mengden.

For å oppnå en symmetrisk fordeling må sannsynligheten for å gå fra $a_i = j$ til $\tilde{a}_i = j'$ være den samme som å gå fra $a_i = j'$ til $\tilde{a}_i = j$. For tilfelle b løses dette ved en uniform fordeling over $\{j | a_i \neq j \forall i, 1 \leq j \leq N_W\}$, mengden av radarpeilinger som ikke er brukt. Den inneholder $N_{W0} = (N_W - N_V) + N_0$ elementer. Den enkleste måten å se dette på er først å la det være like mange av begge typene peilinger, $N_W = N_V$. Da vil antall ukoblede være like mange av begge typer, N_0 , så må det legges til $N_W - N_V$ elementer på grunn av differansen i antall peilinger.

Tilfelle a og c er koblet siden reverseringen av a krever at c utføres, og motsatt. For å beregne disse sannsynlighetene må metodene spesifiseres nærmere:

- a. Velg en AIS-peiling uniformt fra de som har en assosiert radarpeiling og fjern denne assosiasjonen. Det vil si velg i uniformt fra $\{i | a_i = 0, 1 \leq i \leq N_V\}$ og

sett $a_i = 0$.

Størrelsen på mengden er $N_V - N_0$.

- c. Velg en AIS-peiling uniformt fra de som ikke har assosiert radarpeiling, assosier den med en radapeiling valgt uniformt fra de ledige. Det vil si velg i uniformt fra $\{i | a_i \neq 0, 1 \leq i \leq N_V\}$ og j uniformt fra $\{j | a_i \neq j \forall i, 1 \leq j \leq N_W\}$. Sett $a_i = j$.

Størrelsen på mengdene er N_0 og $N_W - N_V + N_0$.

Hvis en da ser på to koblede tilfeller og krever at sannsynligheten P for hver av dem skal være like får en:

For metode (a) la $N_0 = n, a_i = 0$. Sannsynligheten for å endre a_i til j blir da

$$P = P_{-0}^n \frac{1}{n} \frac{1}{N_W - N_V + n}. \quad (102)$$

Hvor faktorene representerer henholdsvis: velg metode c, velg komponent i og velg å endre komponent i til j .

Retur må gjøres ved metode (c) og da er $N_0 = n - 1, a_i = j$. Sannsynligheten for å endre a_i til 0 er

$$P = P_0^{n-1} \frac{1}{N_V - n + 1}. \quad (103)$$

Hvor faktorene er henholdsvis: velg metode a og velg å sette komponent i til 0. Sammensetning av likningene gir

$$P_{-0}^n = \frac{n(n - (N_V - N_W))}{N_V - n + 1} P_0^{n-1}, \quad \max\{0, N_V - N_W\} \leq n \leq N_V. \quad (104)$$

Betingelsen på n kommer av at det ikke være flere nullelementer enn elementer i A , at det ikke kan være færre enn ingen og det ikke kan assosieres flere AIS-peilinger enn antall radarpeilinger, N_W . Fra likning (104) er det klart at en av P_{-0}^n og P_0^{n-1} kan velges fritt, så lenge $0 < P_{-0}^n, P_0^{n-1} < 1$. For å finne aktuelle valg er det verdt å se på ekstremverdiene til faktoren som skiller den,

$$\frac{n(n - (N_V - N_W))}{N_V - n + 1}. \quad (105)$$

På grunn av betingelsene på n er alle faktorene positive. Det er også lett å se at nevneren stiger med n , mens telleren synker. Dermed vil ekstremalverdiene skje ved ekstremalverdiene for n . Høyeste n er N_V som gir

$$\frac{N_V(N_V - (N_V - N_W))}{N_V - n + 1} = N_V N_W. \quad (106)$$

Laveste verdier for n er 1 om $N_W > N_V$, og $N_V - N_W + 1$ ellers. Dette gir henholdsvis

$$\frac{1 - N_V + N_W}{N_V} \quad \text{og} \quad \frac{N_V - N_W + 1}{N_W}. \quad (107)$$

Aktuelle verdier $N_W = N_V = 80$, som gir maksimalverdi 6400 og minimumverdi $1/80$. Dette krever at enkelte sannsynligheter må være mindre enn $1/6400$, noe som vil si at det er veldig vanskelig for kjeden å nå noen områder. Dette er ikke ønskelig, derfor vil det gjøres et nytt, relatert, forsøk på å endre A .

Nesten symmetrisk forslagsfordeling. Siden den mulige symmetriske forslagsfordelingen ved økning eller senking av N_0 ikke ble så bra vil det søkes en alternativ måte å endre A på. Dette gjøres ved å bruke metodene a og c ovenfra med faste sannsynligheter, altså er $P_{-0}^{N_0} = P_{-0}$ og $P_0^{N_0} = P_0$ konstant med hensyn på N_0 . Ved disse metodene må sannsynlighetene beregnes, noe som allerede er utført i likningene (102) og (103). Fra den kan vi se at for \tilde{A} trekt ved metode a, når det er N_0 nullelementer i A , vil forholdet mellom forslagsfordelingene bli

$$\frac{q(\tilde{A}, A)}{q(A; \tilde{A})} = \frac{P_{-0} \frac{1}{N_0+1} \frac{1}{N_W - N_V + N_0 + 1}}{P_0 \frac{1}{N_V - N_0}} = \frac{P_{-0}(N_V - N_0)}{P_0(N_0 + 1)(N_W - N_V + N_0 + 1)}. \quad (108)$$

Siden telleren blir sannsynligheten for å sette et bestemt element i til j når det er $N_0 + 1$ nullelementer, og nevneren er sannsynligheten til å sette element i til null når det er N_0 nullelementer. $q(\tilde{A}, A)$ er kortform for $q(\phi, \tilde{A}, \phi A)$, siden ϕ ikke endres. På tilsvarende måte gir c

$$\frac{q(\tilde{A}, A)}{q(A; \tilde{A})} = \frac{P_0 \frac{1}{N_V - N_0 + 1}}{P_{-0} \frac{1}{N_0} \frac{1}{N_W - N_V + N_0}} = \frac{P_0 N_0 (N_W - N_V + N_0)}{P_{-0} (N_V - N_0 + 1)}. \quad (109)$$

For at beskrivelsen av denne forslagsfordelingen skal være fullstendig må bare ekstremtilfellene behandles. Når N_0 er maksimal/minimal er der vil det ikke være mulig å gå opp/ned eller bytte verdier. Hvis behandlingen av dette endrer P_0 eller P_{-0} vil ikke uttrykkene for forslagsfordelingene i likningene (108) og (109) være gyldige lenger. Dette løses ved å la endring av ϕ overta sannsynligheten, det vil si at om en umulig variant velges så endres ϕ istedet.

Omfattende endring av tilstand. Alle de foregående forslagene til endringer er forsiktige ved at de kun foreslår endringer for mindre deler av tilstanden. Dette gjør at algoritmen kan sette seg fast på lokale topper av $f(\phi, A|V, W)$ siden akseptansannsynligheten α vektet med forholdet mellom gammel og ny verdi av $f(\phi, A|V, W)$. For å kompensere for dette er det konstruert et endringsforslag som omfatter hele tilstanden. Det gjøres på følgende måte

1. Trekk forslag til ny stilling $\tilde{\phi}$ normalfordelt rundt eksisterende, med standardavvik $\sigma_{\phi 2}$, betydelig større enn σ_{ϕ} .
2. Trekk en tilfeldig rekkefølge $J = (j_1, j_2, \dots, j_{N_V})$ for komponentene i \tilde{A} .

3. Utfør følgende for alle komponentene i A , i rekkefølgen J . La \tilde{a}_i få verdien j med sannsynlighet $p_j^i = C \exp\left(\|w_j - r(v_i; \tilde{\phi})\|^2 / 2\sigma_w^2\right)$, om j ikke allerede har blitt tilordnet (ellers sett $p_j^i = 0$). Sett $\tilde{a}_i = 0$ med sannsynlighet $p_0^i = C \exp(L^2 / 2\sigma_w^2)$. Konstanten C er slik at $\sum_{j=0}^{N_W} p_j^i = 1$.

Forslagsfordelingen vil bli symmetrisk med hensyn på ϕ siden den trekkes normalfordelt, det gjør at det ikke trengs å ta hensyn til ϕ ved beregningen av forslagsfordelingen. Bidraget fra A til forslagsfordelingen blir $q(A, \tilde{A}) = \prod_{i=1}^{N_V} p_{\tilde{a}_i}^i$. Retursannsynligheten $q(\tilde{A}, A)$ kan beregnes ved å finne p_j^i etter samme rekkefølge og metode som for $q(A, \tilde{A})$.

4.3.3 Uttrykk for $f(\tilde{\phi}, \tilde{A}|V, W)/f(\phi, A|V, W)$

Akseptansannsynligheten er avhengig av forholdet $f(\tilde{\phi}, \tilde{A}|V, W)/f(\phi, A|V, W)$, og det er derfor nødvendig med uttrykk for dem. Uttrykkene må være på logaritmisk form for å kunne implementeres på en datamaskin. Det logaritmiske uttrykket for fordelingen det ønskes å trekkes fra ble gitt i likning (101), og er

$$\begin{aligned} \ln f(\phi, A|V, W) = & -\frac{1}{2\sigma_\phi^2} \phi^T \phi + N_0 \ln(1 - P_{rad}) - N_0 \ln P_{rad} + \ln(N_W - N_V + N_0)! \\ & + N_0 \ln 2\pi\sigma_v^2 - \sum_{i|a_i \neq 0} \frac{1}{2\sigma_v^2} \|w_{a_i} - r(v_i; \phi)\|^2 - N_0 \ln v_1 v_2 + \text{konstant}. \end{aligned} \quad (110)$$

I beregningen av akseptansannsynligheter er det behov for å se på forholdet mellom forslaget og eksisterende ϕ, A , det vil si differansen på logaritmisk skala. Hele uttrykket trengs når omfattende endringer gjøres. For de andre forslagene endres kun enkelte deler av ϕ eller A , noe som gjør at uttrykket kan forenkles for å spare regnetid.

Endring av ϕ . Anta at $A = \tilde{A}$, slik at kun ϕ endres på. Det betyr at alle ledd som kun avhenger av A og N_0 kan fjernes. Dette gir:

$$\begin{aligned} \ln f(\tilde{\phi}, A|V, W) - \ln f(\phi, A|V, W) = & \frac{1}{2\sigma_\phi^2} \phi^T \phi - \frac{1}{2\sigma_\phi^2} \tilde{\phi}^T \tilde{\phi} \\ & + \sum_{i|a_i \neq 0} \frac{1}{2\sigma_v^2} \|w_{a_i} - r(v_i; \phi)\|^2 - \sum_{i|a_i \neq 0} \frac{1}{2\sigma_v^2} \|w_{a_i} - r(v_i; \tilde{\phi})\|^2. \end{aligned} \quad (111)$$

Normleddene kan forenkles enda mer ved å skrive dem som indreprodukt og bruke vanlige produktregler.

Bytte av verdi for $a_i > 0$. Ved bytte blir N_0 og ϕ konstante, og alle ledd som bare avhenger av dem vil forkortes. La i være nummeret på komponenten som endres og at den endres fra j til j^f . ($a_i = j, \tilde{a}_i = j^f$)

$$\ln \frac{f(\phi, \tilde{A}|V, W)}{f(\phi, A|V, W)} = \frac{1}{2\sigma_v^2} \left(\|w_j - r(v_i; \phi)\|^2 - \|w_{j^f} - r(v_i; \phi)\|^2 \right). \quad (112)$$

Summene er blitt redusert til ett ledd siden de kun er forskjellige for den endrede komponenten.

Endre $a_i > 0$ til null. Her vil ϕ og alle untatt en komponent a_i være konstant. N_0 økes med 1 og a_i endres til null. Dette gir

$$\begin{aligned} \ln \frac{f(\phi, \tilde{A}|V, W)}{f(\phi, A|V, W)} &= \ln(1 - P_{rad}) - \ln P_{rad} + \ln(N_W - N_V + N_0 + 1) \\ &\quad + \ln 2\pi\sigma_v^2 + \frac{1}{2\sigma_v^2} \|w_{a_i} - r(v_i; \phi)\|^2 - \ln v_1 v_2. \end{aligned} \quad (113)$$

Fakultetleddene forenkles ved å bruke $\ln a - \ln b = \ln a/b$.

Endre $a_i = 0$ til verdi større enn null. Her vil ϕ og alle untatt en komponent a_i være konstant. N_0 senkes med 1 og en a_i endres fra null. Dette gir

$$\begin{aligned} \ln \frac{f(\phi, \tilde{A}|V, W)}{f(\phi, A|V, W)} &= -\ln(1 - P_{rad}) + \ln P_{rad} - \ln(N_W - N_V + N_0) \\ &\quad - \ln 2\pi\sigma_v^2 - \frac{1}{2\sigma_v^2} \|w_{a_i^f} - r(v_i; \phi)\|^2 + \ln v_1 v_2 \end{aligned} \quad (114)$$

5 Resultater

I dette kapitlet vil resultatene fra utførte simuleringer presenteres og kommenteres. Alle simuleringer er utført på samme datamaskin, beskrevet i vedlegg A. Kapitlet starter med å beskrive hvordan simuleringene er utført og hvordan parameterne er satt. Så vil de to LMK og MKK metodene fra kapittel 4.1 som opererer på kjente koblinger sammenlignes, noe som gir grunnlaget for hvorfor det ble valgt å bruke MKK som utgangspunkt for NN og GNN. Etter det vil feilmålene J^{NN} og J^{GNN} undersøkes nærmere for å finne en god minimeringsmetode. Denne vil brukes til å sammenligne NN- og GNN-metoden på samme måte som MKK og LMK, og GNN viser seg å være bedre. Det vil så undersøkes hvor godt GNN kan estimere ϕ i situasjonene fra figur 16. Til slutt vil påvirkningen av forskjellige parametere som deteksjonssannsynlighet og antall skip på estimeringsnøyaktigheten bli sett nærmere på.

5.1 Simuleringsmodell

For å teste estimeringsmetodene beskrevet foran er det utført simuleringer. Hovedgrunnen til dette er at det er veldig vanskelig eller umulig å finne de statistiske egenskapene til estimeringsmetodene analytisk, men dette kan estimeres ved gjentatte simuleringer. I tillegg kan simuleringer i bestemte tilfeller gi en god følelse av hvordan metodene fungerer, samtidig som det blir utarbeidet mye kode som kan gi grunnlag for programvare som trengs i en reell anvendelse. Det gir også et grunnlag til å si noe om hvor krevende metodene er med hensyn på kjøretid, selv om det ikke vektlegges i denne oppgaven. Simuleringene vil generere grunnlaget som estimeringsmetodene bruker til å estimere. Det vil si AIS-peilingene i V og radarpeilingene i W sammen med fasitstillingen ϕ^f og koblingene A^f . Ved hjelp av dette kan det beregnes estimater for ϕ^f og ved sammenligning med fasit kan prestasjonen vurderes. I oppgaven brukes to forskjellige metoder for å simulere de nødvendige dataene. En hvor alt genereres tilfeldig og en som delvis baserer seg på data mottatt fra FFI. Videre i dette kapitlet vil dataene fra FFI og hvordan informasjon hentes ut beskrives før det utarbeides et sett med standardparametere for simuleringene, og de to simuleringsmetodene blir beskrevet.

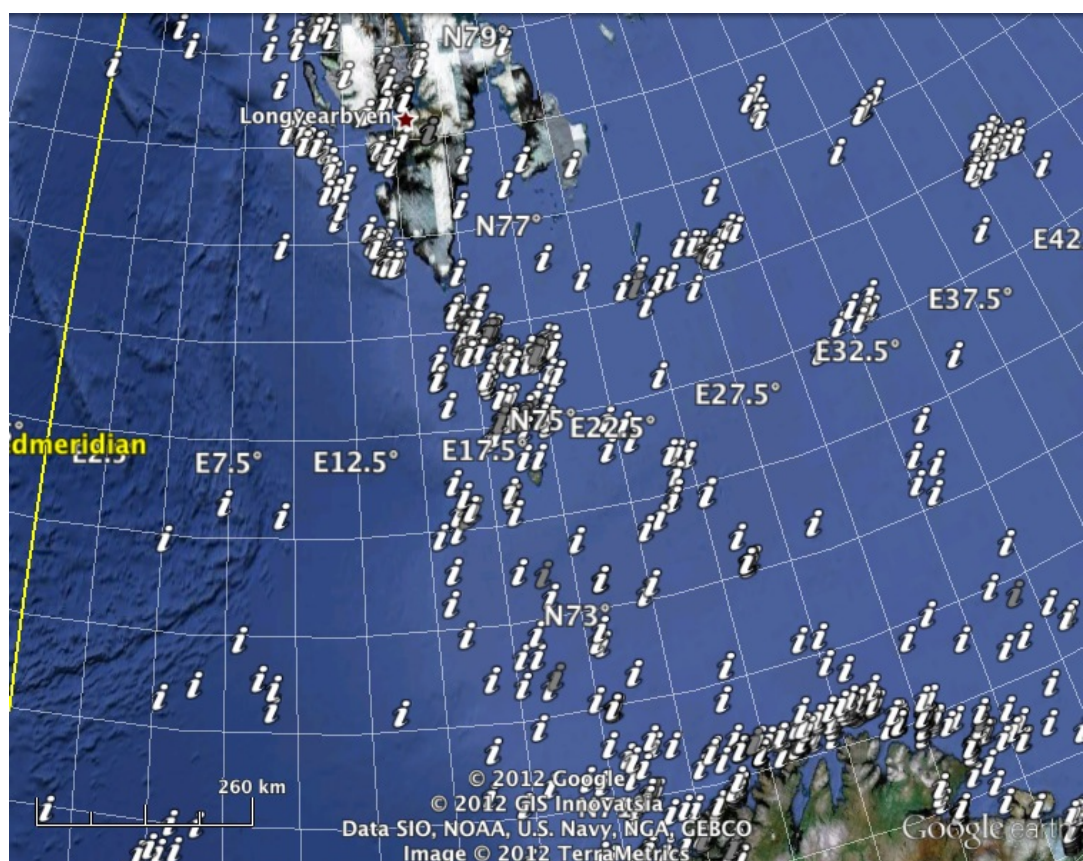
5.1.1 Bestemmelse av parametere fra realistiske data

FFI har driftet AIS-overvåkingssatellitten AISSAT-1 siden 2010, og gjennom den perioden samlet mye data. På grunnlag av dataene er det utført en simulering slik at resultatet kunne bli brukt i denne oppgaven. Denne simuleringen er basert på en modell som har en mer detaljert beskrivelse av utsendelse og mottak av

AIS-meldinger enn simuleringsmodellen i denne oppgaven. Den tar for eksempel hensyn til utsendt signalstyrke og atmosfæriske påvirkninger av signalet. I simuleringen er skipene fordelt på statiske posisjoner over jorden ut fra historisk posisjonsinformasjon som AISSAT-1 har mottatt via AIS. Det simuleres at hvert skip sender ut AIS-meldinger med jevne mellomrom og satellitten mottar dem om posisjon og forhold ligger til rette for det. Lengdene på mellomrommene mellom AIS-meldingene varierer fra skip til skip. Dataene som er mottatt er en liste over alle skipene, med tilhørende posisjon og spesifikasjoner av AIS-utstyret, en liste over når det er mottatt AIS-meldinger og hvilket skip meldingen er fra, samt banen satellitten fulgte i simuleringen. Banen er ikke brukt siden det ville krevd for mye tid å sette seg inn i koordinatsystemet brukt til å beskrive den. Resten av dette kapittelet vil bli brukt til å forklare hvordan disse dataene er behandlet. MATLAB (2012) er brukt til å lese og behandle dataene, og en del resultater er visualisert i Google Earth (2012) ved hjelp av Davis (2011). For denne oppgaven er det skipsfordelingen i nordområdene som er interessant, den er vist i figur 14. I visualiseringen skilles det mellom AIS-sendere av klasse A og klasse B. I følge Kystverket (2011) må alle skip som er pålagt å bære AIS-sender bruke klasse A, mens sendere av klasse B er beregnet for mindre fartøy og kan brukes av skip som ikke er pålagt å bære AIS-sendere. I dataene fra FFI skiller klassene seg fra hverandre ved at skip med sendere av klasse B har lavere signalstyrke og sender ut AIS-meldinger sjeldnere.

Dataene fra FFI vil bli brukt for to formål. Ut fra dem beregnes verdier som vil bli brukt til standardverdier for parametere som brukes i simuleringene, og i noen simuleringer brukes dataene direkte for å finne de mottatte AIS-peilingene, V . Metodene som brukes for de to formålene er de samme, og vil bli beskrevet i dette kapittelet. Hovedfokuset i dette kapittelet vil være å finne parameterverdier, den mer direkte anvendelsen i simuleringene behandles i kapittel 5.1.4. For å finne de nødvendige parameterene trengs det å finne hvilke skip som er innenfor ønsket område, og hvilke av disse satellitten mottar AIS-meldinger fra.

Området som det ønskes å finne skip inni er avgrenset av satellittens åpningsvinkel. Mer spesifikt kan en si at en ønsker å finne de skipene som uten støy vil gi opphav til en peiling som ligger innenfor åpningsvinkelen. For å avgjøre dette er både skipsposisjonene og åpningsvinkelen projisert i et plan og det er så undersøkt hvilke skip som er innenfor den projiserte åpningsvinkelen. Alle projeksjoner er utført ved sentralprojeksjon med sentrum i satellittens posisjon r_s . Det vil si at det finnes skjæringspunkt mellom en linje gjennom r_s og planet ved hjelp av metoden fra kapittel 2.3.3. Hvis et punkt skal projiseres passerer linjen også gjennom dette punktet, og hvis en peiling skal projiseres har linjen samme retning som peilingen representerer, gitt av likning (50). Projeksjonsplanet er valgt til å være normalt på sikteretningen til satellitten (z^{SI} -aksen) og passere gjennom skjæringspunktet mellom sikteretningen og jordoverflaten. Matematisk gjøres dette ved å finne retningsvektoren fra satellitten til skipet og bruke metoden i kapittel 2.3.3. Valget



Figur 14: Oversikt over skipsposisjoner mellom Finnmark og Spitsbergen. Hvite merker indikerer skip med sendere av klasse A og grå merker skip med sendere av klasse B. Bilde eksportert fra Google Earth.



Figur 15: Automatisk utvelging av skip i satellittens åpningsområde. Åpningsvinkel 10 grader i begge retninger, og projeksjonen er gjort ved metoden i kapittel 2.3.2 og er markert med gule linjer. Røde/grønne merker indikerer skip med klasse A/B sender innenfor åpningsvinkelen. Hvit linje viser satellittens sikteretning og senteret er hvor den treffer jordoverflaten. Bildet er eksportert fra Google Earth.

av dette planet gjør at den projiserte åpningsvinkelen tar form av et rektangel, og tre hjørner er tilstrekkelig for å definere området. For 10×10 graders åpningsvinkel brukes peilingene $(5, -5)$, $(-5, -5)$ og $(-5, 5)$. Metoden fra kapittel 2.3.4 brukes så til å finne ut om et skip er innenfor åpningsvinkelen. Resultatet av en slik prosess er vist i figur 15. Gjennom å telle opp hvor mange skip som er innenfor kan det finnes aktuelle verdier for antall skip innenfor åpningsvinkelen, N_s .

For å finne ut hvilke skip det er mottatt AIS-meldinger fra er det sett på tidspunktene meldingene er mottatt. Hvis en ser på tidsperiodene det mottas AIS-meldinger fra skip i et utvalgt område er det intervaller på under 20 minutter hvor det mottas meldinger, adskilt av pauser på over en time. Den beste forklaringen på dette skyldes at satellitten passerer på baksiden av jorden i forhold til området. Tidspunktene gir en rekke intervaller som dermed representerer at satellitten passerer over området, første og siste intervall forkastes for å sikre komplette passeringer. Ut fra hvilke skip det mottas AIS-meldinger fra i løpet av en passering kan verdier for sannsynligheten for å motta AIS-peilinger, P_{AIS} , finnes.

Gjennom diskusjon med FFI er det blitt bestemt at det er hensiktsmessig å se nærmere på tre forskjellige situasjoner:

1. En hvor det er relativt tett med skip.
2. En hvor det er relativt få skip.

P_{AIS}, N_s	Situasjon 1	Situasjon 2	Situasjon 3
10×10 grader	90%, 52	97%, 9	36%, 46
10×40 grader	90%, 89	88%, 21	52%, 147

Tabell 4: Empirisk deteksjonssannsynlighet (P_{AIS}) og antall skip (N_s) for alle situasjonene vist i figur 16. Verdiene er beregnet som gjennomsnitt av verdiene fra passering 5 til 8.

3. En langs kysten.

For alle de tre situasjonene er det aktuelt å se på liten (10×10 grader) og stor (10×40 grader) åpningsvinkel. Det er funnet situasjoner som tilfredsstillende hver av betingelsene 1, 2 og 3 over. En oversikt over dem er vist i figur 16 og nærbilder er vist i figur 17. Det er beregnet deteksjonssannsynligheter for AIS i alle situasjonene, og for alle passeringene. Et typisk forløp, hentet fra situasjon 1 med liten åpningsvinkel, er vist i figur 18. Tilsvarende grafer for de resterende situasjonene og åpningsvinkler kan finnes i vedlegg C. Fra figur 18 er det tydelig at deteksjonssannsynligheten går i en bue med maksimum i midten. Sammen med FFI er det kommet frem til at den mest sannsynlige forklaringen for dette er at satellitten i disse tilfellene har passert i kanten av området, slik at det er vanskeligere å motta meldinger fra en del skip. For å kompensere for dette beregnes gjennomsnittlig deteksjonssannsynlighet kun for passering 5–8. Resultatet av denne beregningen vises sammen med antall skip innenfor åpningsvinkelen i tabell 4. Tabell 4 viser at deteksjonssannsynligheten i situasjon 3 nær kysten er omtrent halvparten av deteksjonssannsynligheten i de andre situasjonene. Dette er sannsynligvis relatert til at forskriftene krever hyppigere utsendte AIS-meldinger fra skip i bevegelse enn fra skip som ligger i ro.

5.1.2 Parameterverdier som brukes til simulering

For å utføre en simulering må det settes tallverdier på en mengde parametere. I dette kapitlet vil det etableres standardverdier som blir brukt hvis ikke noe annet er spesifisert. Bakgrunnen for hvorfor standardene er valgt vil også bli forklart.

Satellittens posisjon r_s er satt til 600km over jordens overflate, fordi det i følge FFI er en typisk høyde for satellitter av denne typen. Standard stilling for FSI er satt lik LH, det vil si at $\Psi = (0, 0, 0)$. Det er valgt fordi det tilsvarer at antennen ser rett ned, noe som er lettere å jobbe med i en del tilfeller. Fasitstillingen ϕ^f skal være ukjent og må trekkes for hver simulering. Den trekkes fra en normalfordeling med forventningsverdi 0 og kovariansmatrise $\sigma_\phi^2 I$. Det er to tilfeller som kan bli lagt til grunn for standardverdien til σ_ϕ . Drift i utstyret som holder orden på stillingen til satellitten skaper økt feil i løpet av en runde rundt jorden, en feil som er anslått til



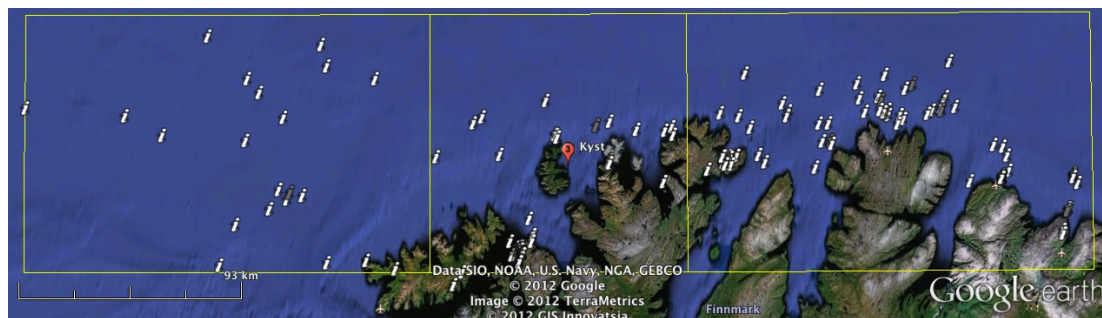
Figur 16: Oversikt over de tre situasjonene. Indre kvadrat representerer åpningsvinkel på 10×10 grader, ytre rektangel 10×40 grader. I alle situasjonene er $\Psi_x = \Psi_y = 0$ som tilsvarer at sikteretningen til satellitten er loddrett ned mot jordens sentrum. For situasjon 1 er Ψ_z er satt til -45 grader, ellers er $\Psi_z = 0$. Bildet er eksportert fra Google Earth.



(a) Situasjon 1 - tett.

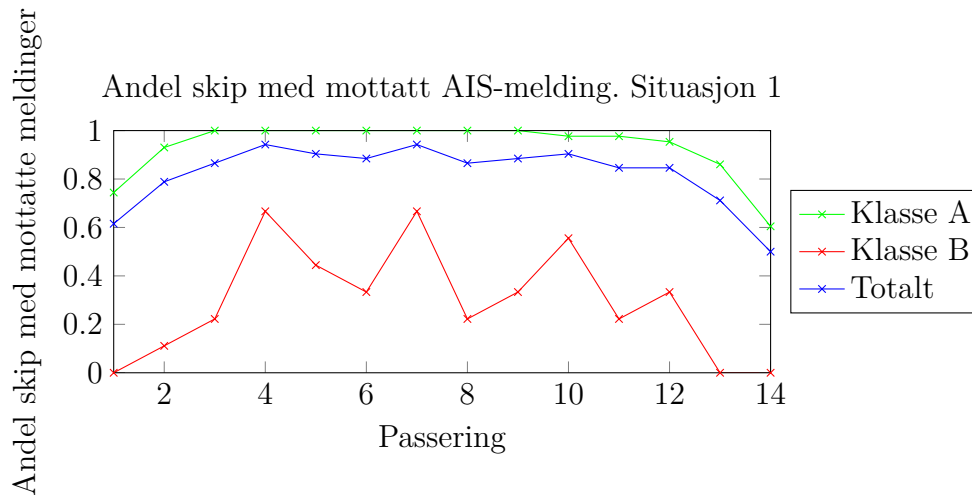


(b) Situasjon 2 - spredt.



(c) Situasjon 3 - kyst.

Figur 17: Oversikt over skipsfordeling i situasjonene. Bilder eksportert fra Google Earth.

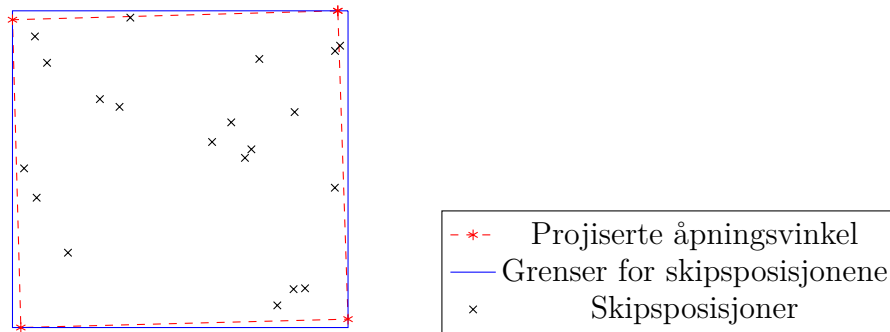


Figur 18: Empirisk deteksjonssannsynlighet etter passering. Klasse A/B er beregnet bare med skipene som har AIS-sender av klasse A/B. Totalt er beregnet med alle skipene i overvåkingsområdet som grunnlag.

å ligge på rundt 2 grader. Det vil også være en initiell usikkerhet ved oppstart som anslås å kunne være opp til 10 grader. Den lave verdien er valgt til standard siden det vil være den verdien som er aktuell oftest, og det gjør at estimeringsalgoritmene for NN og GNN kan kjøres raskere. Standardverdien for σ_ϕ er satt til 1 grad, siden det da er 60% sannsynlig at $\|\phi\|$ er mellom 1 og 3 grader. Feilen i radarpeilingene, ϵ , trekkes fra en normalfordeling med forventningsverdi 0 og kovariansmatrise $\sigma_w^2 I$. Verdien for σ_w er satt til 0.1 grader som standard, på grunnlag av nøyaktigheten som ble oppnådd i eksperimenter på FFI sommeren 2011. Teknikkene som ble brukt der er sammenlignbare med de som tenkes brukt i satellitten, og resultatene regnes derfor som et realistisk forslag for σ_w . Deteksjonssannsynligheten for radar P_{rad} er satt til 90%, ut fra hva FFI forventer at er realistisk. Standard for antall skip N_s og deteksjonssannsynlighet for AIS P_{AIS} er satt ut fra verdiene til situasjon 1 med liten åpningsvinkel, som er funnet i kapittel 5.1.1. Det gir verdiene $N_s = 50$ og $P_{AIS} = 0.9$. Situasjon 1 ble valgt siden den har et middels antall skip og den vanligste deteksjonssannsynligheten. Dette gjør også at 10×10 grader er det mest naturlige valget for standard åpningsvinkelen ν Vindustørrelsen L ble i likning (86) satt til $3\sigma_w$, som også blir brukt som standard.

5.1.3 Tilfeldig simulering

Det er som nevnt brukt to typer simuleringer, og den første som blir forklart her kalles den tilfeldige simuleringen. Simuleringen kalles tilfeldig siden alle dataene blir generert ved hjelp av generatoren for tilfeldige tall i datamaskinen, i motsetning til den andre varianten som i tillegg bruker dataene fra FFI. På



Figur 19: *Plott over projiserte hjørner av åpningsvinkelen, korresponderende område for å plassere skip og eksempel på skipsposisjoner.*

grunn av enklere implementering er det for denne simuleringen brukt en flat jord tilnærming. Det gjør at skipene er fordelt ut over et område i xy -planet i SJ, som tilsvarer jordoverflaten. Området har sentrum i origo og satellitten er plassert rett over. Ut fra standardverdiene i kapittel 5.1.2 gir det at r_s har en z -komponent på 600km mens de andre komponentene er null. LH defineres ved at satellittens fartsretning tenkes å være i SJ sin y -retning, mens jordens sentrum tenkes å være i origo. Dette gir at $\psi = (0, 180, 90)$ grader som representerer at z -aksen snus og x - og y -aksene bytter plass. Det vil si at z^{LH} peker i retningen til $-z^{SJ}$, x^{LH} i retningen til y^{SJ} og y^{LH} i samme retning som x^{SJ} . Sammen med parameterne fra kapittel 5.1.2 gir dette grunnlaget for simuleringen, som så utføres på følgende måte.

Simuleringen startes med å trekke stillingen ϕ^f fra en normalfordeling med kovarians σ_ϕ og forventningsverdi 0. Hjørnepunktene til åpningsvinkelen, peilingene $(v_1/2, v_2/2)$, $(-v_1/2, v_2/2)$, $(v_1/2, -v_2/2)$ og $(-v_1/2, -v_2/2)$, projiseres til jordoverflaten ved å konvertere dem til retninger i SI, som så uttrykkes i SJ og skjæringspunktene mellom retningene og xy -planet finnes ved metoden fra kapittel 2.3.3. Det minste rektangelet i xy -planet som inneholder disse punktene og har akser parallelle med koordinataksene brukes til representere åpningsvinkelen, og N_s skip blir fordelt uniformt over det. Figur 19 viser et eksempel for hvordan dette kan se ut i xy -planet. Ut fra hver av disse posisjonene, satellittposisjonen r_s og stillingsvinklene ψ, Ψ og ϕ^f beregnes potensielle AIS- og radarpeilinger ved fremgangsmåten fra kapittel 3.2. Det trekkes så hvilke skip det mottas AIS- og/eller radarpeilinger fra på en slik måte at et bestemt skip har en sannsynlighet på P_{AIS} for å ha en AIS-peiling og P_{rad} på å ha en radarpeiling. Om et skip mottar AIS-peiling er uavhengig fra om det mottar en radarpeiling. AIS-peilingene som blir mottatt samles i V , og mottatte radarpeilinger samles i W . De potensielle AIS-peilingene som trekningen avgjør at blir mottatt samles i V , og tilsvarende radarpeilinger samles i W . Samtidig beregnes koblingene A^f ved å holde orden på hvilke peilinger som kommer fra samme skip. Etter dette er gjort er alle

nødvendige variable klare til bruk i estimering.

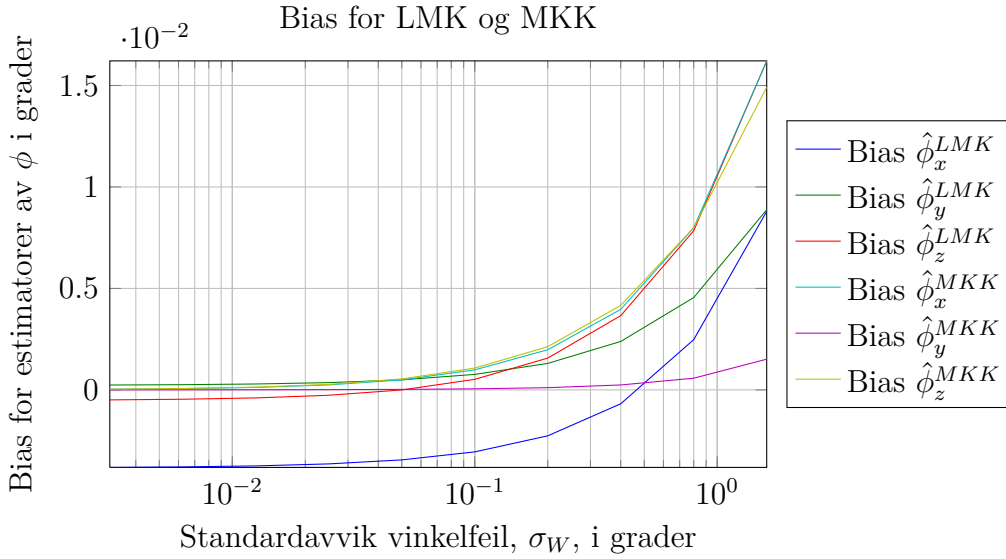
5.1.4 Databasert simulering

Den andre simuleringsvarianten bruker så mye av dataene fra FFI som mulig, og kalles derfor den databaserte simuleringen. Denne simuleringen er basert på en kulemodell av jorden. Satellittposisjonen r_s ligger som standard alltid 600km over overflaten, men hvilket punkt som er rett nedenfor varierer. LH er definert ved at satellitten tenkes å bevege seg i en kulebane med retning rett sørover. Det gjør at LH-koordinatsystemet (og FSI siden $\Psi = 0$) er slik at z -aksen peker mot jordens sentrum og y -aksen følger meridianene. Ved hjelp av dette og metoden fra kapittel 5.1.1 hentes det ut hvilke skip som er inni åpningsvinkelen. Posisjonene uttrykkes i SJ og brukes til å finne retningsvektorer i SI og FSI ved metodene i kapittel 3.2. Radarpeilingene som mottas velges ut og lagres i W på samme måte som for den tilfeldige simuleringen. De potensielle AIS-peilingene beregnes også på samme måte som for den tilfeldige simuleringen, men hvilke som mottas velges ut fra hvilke skip det ble mottatt AIS-meldiger fra i passering nummer 6. Denne utvelgelsen foretas ved metoden beskrevet i kapittel 5.1.1. Dette gir V og ved å holde orden på hvilke peilinger som stammer fra hvilke skip finnes A^f , og simuleringen av de nødvendige variablene er ferdig.

5.2 Sammenligning av MKK og LMK

I dette kapittelet vil til minste kvadraters estimatene ved kjent assosiasjon undersøkes. Siden koblingene er kjente er deteksjonssannsynlighetene uten betydning. Derfor er det i simuleringene latt $P_{AIS} = P_{radar} = 1$ slik at antall skip med både radarpeiling og AIS-peiling holdes konstant. Det første som blir sett nærmere på er statistiske egenskaper og hvordan de varierer med feilen i radarpeilinger σ_W . Testen er utført ved å la alle parametere være som standard, unntatt σ_W som varieres. For hver verdi av σ_W er CR nedre grense beregnet ved fremgangsmåten fra kapittel 2.8. Radarpeilingene w_i tilsvarer z_i , ϕ tilsvarer θ og $r(v_i, \phi)$ tar rollen som $h_i(\phi)$. I tillegg er det trekt skipsposisjoner og beregnet peilinger fra dem 500 ganger ved den tilfeldige simuleringen beskrevet i kapittel 5.1.3. Hver gang er differansen mellom de to estimatorene ($\hat{\phi}^{MKK}$ og $\hat{\phi}^{LMK}$) og faktisk stilling ϕ^f beregnet og lagret. På grunnlag av dette estimeres de statistiske egenskapene. Estimat for forventningsskjevheten regnes ut som gjennomsnittlig differanse mellom fasit og estimat, Bias $\hat{\phi} = \frac{1}{n} \sum_{i=1}^n \hat{\phi} - \phi^f$. Standardavviket estimeres komponentvis ved

$$\text{Std } \hat{\phi}_x = \sqrt{\frac{1}{n-1} \left(\sum_{i=1}^n (\hat{\phi}_x - \phi_x^f)^2 - \left(\sum_{i=1}^n (\hat{\phi}_x - \phi_x^f) \right)^2 \right)}, \quad (115)$$



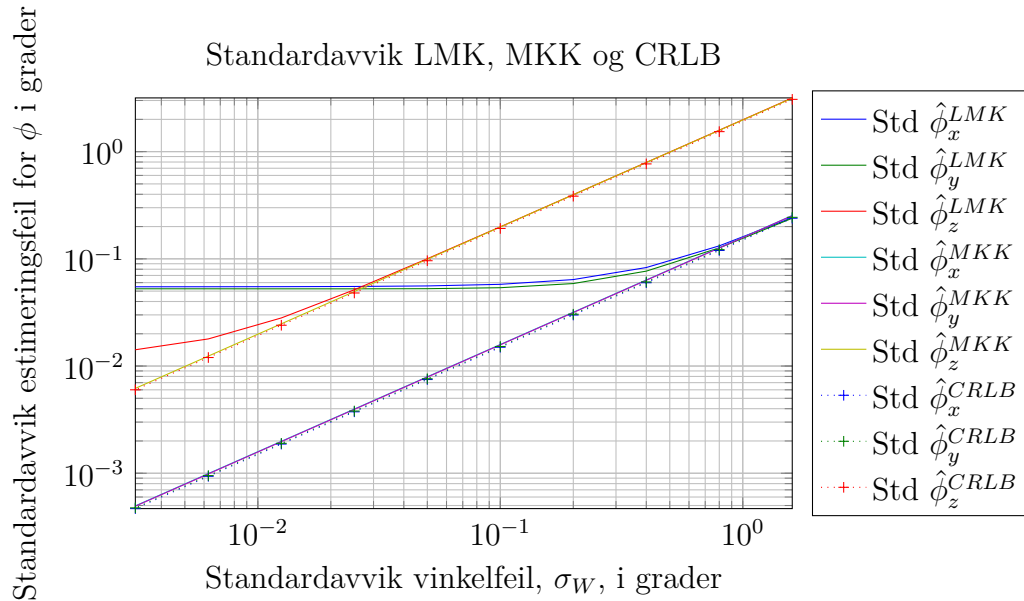
Figur 20: Sammenligning av estimert forventningsskjevhet for MKK og LMK etter målestøy. σ_w antar verdiene $0.1 \cdot 2^n$ grader for $n = -5, -4, \dots, 4$.

eller samlet ved å ta kvadratroten av diagonalelementene fra kovariansmatrisen som beregnes ved

$$\text{Kov } \hat{\phi} = \frac{1}{n-1} \sum_{i=1}^n \left((\hat{\phi} - \phi^f) - \frac{1}{n} \sum_{i=1}^n (\hat{\phi} - \phi^f) \right) \left((\hat{\phi} - \phi^f) - \frac{1}{n} \sum_{i=1}^n (\hat{\phi} - \phi^f) \right)^T. \quad (116)$$

De to metodene er ekvivalente.

Resultatene for estimert forventningsskjevhet og standardavvik vises i figur 20 og 21. Estimert forventningsskjevhet ligger alltid mer enn en faktor ti under standardavviket. Dette betyr at eventuell forventningsskjevhet har liten betydning for totalfeilen i forhold til standardavviket. Fra figur 21 er det tydelig at begge estimatorene presterer like bra med hensyn på standardavvik frem til et punkt hvor den lineariserte versjonen stagnerer. Det regnes som mest sannsynlig at dette skyldes at feilen som skyldes linearisering blir større enn feilen som skyldes målestøy. Utflatingen inntreffer for en høyere verdi av σ_w enn standardverdien på 0.1 grader, noe som gjør at MKK er bedre i aktuelle områder. MKK er derfor lagt til grunn for metodene som utvikles i de neste kapitlene. Det er også viktig å merke seg at standardavvikene for MKK ligger nært CR grensene, altså er $\hat{\phi}^{MKK}$ optimal eller nær optimal. Estimerte standardavvik for $\hat{\phi}_x^{MKK}$ og $\hat{\phi}_y^{MKK}$ ligger en faktor 6.3 under σ_w , mens det tilsvarende for CR er 6.7 for x -komponenten og 6.6 for y . For z -komponenten er estimert standardavvik henholdsvis en faktor 2.0 og 1.9 større enn σ_w for MKK og CR. Dette tilsvarer at $\text{Std } \hat{\phi}_x^{MKK}$ er 6.1% større enn $\text{Std } \hat{\phi}_x^{CR}$, og tilsvarende 3.6% og 3.0% større for y og z komponentene. Disse forholdstallene er beregnet ved å finne forholdet mellom de



Figur 21: Sammenligning av estimert standardavvik for MKK og LMK etter målestøy. σ_w antar verdiene $0.1 \cdot 2^n$ grader for $n = -5, -4, \dots, 4$. x og y komponentene ligger langs nederste linje og er vanskelig å skille.

aktuelle verdiene for alle verdiene av σ_w og så ta gjennomsnittet. Etter å ha sett taylorutviklingen i likning (136) er det ikke overraskende at ϕ_z er vanskeligere å estimere, siden den påvirker peilingene i mindre grad. Peilingene påvirkes direkte av ϕ_x og ϕ_y , mens ϕ_z ganges sammen med en peiling som har en maksimal størrelse på 0.09 ved 10 graders åpningsvinkel. Størrelsene på ϕ_x og ϕ_y er typisk enda mindre enn dette igjen, og gir dermed enda mindre påvirkning. Området som standardavviket for x - og y -komponentene ligger i er delvis forventede siden de faktorene er nær $7.1 = \sqrt{50}$, og standardavviket for mange enkle estimatører faller med roten av antall målinger. Det er mulig at størrelsen til standardavviket til z -komponenten i forhold til de andre komponentene kan kobles direkte til størrelsen på peilevinklene. Forholdet mellom det estimerte standardavviket til x - eller y -komponenten og z -komponenten er 0.08, mens maksimalverdien til peilingene ved denne åpningsvinkelen er 0.09.

5.3 Undersøkelse av $J^{GNN}(\phi)$ og $J^{NN}(\phi)$

I kapittel 4.2 ble det forklart hvilke problemet vinduet skaper for direkte bruk av `fminsearch` for å finne globalt minimum. For å finne ut hva som kan fungere er det sett nærmere på formen til feilfunksjonene ved å plote $J^{GNN}(\phi)$ og $J^{NN}(\phi)$ over forskjellige områder. Siden ϕ har tre komponenter blir det fullstendige plottet firedimensjonalt. For å komme rundt dette er en av komponentene holdt fast

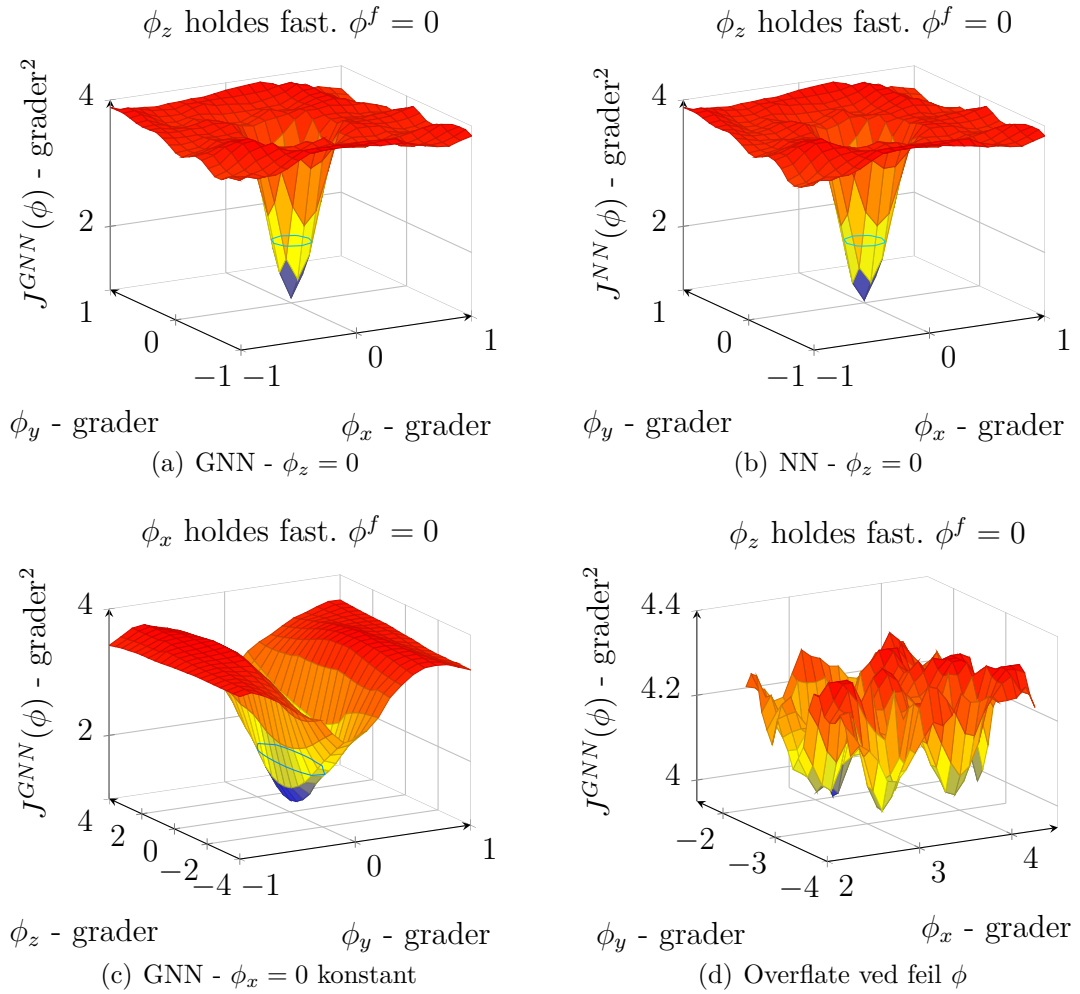
mens de andre varierer. Eksempler basert på tilfeldig simulering er vist i figur 22. Figur 22(a) viser $J^{GNN}(\phi)$ når $\phi_z = 0$, figur 22(b) viser det samme med $J^{NN}(\phi)$ istedet for $J^{GNN}(\phi)$, og figur 22(c) viser $J^{GNN}(\phi)$ når ϕ_x holdes fast. Figur 22(d) viser et nærbilde av hvordan overflaten det kuperte området utenfor søkket rundt ϕ^f i figur 22(a) ser ut. Figur 22(a) og 22(b) viser at J^{GNN} og J^{NN} har forskjeller på detaljnivå, men grovt sett er de like. Denne trenden holder seg for de andre plottene, og derfor er det ikke tatt med plott tilsvarende figur 22(c) og 22(d) for NN. Plottet for GNN hvor ϕ_y holdes fast er også utelatt siden det er likt figur 22(c). Til sammen viser figurene at J^{GNN} og J^{NN} har dypt søkk (minimumsverdi rundt 1) rundt ϕ^f og at feilmålet har verdier rundt 4 utenfor dette søkket. Dette betyr at søkket rundt ϕ^f skiller seg ut fra alle de andre små søkkene med tilhørende lokale minimum. At det finnes så mange lokale minimum gjør at direkte bruk av `fminsearch` ikke vil fungere. Men hvis det er mulig å finne et punkt i søkket som inneholder ϕ^f vil `fminsearch` kunne brukes til å finne globalt minimum ved å starte i dette punktet. Dette reduserer problemet til å finne et punkt i søkket, et problem som er løst på 3 måter. Erfaringsmessig har GNN en kjøretid på 5–10 ganger den som NN har. For eksempel tok det 3.5 sekunder å beregne dataene til figur 22(a), men kun 0.5 sekunder for dataene til figur 22(b). Siden søkket er på samme sted uansett vil NN brukes for å finne søkket, så kan finsøket etterpå eventuelt bruke GNN.

5.3.1 Enkelt rutenettsøk

En enkel idé for å finne et punkt i søkket er å lage et rutenett som er stort nok til at det nesten helt sikkert inneholder ϕ^f og er så tett at det er minst ett punkt i søkket. Ved å beregne $J^{NN}(\phi)$ i alle punktene kan søkket finnes ved minimumsverdien. Masketettheten kan for eksempel bestemmes ved å se på hvor bredt søkket er i typiske tilfeller. Konturlinjene i figur 22 som har høyde på halvparten av maksimalverdien er en mulighet til å gjøre dette. Ut fra dem bør tetthet på 0.2, 0.2 og 2 grader bør fungere for henholdsvis ϕ_x , ϕ_y og ϕ_z . Hvor bredt søkket skal være avgjøres av σ_ϕ , siden den avgjør spredningen til ϕ^f . Ved hjelp av likning (28) kan det finnes en størrelse som gjør det tilstrekkelig sannsynlig at ϕ^f er innenfor.

5.3.2 Smart rutenettsøk

Rutenettsøket beskrevet over vil fungere, men det er flere egenskaper ved problemet som kan utnyttes for å gjøre den mer effektiv. Det viktigste er at søkkets dybde er unik, slik at om en finner et punkt hvor J^{NN} har lav nok verdi kan en stoppe der og la være å sjekke resten av rutenettet. Stillingen er også normalfordelt rundt 0, noe som gjør at sannsynlighetstettheten synker når $\|\phi\|$ stiger. Altså er



Figur 22: Plott over oppførselen til J^{GNN} og J^{NN} . Konturlinje angir halvparten av maksimal verdi for feilmålet, diameter er omtrent 0.4 grader i x - og y -retning og 4 grader i z retning. Merk at aksene er forskjellige.

det smartest å søke for små $\|\phi\|$ først. For å forklare hvordan dette er gjort la rutenettet bestå av alle punkter som har heltall som koordinater. Det er så mulig å starte i $\phi = (0, 0, 0)$ og jobbe seg utover ved å først ta alle koordinater hvor $|\phi_x| + |\phi_y| + |\phi_z| = 1$, så med summen lik 2 og så videre oppover. Ved å se bort fra alle koordinatene hvor ϕ_z ikke er delelig med 10 fås et rutenett som har ti ganger mindre tetthet i z -retning, og ved å gange alle resterende punkt med 0.2 oppnås tettheten fra forrige kapittel.

Dette er implementert ved å la ϕ' løpe over alle rutenettspunkt med positive koordinater, og for hver av dem la ϕ løpe over alle mulige fortegnsvarianter. Algoritmen som bestemmer ϕ' er som følgende, når $n = |\phi_x| + |\phi_y| + |\phi_z|$: Start med $\phi' = (0, 0, 0)$, og bestem det neste punktet $\tilde{\phi}'$ etter følgende regler

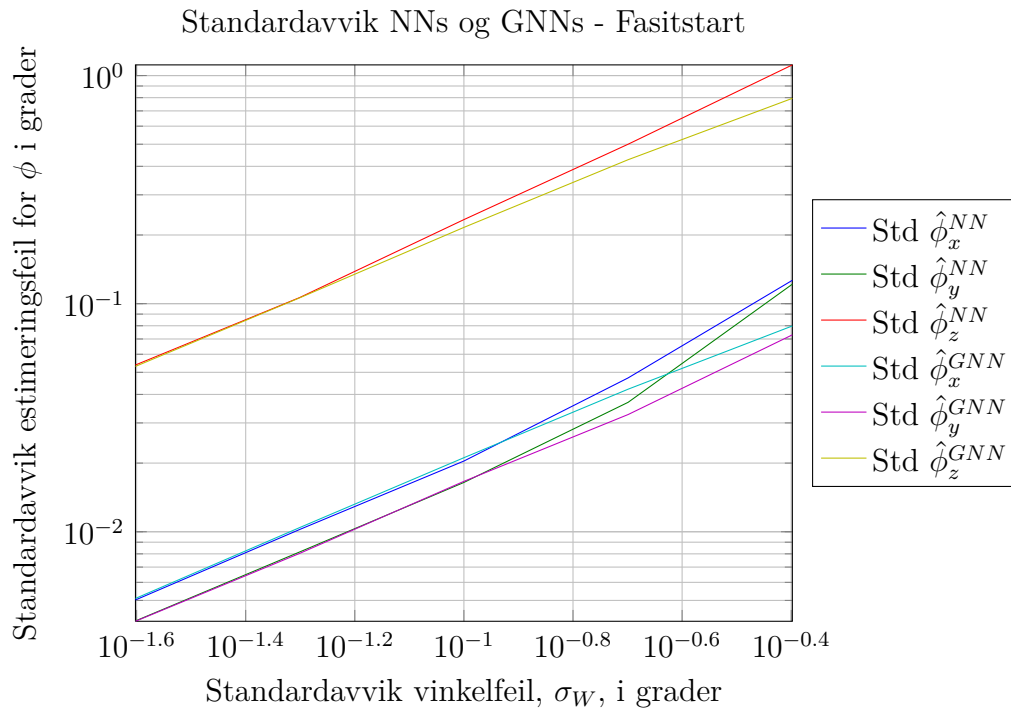
- Hvis $\phi'_z = n$ sett $\tilde{\phi}' = (n + 1, 0, 0)$.
- Hvis $\phi'_y = 0$ og $\phi'_z < n$ la $\tilde{\phi}' = (\phi'_x - 1, n - \phi'_x, 0)$.
- Hvis $\phi'_y > 0$ la $\tilde{\phi}' = (\phi'_x, \phi'_y - 1, \phi'_z + 1)$.

5.3.3 Start i fasit

Siden søkket skiller seg ut såpass mye, vil et godt rutenettsøk finne den. Dette dreier seg om simuleringer, så det finnes et tilgjengelig punkt inni søkket, nemlig fasitstillingen ϕ^f . Resultatet ved et søk som starter i ϕ^f vil være det samme som ethvert annet punkt, siden søkket kun har ett minimum. Derfor kan estimatet bestemmes ved å starte i fasit istedet for ved et krevende rutenettsøk.

5.3.4 Fordeler, ulemper og valg av søkemetode

Alle metodene skal gi samme resultat, så lenge formen på J^{NN} holder seg, så andre faktorer må avgjøre valget. Å starte i fasit er naturligvis ikke mulig i en reell situasjon, men det er ikke en viktig innvending her siden spørsmålet er hvor godt en estimator kan fungere. Kjøretiden til programmet er ikke med i resultatene for denne oppgaven, men påvirker hvor praktisk det er å utføre simuleringene. I estimeringen er det beregningen av J^{NN} og/eller J^{GNN} som tar tid, derfor er den viktigste faktoren for kjøretiden antall beregninger av feilmålene som kreves. For dette er rutenettsøket klart dårligst, den smarte varianten en del bedre og start i fasit klart best. En annen faktor er hvordan figur 22 vil se ut for andre parametere. Endring av vindustørrelse, σ_w eller annet kan tenkes å skape mindre forskjell mellom globalt minimum og resten, da vil det være vanskeligere å velge verdien som gjør at det smarte rutenettsøket stopper. Det kan også bli behov for tettere rutenett, som gir lengre kjøretid. Fasitstart vil finne det globale minimumet så lenge det er i nærheten av den riktige stillingen ϕ^f . Hvis det ikke er slik er det



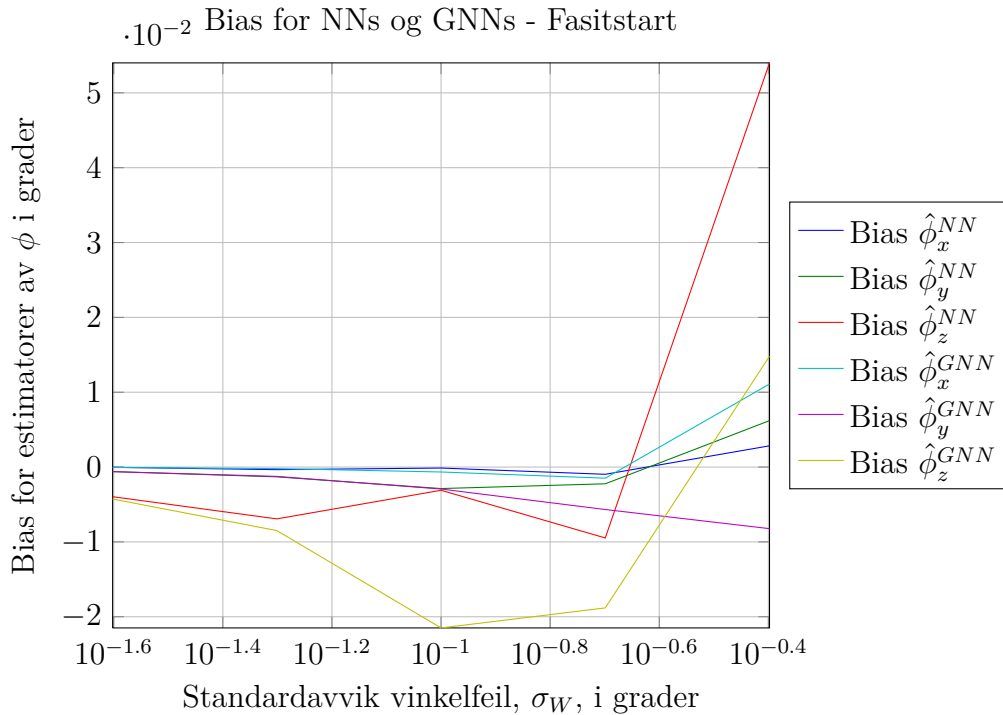
Figur 23: Empirisk bestemte standardavvik for NN og GNN.

uansett ikke håp uansett metode. Derfor er metoden med å starte i fasit den beste i mitt tilfelle, og vil bli brukt oftest. Resultatet av estimeringen ved hjelp av NN og GNN kan uansett ikke bli bedre enn dette. For å sjekke at resultatene blir like ved bruk av fasitstart og rutenettsøkene, vil det i noen tilfeller bli brukt både smart rutenettsøk og fasitstart.

5.4 Sammenligning av NN og GNN

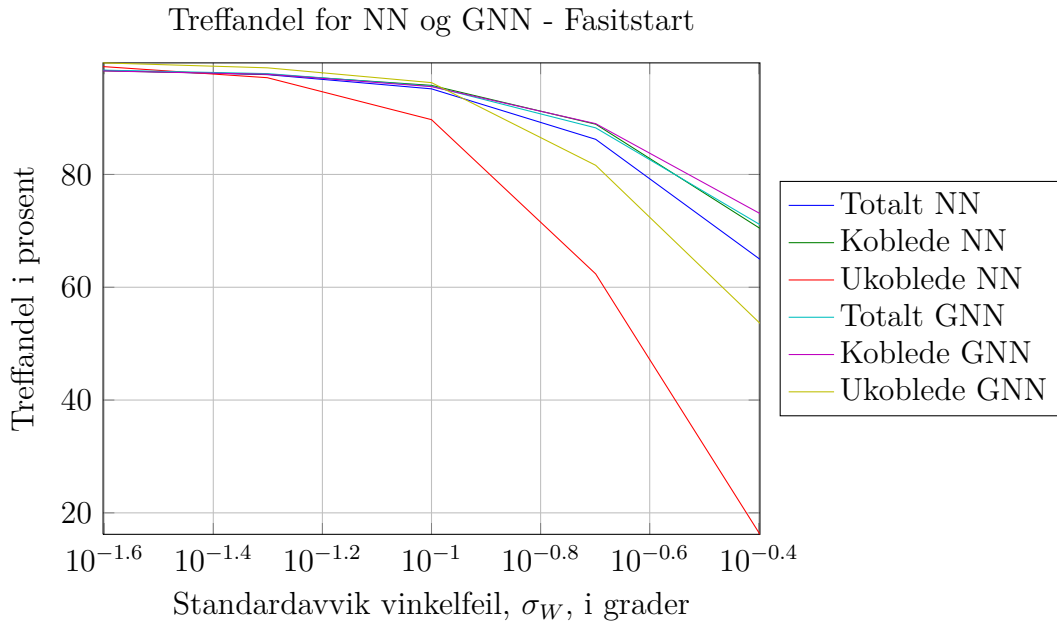
Det er to aktuelle estimatorer når koblingene er ukjente, NN og GNN. For å avgjøre hvilken som presterer best er det gjort tilsvarende simulering som i kapittel 5.2. For $\sigma_w = 0.4, 0.2, \dots, 0.025$ grader er det simulert AIS- og radarpeilinger 100 ganger. Ut fra peilingene er ϕ estimert ved å minimere $J^{GNN}(\phi)$ og $J^{NN}(\phi)$ ved å bruke fasitstartmetoden fra kapittel 5.3.3. Standardavvik og forventningsskjevhet er beregnet på samme måte som i kapittel 5.2, og er vist i figur 23 og 24.

Estimert forventningsskjevhet er jevnt over minst en faktor ti mindre enn estimert standardavvik. Altså ser det ut til at estimatorene er forventningsrette eller svært nær forventningsrette. De estimerte standardavvikene for GNN-estimatorene er lavere enn de estimerte standardavvikene for NN ved høye σ_w , og like for de lave σ_w . Dette er rimelig siden NN og GNN vil gjøre samme assosiasjonene om det er



Figur 24: Empirisk bestemte forventningskjevheter for NN og GNN.

en eller ingen radarpeiling innenfor vinduet til en AIS-peiling, mens forskjellene vil gjøre seg gjeldende ved flere innenfor. Siden tettheten av skip er den samme hele tiden vil økende σ_w , og dermed økende vindu, gjøre at flere skip vil havne innenfor. Opptelling etter en simulering gir at det i gjennomsnitt er 0.96 peilinger innenfor vinduet når σ_w er 0.025 grader mot 2.4 peilinger når σ_w er 0.4 grader. Det er også en trend for at ϕ_y estimeres bedre enn ϕ_x for begge metodene, selv om trenden ser ut til å gi seg når σ_w nærmer seg 0.4 grader. En forklaring på dette kan være rotasjonsrekkefølgen. Siden rotasjonen om y -aksen utføres før rotasjonene om x -aksen vil ϕ_y ha en liten ekstra effekt på stillingen, og dermed også peilingene. Dette holdes som den mest lovende forklaringen, men har ikke prioritert å teste den grundigere. En annen ting som er interessant er hvor stor andel av assosiasjonene som er korrekte. Dette kalles treffandelen og er beregnet totalt, for koblete AIS-peilinger og for ukoblete AIS-peilinger. Figur 25 viser resultatet fra simuleringene. Den totale treffandelen er beregnet ved å ta totalt antall riktige assosiasjoner over de 100 simuleringene og dele det på totalt antall AIS-peilinger. For de andre treffandelene er det telt opp hvor mange AIS-peilinger som hadde koblete radarpeilinger og hvor mange som ikke hadde det, for så å se på andelen av dem som er blitt korrekt assosiert. Det tydeligste trendene her er at GNN ligger over NN for alle typer, og at treffprosenten for ukoblete AIS-peilingen synker veldig raskt for NN. Det skyldes at NN ikke hindrer at en radarpeiling kobles til flere AIS-peilinger, og økende vindu gjør det mye mer sannsynlig å få en

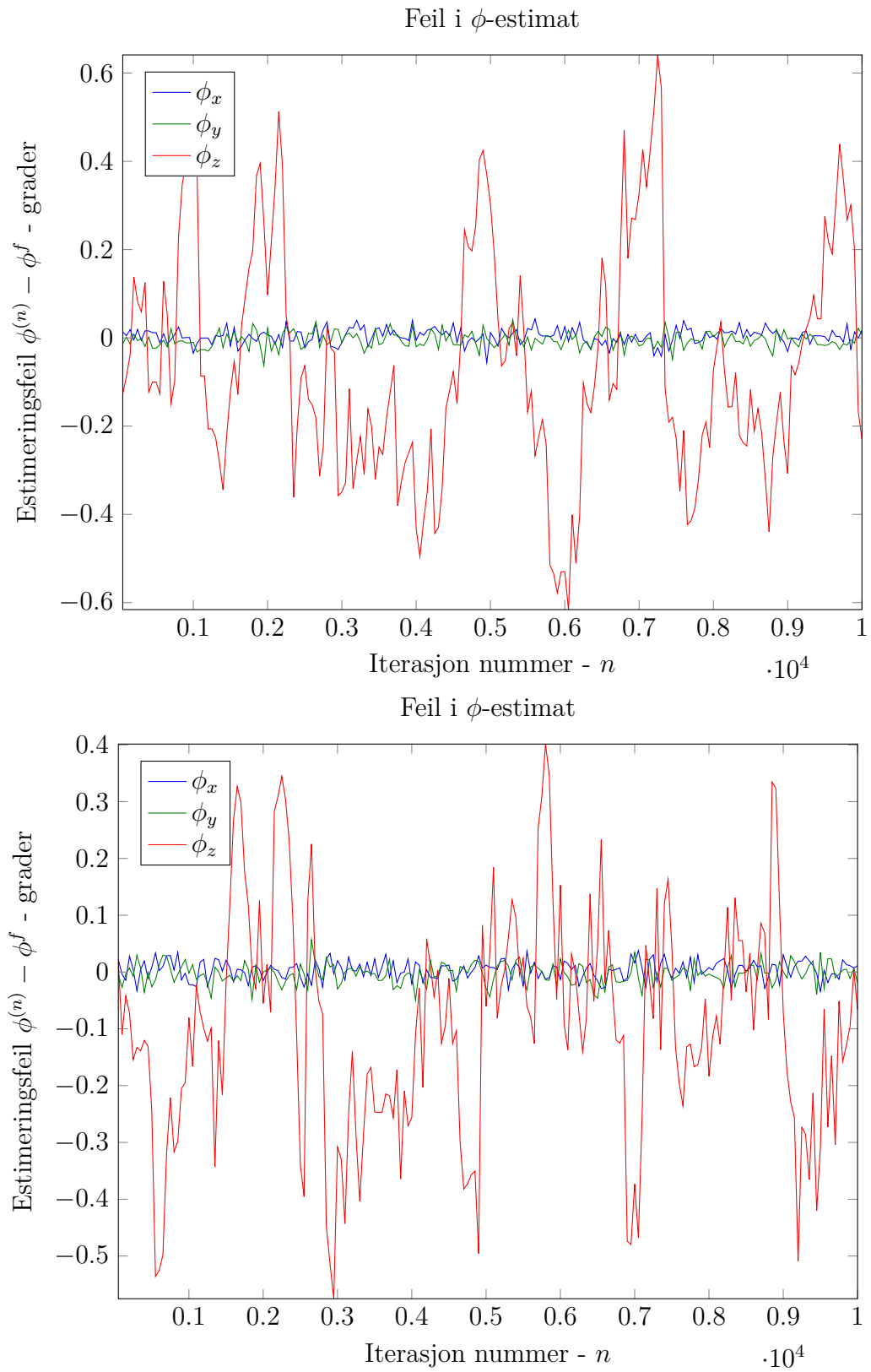


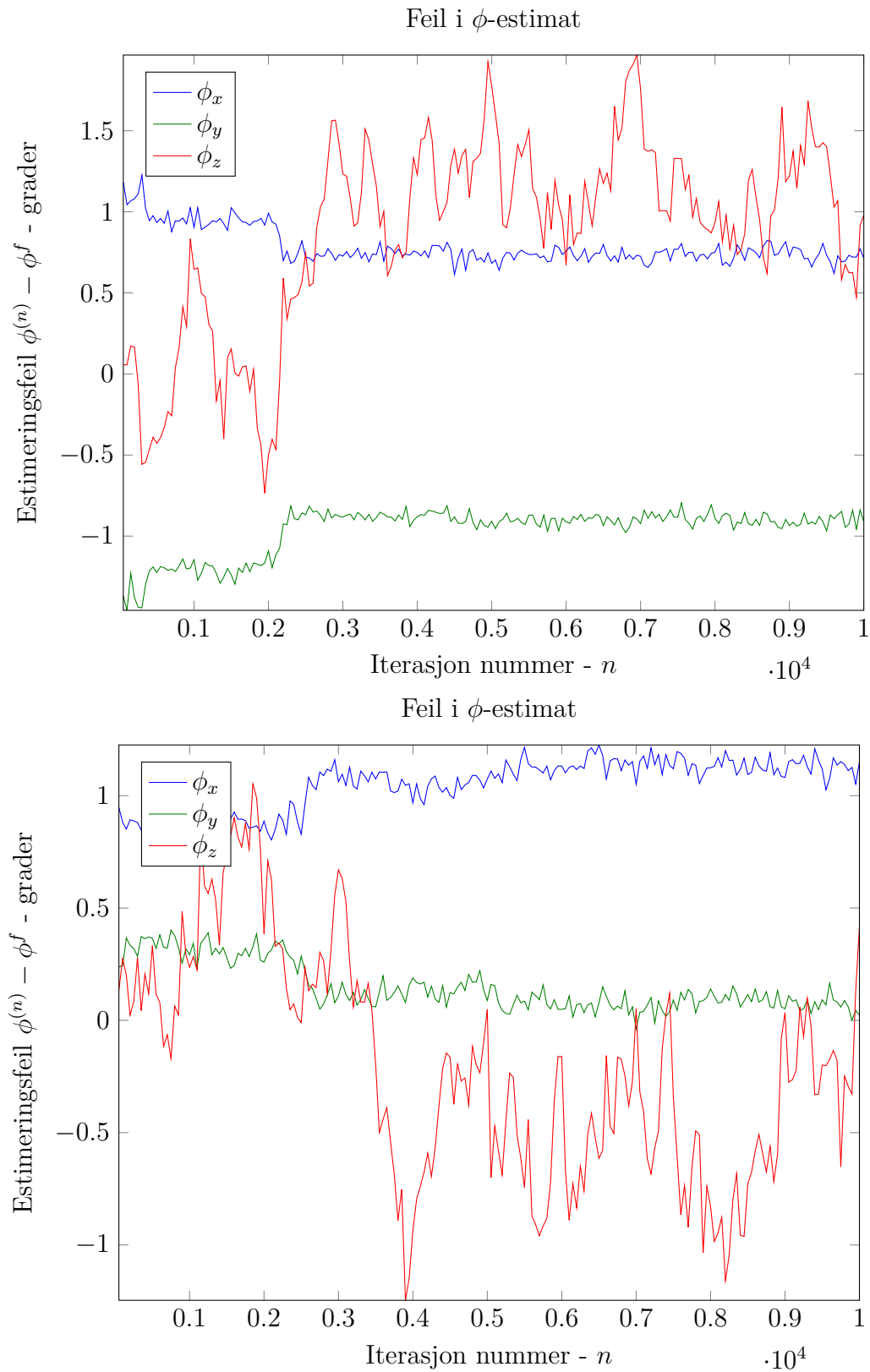
Figur 25: Andel riktige assosiasjoner for NN og GNN, målt i prosent.

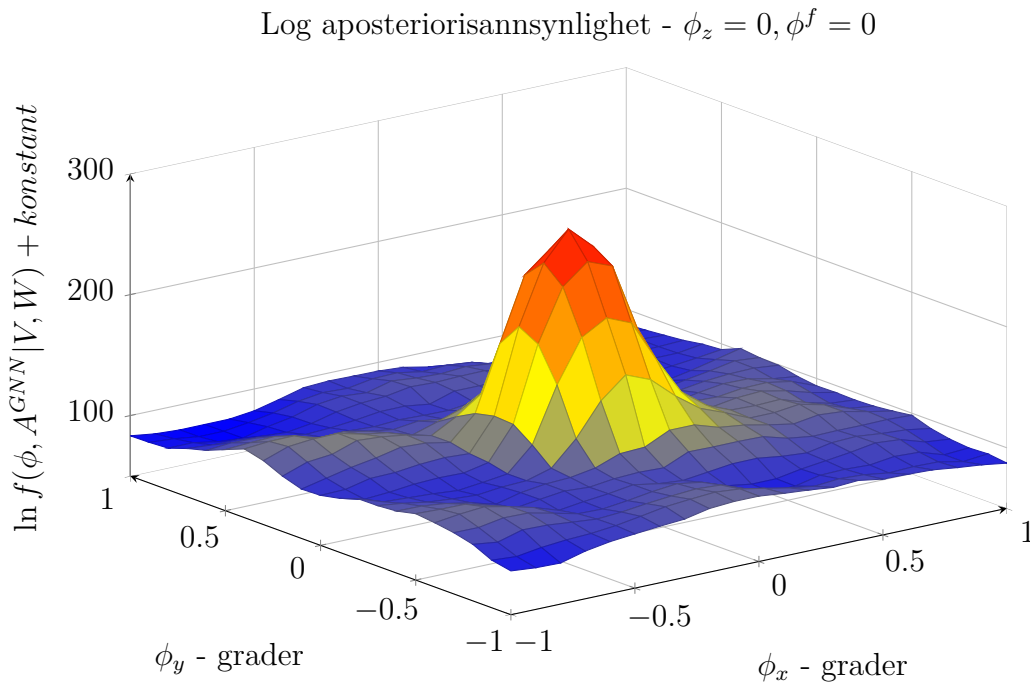
peiling innenfor. Resultatet er at GNN er best på alle fronter, og vil derfor være metoden som blir brukt videre.

5.5 Resultater fra MH-simulering

Når MH algoritmen fra kapittel 4.3 skal startes er det flere valg for hvordan starttilstandene $\phi^{(0)}$ og $A^{(0)}$ kan velges. Algoritmen skal konvergere uansett, men simuleringer av kjeden viser at dette ikke skjer innen rimelig tid. Simuleringene er utført som beskrevet i kapittel 2.9 ved hjelp av fordelingene beskrevet i kapittel 4.3. Figur 26 viser to kjeder som starter i fasit, mens figur 27 viser to kjeder som startet et stykke unna fasit. Alle fire kjedene har brukt de samme simulerte peilingene, slik at alle forskjeller kun skyldes startpunkt og at kjedene er stokastiske. Det er simulert 10 000 runder for hver av kjedene, og tilstanden fra hver femtiende runde er tegnet inn i grafene. Fra disse figurene er det klart at alle kjedene etter hvert stabiliserer seg rundt en verdi, som ikke nødvendigvis er riktig. Det er også utført simuleringer med flere runder og det er ingen tegn til at denne trenden brytes. For å undersøke hva dette skyldes er strukturen på aposteriorifordelingen $f(\phi, A|V, W)$ undersøkt nærmere. Som for figur 22 av J^{GNN} og J^{NN} er dette gjort ved å lage overflateplott mens et redusert antall komponenter varieres. Dette er gjort ved å holde ϕ_z fast i fasitverdien $\phi_z^f = 0$, og bruke GNN-assosiasjonene i punktet for A . Resultatet er figur 28, et overflateplott av $\ln f(\phi, A^{GNN}|V, W) + \text{konstant}$, logaritmen til et uttrykk som er proporsjonalt med aposteriorifordelingen. Den

**Figur 26:** *Utvikling ved start i fasit*

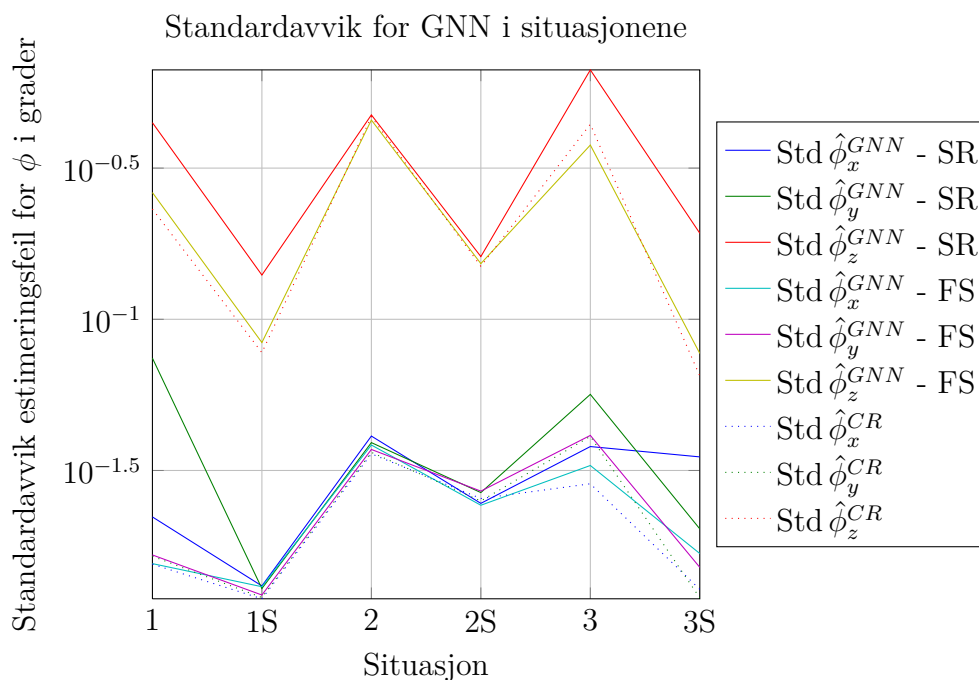
**Figur 27:** Utvikling ved start annet sted enn fasit



Figur 28: *Plott over logaritmen til aposteriorfordelingen opp til en konstant.*

høyeste toppen til overflaten er 250, og den kuperte flaten ved foten av den har verdier fra 50 til 70. Siden plottet har logaritmisk skala betyr dette at $f(\phi, A|V, W)$ er en faktor $e^{250-70} \approx 10^{78}$ høyere ved den høyeste toppen enn på flaten rundt, og at det er et forhold på $e^{50-70} \approx 10^8$ mellom toppene og bunnene på flaten.

Akseptansannsynligheten, α , påvirkes som likning (38) viser av forholdet mellom verdien aposteriorfordelingen for forslaget og verdien nå. Endringsforslagene algoritmen foreslår for ϕ er små relativt til utbredelsen av toppene, og derfor vil det være veldig vanskelig for algoritmen å gå fra en topp til en annen. Altså kan sammensetningen av endringsforslagene og formen til aposteriorfordelingen forklare at kjedene i figur 26 og 27 stabiliserer seg ved at de finner et lokalt maksimum. Ved å forbedre endringsforslagene skal det være mulig å unngå dette problemet. En mulighet er å innføre mange verdier av ϕ med stor spredning og velge forslaget som skal brukes ut fra verdien av aposteriorfordelingen i punktet, lignende måten A velges på ved den drastiske endringsmetoden. En enklere mulighet kan være å starte i fasit, og anta at kjeden da vil oppnå konvergens. Dette er en rimelig antakelse siden toppen rundt fasit er veldig (faktor 10^{78}) mye høyere enn områdene rundt, noe som gjør at sannsynligheten for at en konvergent kjede oppholder seg utenfor toppen rundt ϕ^f trolig er neglisjerbar. Dermed bør resultatet man får med å starte den eksisterende algoritmen i fasit være likt det en ville fått fra en algoritme med forbedret forslagsfordeling. Men på grunn av tidshensyn og at GNN virket mer lovende er ikke dette undersøkt nærmere.



Figur 29: Empirisk bestemte standardavvik for GNN i situasjonene beskrevet i kapittel 5.1.1.

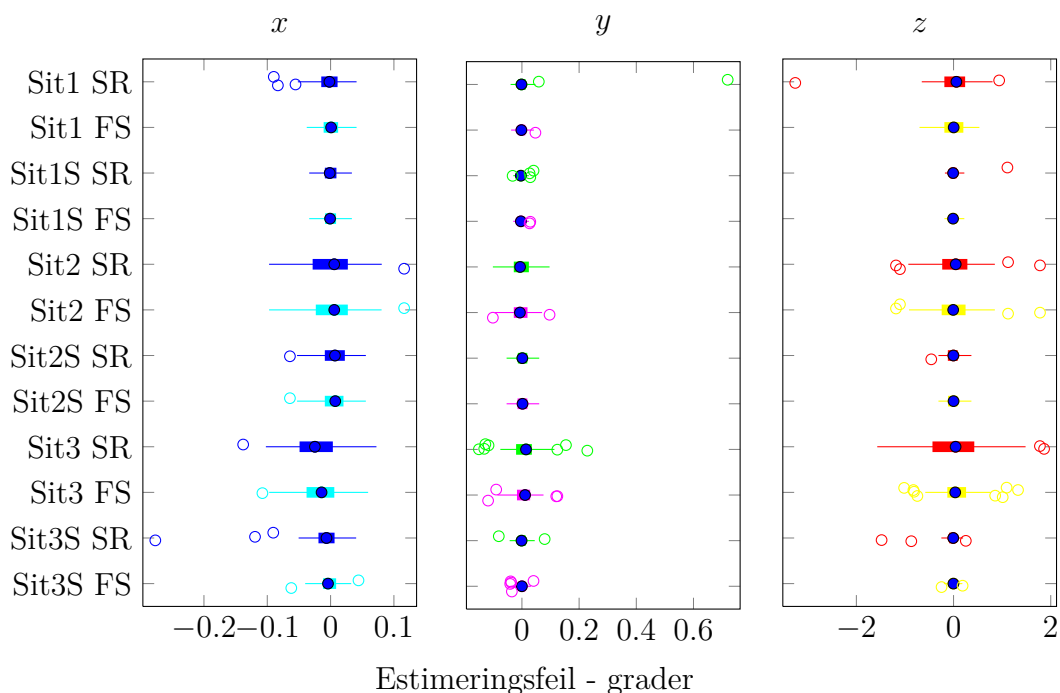
5.6 Simuleringer basert på realistiske data

Dette kapitlet vil vise og diskutere resultatene fra databaserte simuleringer, som beskrevet i kapittel 5.1.4. På grunn av at GNN i kapittel 5.4 viste seg å være den beste estimatoren vil den bli brukt her. Minimeringen av J^{GNN} vil bli utført ved hjelp av fasitstart (FS) metoden fra kapittel 5.3.3 i tillegg til det smarte rutenettsøket (SR) fra kapittel 5.3.2. FS brukes siden den i kapittel 5.3.4 ble valgt til den beste metoden for denne oppgaven. SR er også tatt med for å undersøke om antakelsene gjort i kapittel 5.3 holder slik at FS og SR oppnår like resultater. Simuleringene er utført for stor og liten åpningsvinkel i alle tre situasjonene vist i figur 16. Radar og AIS-peilinger er generert, assosiasjoner er utført og stillingen er estimert 100 ganger for hver åpningsvinkel i alle situasjonene. Resultatene er lagret og analysert. For å lettere skille mellom åpningsvinklene vil S brukes til å angi stor åpningsvinkel, mens det brukes liten åpningsvinkel om ikke annet spesifiseres. Dermed vil situasjon 1S bety situasjon 1 med åpningsvinkel på 10×40 grader mens situasjon 3 vil bety situasjon 3 med åpningsvinkelen 10×10 . Diskusjonen av resultatene vil starte med estimerte standardavvik før den beveger seg over til estimerte forventningsskjevheter og avsluttes med å se nærmere på andelen riktige assosiasjoner.

Estimerte standardavvik for FS og SR vises i figur 29 sammen med tilnærmede

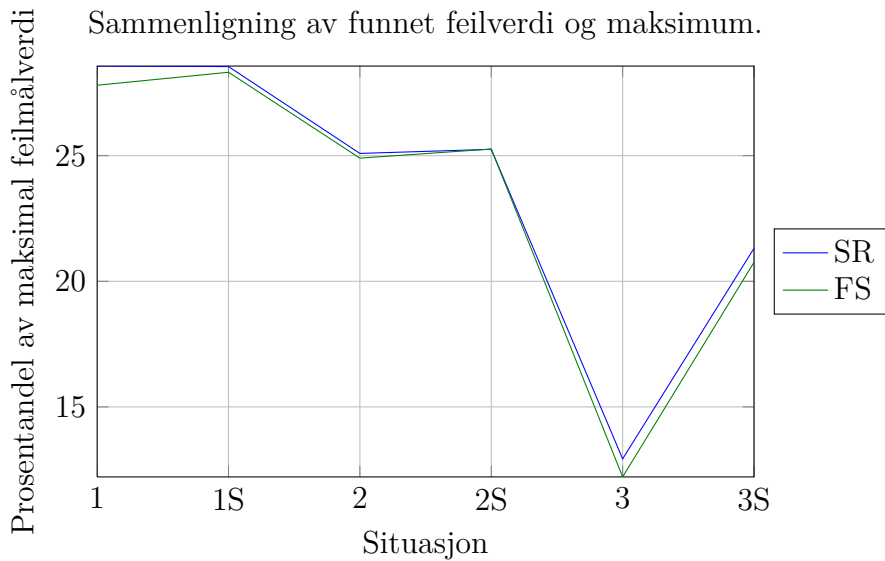
CR-grenser. CR-grensene er beregnet ved metoden i kapittel 2.8 under antakelsen at koblingene, A^f , er kjent. Koblingene endres hver gang radarpeilingene beregnes på nytt, og det er brukt verdien av A^f fra den siste runden for hver situasjon og åpningsvinkel. Fra figur 29 er det tydelig at de estimerte standardavvikene for GNN ved bruk av FS følger CR-grensene tett. I gjennomsnitt er de estimerte standardavvikene ved bruk av FS 6.8% større enn CR-grensene, 9.7% for $\hat{\phi}_x^{GNN}$, 6.6% for $\hat{\phi}_y^{GNN}$ og 4.2% for $\hat{\phi}_z^{GNN}$. Maksimalverdien er at estimert standardavvik for $\hat{\phi}_x^{GNN}$ i situasjon 3S er 31% over CR-grensen og minimalverdien er at estimert standardavvik for $\hat{\phi}_z^{GNN}$ er 15% under. Estimert standardavvik er under CR-grensen i to andre tilfeller, 5% under for $\hat{\phi}_x^{GNN}$ i situasjon 2S og 3% under for $\hat{\phi}_y^{GNN}$ i situasjon 2. Siden CR skal gi en nedre grense er dette bekymringsverdig, men det finnes flere mulige forklaringer. For det første må en estimator være forventningsrett for at CR-grensen skal holde, og nedenfor vises det at dette ikke ser ut til å holde for $\hat{\phi}^{GNN}$. I tillegg er alle standardavvikene til $\hat{\phi}^{GNN}$ estimerte, slik at det er mulig at estimatene er for lave men standardavviket egentlig er over grensen. CR-grensen er også kun en tilnærming, slik at den kan være litt forskjellig fra korrekt verdi. For eksempel hvis A^f fra siste runden er slik at de koblede peilingene egner seg dårligere til estimering enn normalt, vil CR-grensen være høyere enn den burde være. En siste mulighet er at bruk av fasitstillingen, ϕ^f , gjør at det estimerte standardavviket er lavere enn det er mulig å få til uten å bruke ϕ^f . Dette regnes ikke som en sannsynlig forklaring siden estimerte standardavvik for SR i mange tilfeller følger de estimerte standardavvikene for FS tett, faktisk tettere enn figur 29 tilsier av grunner som forklares nedenfor. Bruk av ϕ^f kan uansett ikke være en fullstendig forklaring siden estimert standardavvik for $\hat{\phi}_x^{GNN}$ ved bruk av SR er 4% under CR-grensen i situasjon 2S, og SR bruker ikke ϕ^f i estimeringen. Lettere undersøkelser har ikke vist tegn på at den estimerte CR-grensen skal være for høy. På grunn av dette holdes forventningsskjevhet og usikkerhet i estimerte standardavvik som de beste forklaringene.

I situasjon 2 og 2S ligger de estimerte standardavvikene for SR og FS tett i figur 29, det gjelder også for $\hat{\phi}_x^{GNN}$ og $\hat{\phi}_y^{GNN}$ i situasjon 1S. Ellers er det tydelige avvik av forskjellig størrelse. Det mest ekstreme er for $\hat{\phi}_y^{GNN}$ i situasjon 1 hvor estimert standardavvik for SR er 4.5 ganger større enn for FS. Dette ser i stor grad ut til å skyldes at SR i noen tilfeller ikke klarer å finne riktig søkk, og får sjeldne, store, feil som resultat. Disse feilene trekker opp standardavviket kraftig. I eksempelet over finnes det en feil i y -komponenten på 0.72 grader, når nest største feil er 0.06 grader. Hvis en ser bort fra denne runden ved estimering av standardavvik synker verdien fra 4.5 ganger større enn estimatet for FS til kun 0.14 ganger større enn estimatet. Dette er gjentatt for flere av de største avvikene med tilsvarende resultat, og det er mulig å se dette ved hjelp av figur 30. Figuren viser et såkalt boksplott av estimeringsfeilen for alle komponentene, for alle situasjoner og åpningsvinkler. Sirkelen i midten viser medianfeilen, og den



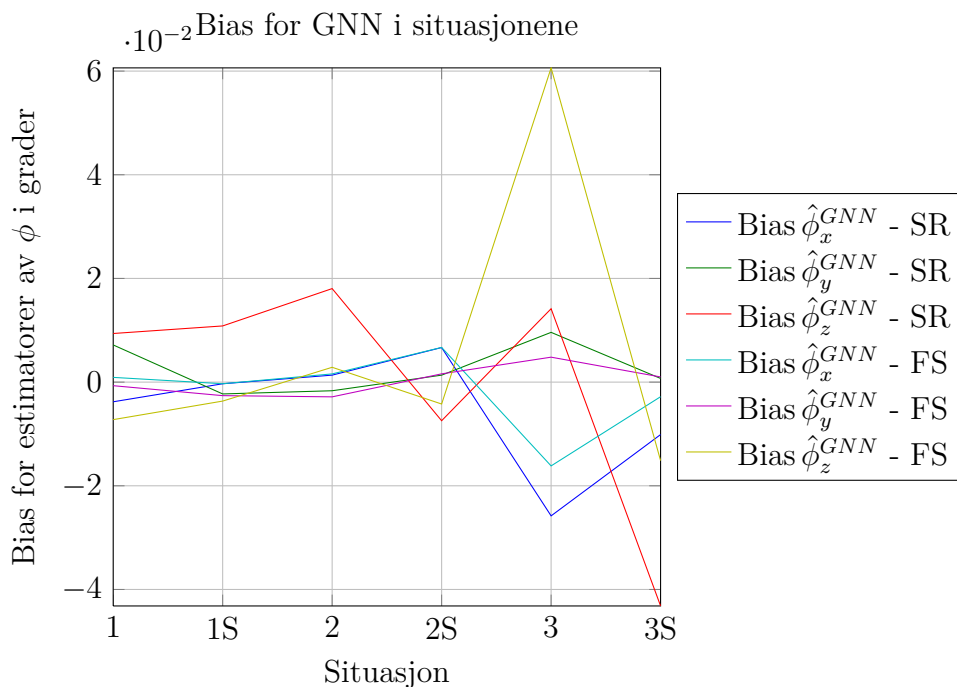
Figur 30: Boksplokk for estimeringsfeilen for de tre komponentene.

tykke linjen strekker seg fra 25% kvantilen til 75% kvantilen. Lengden på den tykke linjen kalles den interkvartile variasjonsbredden, og en feil regnes som en utligger om den avstanden til den mørke linjen er større enn 1.5 ganger den interkvartile variasjonsbredden. Utliggerene tegnes som sirkler og de tynne linjene strekker seg ut til den største feilen som ikke er en utligger. Hvis feilene var normalfordelt ville kun 0.7% av dem vært utliggere, i følge dokumentasjonen til funksjonen `boxplot` i MATLAB (2012), som er brukt til å tegne boksplokket. Fra boksplokket kan vi se en korrelasjon mellom store utliggere og at estimert standardavvik for SR er større enn estimert standardavvik for FS. Sammen med resultatet av fjerning av utliggere i enkelttilfeller tilsier dette at hvis SR forbedres kan den oppnå resultater tilsvarende resultatene fra FS. At SR kan forbedres kan sees tydelig ved å se på verdien av $J^{GNN}(\hat{\phi}^{GNN})$. For utliggeren for $\hat{\phi}_y^{GNN}$ i situasjon 1 er verdien 3.2 ganger høyere for SR estimatet enn for FS estimatet. Verdien er også dobbelt så stor som nest største verdi av feilmålet for alle de 100 rundene, og 2.8 ganger større enn gjennomsnittlig verdi av $J^{GNN}(\hat{\phi})$. Under undersøkelsene av andre situasjoner med utliggere ble det observert at de største utliggerne alltid hadde størst verdi for J^{GNN} , og er større enn tilsvarende verdi for FS estimatet. Dette har to konsekvenser. Dårlige estimat kan gjenkjennes ved størrelsen på feilmålet og SR ikke finner globalt minimum, og derfor kan forbedres. Det siste poenget understrekes av figur 31 som viser at estimatet beregnet ved SR i gjennomsnitt har større verdi på feilmålet enn estimatet som fås ved bruk av FS.



Figur 31: Gjennomsnittlig andel beregnet minimum av feilmålet utgjør av maksimal verdi ($N_v L^2$) for situasjonene beskrevet i kapittel 5.1.1.

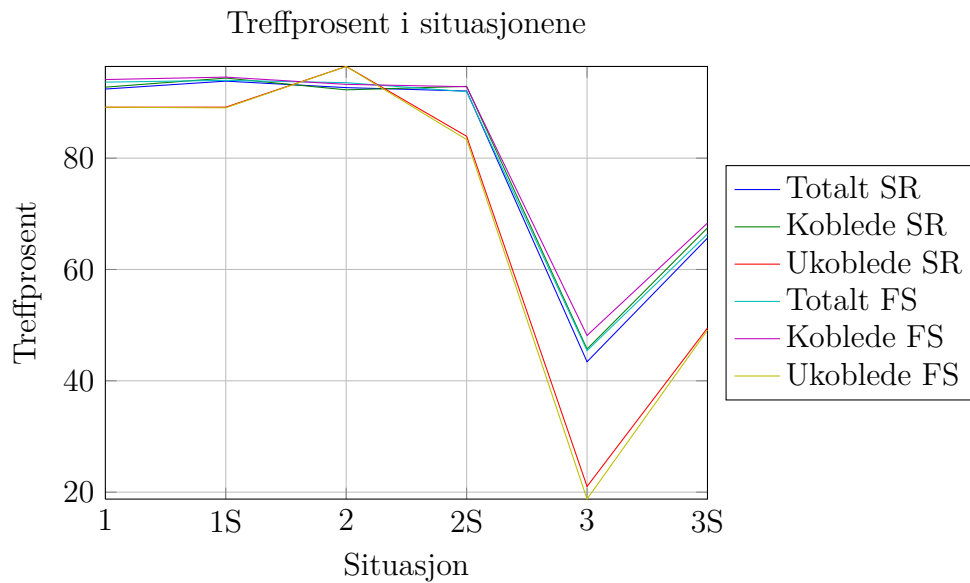
Estimerte forventningsskjevheter er beregnet på samme måte som i tidligere kapitler og er vist i figur 32. Det første som er verdt å legge merke til er at det største estimerte standardavviket er for $\hat{\phi}_z^{GNN}$ beregnet ved FS i situasjon 3, samme situasjon, metode og komponent som ga opphav til estimert standardavvik 15% lavere enn CR-grensen. Dermed kan dette tyde på forventningsskjevhet forklarer at estimert standardavvik er lavere enn CR-grensen, i det minste i dette tilfellet. For å undersøke hvor stor forventningsskjevheten kan forventes å være er der funnet en tilnærmet fordeling for forholdet mellom estimert forventningsverdi og standardavvik. Ved hjelp av sentralgrenseteoremet og Slutskys teorem vises det i Casella og Berger (2002, s. 240) at $\sqrt{n} \text{Bias}(\hat{\phi}_i^{GNN}) / \text{Std} \hat{\phi}_i^{GNN}$, $i = x, y, z$ konvergerer i fordeling til standardnormalfordelingen hvis $\hat{\phi}_i^{GNN}$ er forventningsrett og $n \rightarrow \infty$. Ut fra boksplottet i figur 30 og undersøkelse av histogrammer over estimeringsfeilene ser fordelingen av dem ut til å være rimelig symmetrisk og uten lange haler, slik at estimatet kan forventes å være godt ved $n = 100$. Estimater gir da at det er 95% sannsynlig at $\text{Bias}(\hat{\phi}_i^{GNN}) / \text{Std} \hat{\phi}_i^{GNN}$ er i intervallet $\pm z_{0.025} / \sqrt{100} = \pm 0.20$, hvor z_α er α -kvantilen til standardnormalfordelingen. Altså er det tilnærmet 95% sannsynlig at estimert forventningsskjevhet utgjør mindre enn 20% av estimert standardavvik. Det finnes tre tilfeller hvor denne grensen brytes mens FS brukes, da utgjør estimert forventningsskjevhet henholdsvis 49%, 28% og 21% av estimert standardavvik. Dette skjer for ϕ_x^{GNN} i situasjon 3 og 2S, og for ϕ_y^{GNN} i situasjon 1S. Siden det utføres 18 sammenligninger er det ikke overraskende om det kun var en utenfor på 21%, men 3 utenfor og to større virker ikke sannsynlig. Beregning av p -verdiene gir $9.6 \cdot 10^{-7}$ for 49% og $5.1 \cdot 10^{-3}$ for 28%. Altså kan det med stor sikkerhet konkluderes at $\hat{\phi}^{GNN}$ ikke



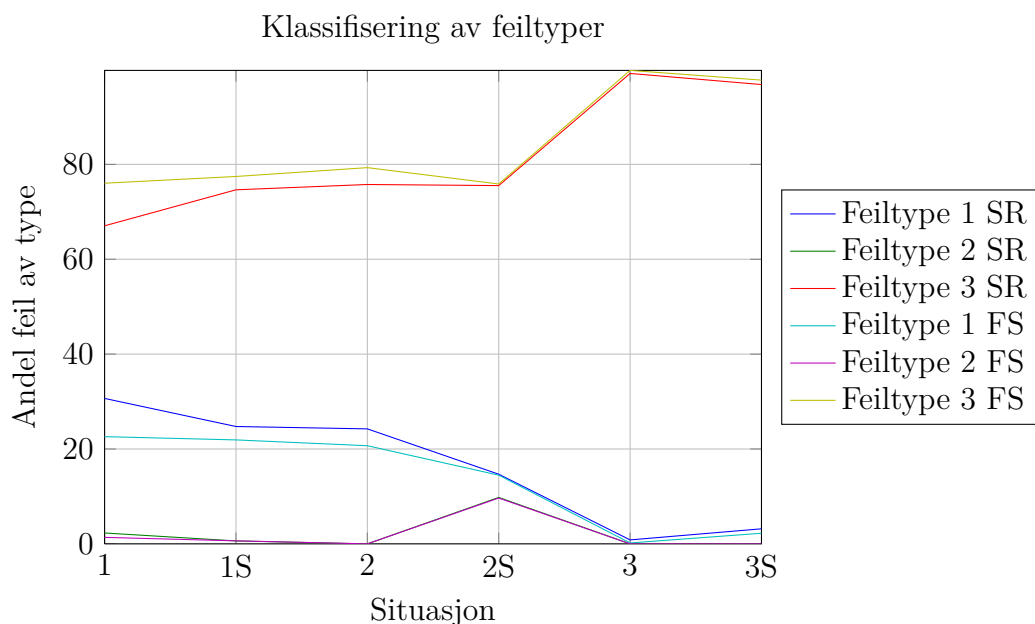
Figur 32: Empirisk bestemte forventningsskjevheter for GNN i situasjonene beskrevet i kapittel 5.1.1.

er forventningsrett i alle situasjonene når den beregnes ved hjelp av FS. Den estimerte forventningsskjevheten i situasjon 3 for $\hat{\phi}_z^{GNN}$ utgjør kun 16% av estimert standardavvik, og er derfor ikke overraskende stort. Men denne tilnærmingen kan ikke bevise at estimatoren er forventningsrett, altså er det fortsatt mulig at dette forklarer differansen, selv om denne forklaringen står litt svakere etter denne analysen.

Andelen korrekte assosiasjoner er vist i figur 33. Hver av feilassosiasjonene er undersøkt og klassifisert i forhold til feiltyperne definert på side 33 i kapittel 4.2.4. Resultatene for SR og FS er rimelig like, så kun resultatene fra FS vil bli diskutert her. Den tydeligste differansen i dataene er at andelen treff er mye lavere for situasjon 3 og 3S enn de andre situasjonene. I situasjon 1, 1S, 2 og 2S er andelen korrekte assosiasjoner omtrent 93%, mens i situasjon 3 og 3S er den henholdsvis 45% og 66%. Dette skyldes at det i situasjon 3 og 3S finnes punktvis veldig høye tettheter av skip, noe som kan sees i figur 17(c). Disse høye tetthetene opptrer i nærheten av byer eller tettsteder og er sannsynligvis skip som ligger i ro. Den store gruppen skip helt sør i situasjon 3 er omkring Hammerfest og her befinner 23 skip i et område som tilsvarer en åpningsvinkel på 1×1 grad. Dette er halvparten av de 46 skipene som er innenfor den lille åpningsvinkelen. Det finnes flere slike områder og for stor åpningsvinkel, situasjon 3S, hvor det for eksempel finnes 15 skip nær Honningsvåg innenfor et område som tilsvarer en åpningsvinkel på 0.5×0.5 grader.



Figur 33: Empirisk bestemt treffprosent for GNN i situasjonene beskrevet i kapittel 5.1.1. Koblet betyr andel riktige assosiasjoner for AIS-peilinger som har en koblet radarpeiling, og ukoblet betyr tilsvarende for AIS-peilingene uten koblet radarpeiling. Totalt betyr andelen korrekte assosiasjoner når alle AIS-peilingene er tatt med.



Figur 34: Klassifisering av feilassosiasjonene for GNN i situasjonene beskrevet i kapittel 5.1.1. Feiltypeene er definert i kapittel 4.2.4 og illustrert i figur 11.

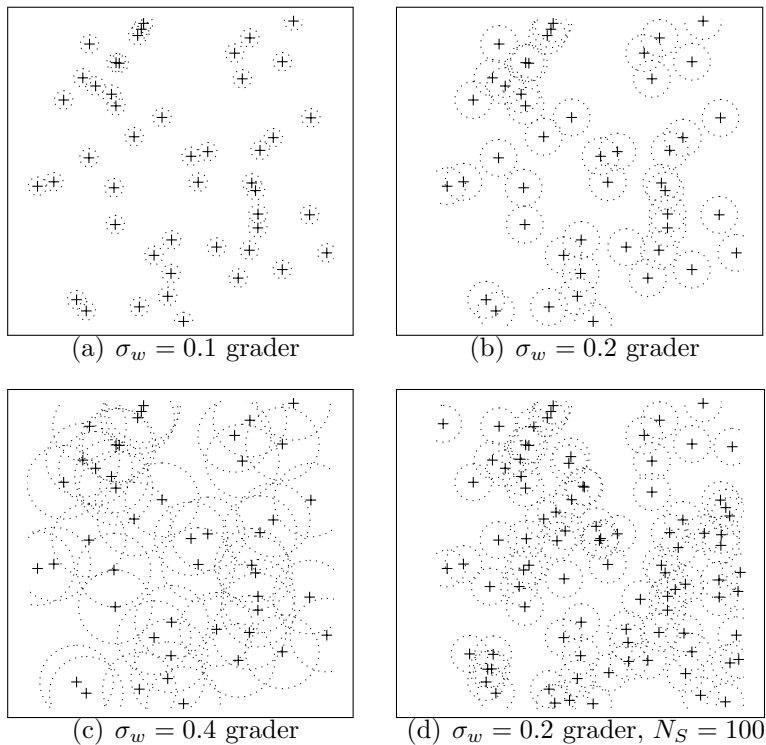
Ved slike tettheter vil det kreves meget høy nøyaktighet på peilingene for å kunne skille skipene, og det er derfor ikke overraskende at andelen riktige assosiasjoner i figur 33 synker. Dette stemmer også overens med at det nesten utelukkende er type 3 feil i situasjon 3 og 3S, som tilsvarer at peilinger er byttet om.

Områdene med veldig høy tetthet gjør at situasjon 3 og 3S ikke oppfyller antakelsen om uniformt fordelte peilinger som ble brukt ved beregning av sannsynlighet for feil i kapittel 4.2.4. De andre situasjonene ser ut til å ha fordelinger av skip som er nærmere, selv om det også i figur 17(a) og 17(b) kan sees tegn til grupper av skip. Det ble i kapittel 4.2.4 estimert at NN skulle få 96% korrekte assosiasjoner hvis vindustørrelsen ble satt til $L = 3\sigma_w$ og argumentert for at GNN skulle gjøre det bedre. Feiltype 1 skulle utgjøre 22% av feilene, mens type 2 og 3 skulle bidra med henholdsvis 26% og 52% av feilene. De empiriske verdiene, beregnet som gjennomsnittet over situasjon 1, 1S, 2 og 2S, er totalt 93% korrekte assosiasjoner hvor 26% er feil av type 1, 2% av type 2 og 72% av type 3. Det tydeligste avviket er at andelen type 3 feil har høyere frekvens enn beregner, mens type 2 feil har lavere frekvens. Det er funnet tre forslag til forklaringer. Avvik fra uniformfordeling vil føre til punktvis høyere tettheter. Dette ville forklart økt forekomst av type 3 feil, siden høyere tetthet vil gjøre det lettere å forveksle målinger. Men det burde også øke forekomsten av type 2 feil siden det er flere radarpeilinger i nærheten, og det bør bli færre type 1 feil, fordi det er mindre sannsynlig å ikke få en ukoblet peiling innenfor vinduet. De globale hensynene til GNN gjør at feil av type 1 og 2 ikke skjer like lett som for NN, men effekten vil fremdeles forventes å trekke i samme retning. I tillegg er det i beregningene gjort en uavhengighetsantakelse som ikke holder. Om dette kan forklare avviket er usikkert, for å avgjøre det er det behov for en mer nøyaktig analyse.

Korrelasjonskoeffisientene er beregnet, men ikke analysert nærmere. De er vist i vedlegg E, og kan brukes til å undersøke korrelasjonen mellom komponentene.

5.7 Effekt av enkeltparametere

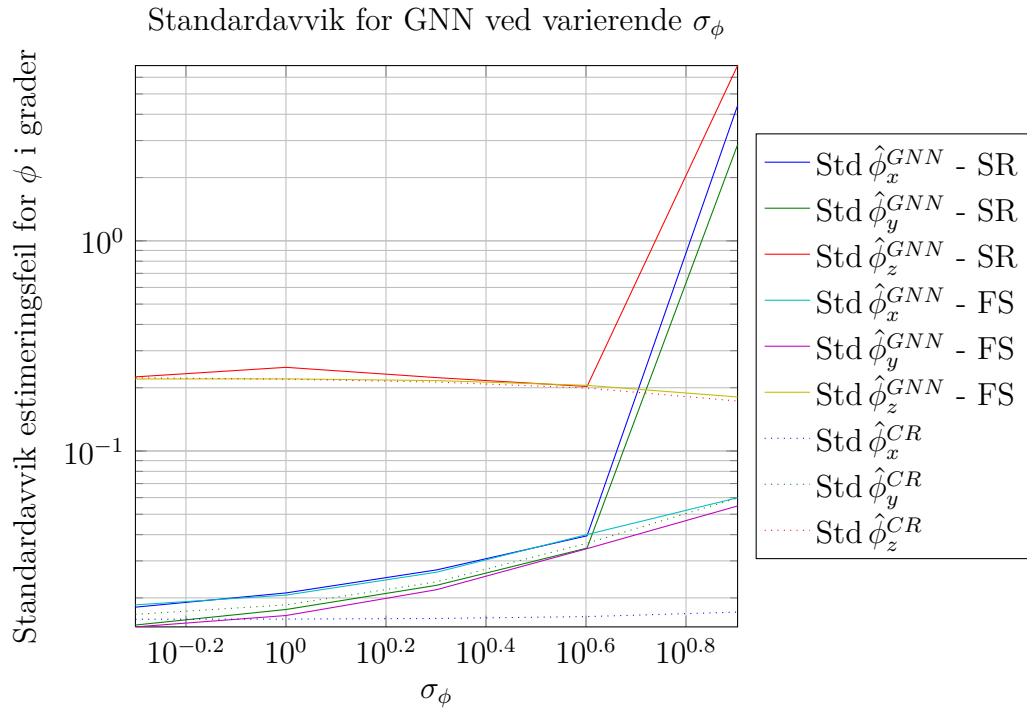
Etter å ha sett på hvordan estimatoren oppfører seg i de forskjellige situasjonene vil fokuset endres til på effekten enkeltparametere har på resultatet. En god måte å gjøre dette på er ved å utføre gjentatte simuleringer mens en parameter varieres, og alle andre parametere holdes på sin standardverdi. Gjennom sammenligningen av NN og GNN i kapittel 5.4 er dette allerede utført for σ_w , og egne simuleringer er utført for σ_ϕ . I tillegg kan forskjellene på situasjonene brukt til simuleringene fra forrige kapittel og kunnskap om GNN sin oppbygning brukes til å diskutere effekten av andre parametere.



Figur 35: Oversikt over hvor mye av åpningsvinkelen som er dekket av vinduene rundt AIS-peilingene. Kryssene indikerer AIS-peilinger og sirklene størrelsen på tilhørende vinduer. Plottene viser hele åpningsvinkelen på 10×10 grader.

5.7.1 Variasjon av feil i radarpeilingene og relaterte parametere

Fra resultatene i kapittel 5.4 kan effekten av å endre feilen i radarpeilingene, σ_w , leses ut. Figur 23 viser rette linjer for GNN, noe som indikerer at standardavvikene for estimeringsfeilene i de forskjellige komponentene er proporsjonale med σ_w^γ , for en verdi av γ , i det undersøkte området. Sammenligning av estimert standardavvik og σ_w viser at γ med god tilnærming kan settes til 1, slik at estimert standardavvik er proporsjonalt med standardavviket til feilen i radarpeilingene. Det forventes derimot at denne trenden vil svekkes siden figur 25 viser at treffprosenten faller kraftig for de høyeste verdiene av σ_w , og dette må etter hvert få konsekvenser for estimeringsnøyaktigheten. For å bedre vise en av effektene endring av σ_w har på GNN er figur 35 produsert. Figur 35(a) til 35(c) viser et eksempel på fordeling av AIS-peilinger og området vinduene dekker ved σ_w lik 0.1, 0.2 og 0.4 grader. Det vises tydelig at ved $\sigma_w = 0.1$ grader er det liten grad av overlapp mellom vinduene, og dette endres til stor grad av overlapp når σ_w når 0.4 grader. Stor grad av overlapp med vinduene vil øke sannsynligheten for forveksling av radarpeilinger. Dette regnes som den mest sannsynlige forklaringen for den fallende treffprosenten. Figur 35(d) demonstrerer at betydningen av σ_w er koblet med antall skip, og er



Figur 36: Empirisk bestemte standardavvik ved varierende σ_ϕ .

laget ved å legge til 50 AIS-peilinger på figur 35(b). Ut fra dette ser det ut som at flere AIS-peilinger og høyere usikkerhet i radarpeilingene ha en tilsvarende påvirkning på GNN, og senking av den ene kan trolig kompensere for økning av den andre.

5.7.2 Simuleringer ved varierende σ_ϕ

For å undersøke effekten av økt usikkerhet i stillingen er det brukt tilfeldig simulering. Simuleringene er utført for $\sigma_w = 0.5, 1, 2, 4, \text{ og } 8$ grader. Empiriske standardavvik beregnet fra simuleringene er vist i figur 36, CR-grensene er beregnet på samme måte som i kapittel 5.6. De sterkt avvikende resultatene for SR ved $\sigma_\phi = 8$ grader skyldes at rutenettsøket ikke var bredt nok, og vil derfor sees bort fra i det følgende. Figur 36 viser en stigende trend for standardavviket for estimert x og y -komponentene, og en svakere synkende trend for z -komponenten. I gjennomsnitt øker $\text{Std } \hat{\phi}_x^{GNN}$ med 35% når σ_ϕ dobles. For $\text{Std } \hat{\phi}_y^{GNN}$ er verdien 41% og $\text{Std } \hat{\phi}_z^{GNN}$ synker i gjennomsnitt med 5%. Dette innebærer at økning av usikkerheten i stillingen har mindre effekt på estimeringsfeilen enn usikkerheten i radarpeilingene. Endringen i størrelsen på estimerte standardavvik kan ha sammenheng med at området skipene plasseres i avgjøres av stillingen, og ved større vinkler vil fordelingen til peilingene endres på grunn av at sikteretningen

til satellitten blir skjevere med hensyn på planet skipene fordeles i.

Estimerte standardavvik for $\text{Std } \hat{\phi}_y^{GNN}$ og $\text{Std } \hat{\phi}_z^{GNN}$ ligger nær CR-grensene for hele intervallet, men det gjelder ikke $\text{Std } \hat{\phi}_x^{GNN}$. CR-grensen for $\text{Std } \hat{\phi}_x^{GNN}$ er tilnærmet konstant, mens det estimerte standardavviket øker på tilsvarende måte som for $\text{Std } \hat{\phi}_y^{GNN}$. Det er ikke funnet noen tilfredsstillende forklaring på hvorfor CR-grensen til x -komponenten skal oppføre seg forskjellig fra y -komponenten, eller hvorfor GNN estimatoren ikke lenger følger CR-grensen tett. De to komponentene spiller nesten symmetriske roller i problemet, det er kun rotasjonsrekkefølgen som gir dem litt forskjellige egenskaper. Denne forskjellen kan kanskje ha innvirkning her.

5.7.3 Andre parametere

Det har ikke blitt utført grundige simuleringer for å undersøke flere parametere. Dette kapitlet vil utnytte informasjonen som finnes i simuleringene som allerede er beskrevet for å belyse effekten av parametere som ikke er diskutert. Det vil også bli gjort noen kvalifiserte gjetninger på hva effekten blir om de to fordelingene som brukes blir endret.

De estimerte standardavvikene for $\hat{\phi}^{GNN}$ for de forskjellige situasjonene i figur 29 viser at de estimerte standardavvikene alltid synker når åpningsvinkelen øker fra 10×10 grader til 10×40 grader. Økning av åpningsvinkelen har to effekter som kan tenkes å forbedre estimeringen, antall skip innenfor åpningsvinkelen øker og utstrekningen til området som peilingene kan mottas fra utvides. Flere skip betyr mer informasjon, så lenge det er mulig å assosiere korrekt, og bør derfor bidra til bedre estimat. Taylorutviklingen av $r(v; \phi)$ i likning (54) viser at større verdier for peilingene vil gjøre at samme endring av ϕ_z gir større endring av peilingene. Det tyder på at økning av området med peilinger vil ha større effekt på estimeringen av ϕ_z^f enn de andre komponentene. Dette underbygges av figur 29 som viser at det estimerte standardavviket for ϕ_z^{GNN} ved bruk av FS i situasjon 2S er lavere enn for situasjon 1, men motsatt for de to andre komponentene. Dette holder selv om tabell 4 viser at det er færre skip innenfor åpningsvinkelen i situasjon 2S enn i situasjon 1.

Det er også mulig å få informasjon om effekten av flere parametere ved å sammenligne figur 29 med tabell 4. Alle situasjonene ved samme åpningsvinkel har brukt det samme settet med parametere bortsett fra skipsposisjoner, antall skip og andel skip det mottas AIS-meldinger fra, siden dette kan hentes fra dataene fra FFI. Dermed må forskjellene mellom dem skyldes dette. Estimerte standardavvik ved bruk av FS for situasjon 2 og 3 er rimelig like, og større enn de estimerte standardavvikene for situasjon 1. Situasjon 1 har 52 skip innenfor åpningsvinkelen, mens situasjon 2 har 9 skip og situasjon 3 har 46 skip innenfor. Deteksjonssann-

synligheten for AIS er henholdsvis 90%, 97% og 36% for de tre situasjonene, som betyr at det mottas AIS-meldinger fra 49, 9 og 17 skip. Av disse verdiene er det antall mottatte AIS-meldinger som har størst korrelasjon med størrelsen på de estimerte standardavvikene, siden verdiene for situasjon 2 og 3 er nærmest hverandre der. Men når det er mottatt dobbelt så mange AIS-peilinger i situasjon 3 som situasjon 2 synes det som at fordelingen av skip sannsynligvis også har en rolle. En kan få et lignende resultat ved å sammenligne situasjon 1S og 3S som har motatt AIS-meldinger fra henholdsvis 80 og 76 skip. Her er også de estimerte standardavvikene rimelig like. For situasjon 1S ligger standardavvikene for x - og y -komponentene omtrent 20% lavere enn 3S, mens det estimerte standardavviket for z -komponenten i situasjon 1S er 9% høyere standardavviket fra situasjon 3S.

Et annet spørsmål hvordan endring av de to fordelingene som er brukt vil påvirke problemet. For stillingen, ϕ^f , forventes det at størrelsen på stillingen, $\|\phi^f\|$, har den største påvirkningen. Dette er fordi $J^{GN}(\phi)$ er lik uansett hvilken fordeling ϕ^f er trekt fra. Effekten endring av størrelsen til $\|\phi^f\|$ har på estimatet ble diskutert i forrige kapittel. Fordelingen til feilen på radarpeilingene, ϵ_i , forventes å ha en større betydning. At $E \epsilon_i = 0$, eller i det minste svært nær null, er essensielt siden feilmålene er bygd på antakelsen at $E w_i = r(v_i, \phi^f)$ når v_i og w_i stammer fra samme skip. At fordelingen inneholder slengere, det vil si en viss sannsynlighet for store³ avvik, forventes å ha samme effekt som å senke P_{rad} og P_{AIS} samtidig som antall skip økes. Dette er fordi at i de tilfeller slengere oppstår vil de havne utenfor vinduet, noe som for algoritmen sin del er det samme som å få kun AIS-peiling fra skipet og en ekstra radarpeiling (men ikke AIS-peiling) fra et annet skip. For feil innenfor vinduet vil symmetri om forventningsverdien være viktig. Det samme gjelder at feilen i de to komponentene er uavhengige og likt fordelte. Det skyldes at feilmålene legger like stor vekt på feil i like stor avstand, uansett retning. Hvis den siste forutsetningen brytes kan det være mulig å tilpasse feilmålene ved hjelp av vektet minste kvadraters metode, som gir mulighet til å legge større vekt på en komponent og kan ta hensyn til korrelasjon mellom feilene.

³I det minste større enn vinduet

6 Oppsummering

I denne oppgaven har undersøkt om det er mulig å forbedre kunnskapen om en overvåkings satellitts stilling ved å assosiere posisjonsinformasjon fra AIS med radarpeilinger. Det er utarbeidet en matematisk beskrivelse av problemet, som har gitt modellen i kapittel 3. Ut fra denne modellen er det i kapittel 4 utarbeidet flere estimatorene for å estimere satellittens stilling, basert på forskjellige antakelser og metoder. For å teste estimatorene er det utviklet to simuleringsverktøy i MATLAB (2012) basert på modellen, et som tar utgangspunkt i realistiske data og et som genererer alle nødvendige data. Basert på de tilfeldige simuleringene ble det avgjort at numerisk minimering av et feilmål oppnådde bedre resultater enn analytisk minimering av en lineær tilnærming av feilmålet, når koblingene mellom AIS- og radarpeilingene er kjent. Når koblingene er ukjente ble det på samme måte vist at GNN var bedre enn NN ved store feil i peilingene, men at NN med fordel kan brukes for å finne et startpunkt for GNN på grunn av kortere kjøretid. Ved hjelp av de databaserte simuleringene ble det funnet at x og y komponenten til stillingen kan estimeres med nøyaktighet på 0.012–0.041 grader, avhengig av situasjonen, når usikkerheten i radarpeilingene er på 0.1 grader. For z -komponenten er de tilsvarende verdiene 0.077–0.46 grader. Det ble også beregnet tilnærmede CR nedre grenser for disse situasjonene, under antakelsen om kjente koblinger. Grensene er 0.012–0.041 for x - og y -komponentene og 0.065–0.46 for z -komponenten. Nøyaktigheten i stillingsestimateret er funnet proporsjonalt med usikkerheten i radarpeilingene i intervallet 0.025–0.4 grader, men det forventes ikke at denne trenden fortsetter. Større åpningsvinkel har gitt bedre estimat av stillingen for alle undersøkte situasjoner, noe som nok har sammenheng med økt antall skip. Ved tilstrekkelig lav tetthet av skip og usikkerhet i radarpeilinger oppnår GNN riktig assosiasjon i over 90% av tilfellene. I kystnære områder, med punktvis høy tetthet av skip, synker andelen riktige assosiasjoner til mellom 50%–70%. Usikkerheten i radarpeilingene ser ut til å være parameteren som påvirker nøyaktigheten i estimeringen sterkest. Økt usikkerhet i stillingen ser ikke ut til å senke andelen riktige assosiasjoner, og har varierende effekt på estimeringsfeilen. Standardavviket for x og y -komponentene av stilingen øker i gjennomsnitt med henholdsvis 35% og 41% når usikkerheten i stillingen doubles, mens standardavviket for z -komponenten synker med 5%.

Samlet viser dette at det under betingelsene brukt i simuleringene er mulig å bruke assosiasjoner mellom AIS-posisjoner og radarpeilinger til å forbedre kunnskapen om satellittens stilling. Estimeringsmetodene utviklet i denne oppgaven utfører estimeringen på en optimal eller nær optimal måte.

7 Anbefalinger

Fra de realistiske situasjonene som er undersøkt ser det ut til at punkt med meget høy tetthet av skip, typisk nær byer eller tettsteder, er det som gir størst problemer. Siden det uansett er umulig å assosiere korrekt i disse områdene (med mindre feilen i radarfeilen senkes godt under 0.1 grad) kan det kanskje være nyttig å erstatte disse gruppene med gjennomsnittet. Hvis dette gjøres må det vurderes om feil i denne peilingen skal tillegges større eller mindre vekt enn vanlige peilinger.

Grupper på en litt større skala kan kanskje også brukes til å reduserer arbeidet som kreves ved et rutenettsøk. Siden grupper av AIS-peilinger bør ha en tilsvarende gruppe av radarpeilinger bør det være mulig å oppnå en forbedring av kjennskapen til stillingen ved å kjøre algoritmen på gruppene istedet for alle peilingene. Reduksjon av antall vil gi redusert kjøretid for grovinnretting, og forbedring i kunnskapen om stillingen reduserer nødvendig størrelse på rutenettet.

En tilsvarende fremgangsmåte kan også være mulig hvis det ønskes å ta hensyn til at det kan komme flere radarpeilinger fra samme skip. Hvis radarpeilingene som kommer fra samme skip kan identifiseres, kan det være en fordel å se på dem som en gruppe som skal assosieres med en AIS-peiling. Totalsummen $\sum_i \|w_i - r(v_j, \phi)\|^2$ hvor i løper over verdier som er slik at w_i alltid kommer fra samme skip, kan for eksempel brukes istedet for de individuelle avstandene $\|w_i - r(v_j, \phi)\|$. En måte å identifisere peilingene fra et skip er ved å se på frekvens og kjennetegn ved radarpulsene som sendes ut, noe som fungerte i liten skala under eksperimenter ved FFI.

I en praktisk anvendelse er det nødvendig å ta hensyn til at satellittens posisjon endres over tid, noe som vil endre hvordan AIS-peilingene må beregnes. Hvis utviklingen til skjevstillingen til satellitten er slik at den kan antas konstant under en passering er det mulig å behandle dette problemet ved å gjøre AIS-peilingene tidsavhengige. Hvis radarpeilingen w er mottatt ved tidspunktet t vil sammenhengen med den koblede AIS-peilingen $v(t)$ være $w = r(v(t); \phi^f) + \epsilon$. Gitt posisjonen til satellitten ved tidspunktet t kan AIS-peilingen $v(t)$ beregnes på samme måte som i kapittel 3.2. Ved denne metoden kan NN og GNN-metodene anvendes direkte, og siden problemene er matematisk sett like forventes det at resultatene i stor grad vil være overførbare. Men metoden vil kreve mer regnetid siden det må beregnes ett sett AIS-peilinger for hver radarpeiling.

8 Referanser

- Adams, R. A. (2006). *Calculus : A Complete Course*, Pearson/Addison-Wesley, Toronto, Ont.
- Andersen, P. H. (2005). Geolocation algorithms, *Technical Report RES 6300.1*, Forsvarets forskningsinstitutt.
- Bertsekas, D. (1991). *Linear Network Optimization: Algorithms and Codes*, MIT Press.
- Casella, G. og Berger, R. L. (2002). *Statistical Inference*, Duxbury, Pacific Grove, Calif.
- Castanon, D. A. (1992). New assignment algorithms for data association, *Signal and Data Processing of Small Targets 1992* **1698**(1): 313–323.
- Davis, S. L. (2011). Google earth toolbox. Sist sjekket 27. mai 2012.
URL: <http://www.mathworks.com/matlabcentral/fileexchange/12954>
- Duda, R. O., Hart, P. E. og Stork, D. G. (2001). *Pattern Classification*, Wiley, New York.
- Gamerman, D. og Lopes, H. F. (2006). *Markov Chain Monte Carlo : Stochastic Simulation for Bayesian Inference*, Taylor & Francis, Boca Raton, Fla.
- Google Earth (2012). *Versjon 6.2.2.6613*, Google inc.
- Hein, J. (2009). Example: The 3dplot package. Sist sjekket 8. mars 2012.
URL: <http://www.texample.net/tikz/examples/the-3dplot-package/>
- Kystverket (2011). Ais regelverk og brukarkrav. Sist sjekket 27. mai 2012.
URL: <http://www.kystverket.no/Maritime-tjenester/Meldings-og-informasjontjenester/Automatisk-identifikasjonssystem-AIS/AIS-regelverk-og-brukerkrav/>
- Lagarias, J. C., Reeds, J. A., Wright, M. H. og Wright, P. E. (1998). Convergence properties of the Nelder–Mead simplex method in low dimensions, *SIAM Journal on Optimization* **9**(1): 112–147.
- Lehmann, E. L. og Casella, G. (1998). *Theory of Point Estimation*, Springer-Verlag New York, Inc., New York, NY.
- MATLAB (2012). *Versjon 7.14.0.739 (R2012a)*, The MathWorks inc., Natick, Massachusetts.
- McDonald, J. og Weiss, N. (1999). *A Course in Real Analysis*, Academic Press.

Ross, S. M. (2010). *Introduction to Probability Models*, 10 edn, Academic Press.

Trefethen, L. og Bau, D. (1997). *Numerical Linear Algebra*, Society for Industrial and Applied Mathematics.

Walpole, R. E., Myers, R. H., Myers, S. L. og Ye, K. (2007). *Probability & Statistics for Engineers & Scientists*, 8 edn, Pearson Prentice Hall, Upper Saddle River, N.J.

A Spesifikasjoner for maskinvare

Alle simuleringer er utført på en 13" MacBook Pro fra våren 2011 med 8GB 1333MHz DDR3 RAM og 2,3 GHz Intel Core i5 prosessor.

B En auksjonsalgoritme for å finne GNN-assosiasjon

En klasse algoritmer for å finne GNN-assosiasjon er kalt auksjonsalgoritmer. De er basert på en alternativ matematisk formulering fra feltet "lineær programmering", og i den konteksten har det fått navnet det "asymmetriske tilordningsproblemet" ("Asymmetric Assignment Problem"). Navnet asymmetrisk kommer fra at det er ulik størrelse på de to gruppene, som i eksempelet fra figur 4 hvor P større enn Q. Som nevnt i kapittel 2.5 er det i denne oppgaven er det brukt en algoritme hentet fra Castanon (1992). Beskrivelsen av det generelle problemet er basert på Bertsekas (1991).

I det asymmetriske tilordningsproblemet ønskes det å tilordne hver av m personer en av n objekter, når det er flere objekter enn personer ($n > m$). Det er knyttet en gevinst b_{ij} til å tilordne objekt j til person i , og det ønskes å gjøre tilordningen slik at total gevinst blir maksimal. For eksempelet i kapittel 2.5 vil gevinsten være $b_{ij} = -\|q_i - p_j\|$. Det er i tillegg en begrensing som sier at person i kun kan tilordnes objekt j om paret (i, j) er med i mengden av tillatte par \mathcal{A} . Med en tilordning S menes en mengde av tillatte par $(i, j) \in \mathcal{A}$ hvor det maksimalt er et par med en bestemt person og det samme for et bestemt objekt, det vil si at $(i, j), (i', j') \in S \Rightarrow i \neq i' \wedge j \neq j'$. Tilordningen S kalles brukbar om alle personer har blitt tilordnet et objekt under S . Tilordningen kan også representeres med variablene x_{ij} , hvor $x_{ij} = 1$ og $x_{ij} = 0$ betyr henholdsvis i tilordnet og ikke tilordnet j det vil si $x_{ij} = 1 \Leftrightarrow (i, j) \in S$. Med dette grunnlaget kan GNN-problemet formuleres som

$$\text{maksimer } \sum_{(i,j) \in \mathcal{A}} b_{ij} x_{ij}, \quad (117a)$$

med hensyn på

$$\sum_{i|(i,j) \in \mathcal{A}} x_{ij} = 1 \quad \forall i = 1, \dots, m, \quad (117b)$$

$$\sum_{j|(i,j) \in \mathcal{A}} x_{ij} \leq 1 \quad \forall i = 1, \dots, n, \quad (117c)$$

$$x_{ij} \in \{0, 1\} \quad \forall (i, j) \in \mathcal{A}. \quad (117d)$$

Betingelsene (117b) og (117c) representerer at alle personer skal tilordnes et objekt og at et objekt maksimalt kan tilordnes en person. Auksjonsalgoritmene opererer

ikke på denne formuleringen, men på et såkalt dualproblem. Ved å innføre en kostnad k_j for alle objektene og en fordel κ_i for hver person kan dualproblemet formuleres som

$$\text{minimer } \sum_{i=1}^m \kappa_i + \sum_{j=i}^n k_j - (n - m)\lambda \quad (118a)$$

med hensyn på

$$\kappa_i + k_j \geq b_{ij}, \quad \forall (i, j) \in \mathcal{A}, \quad (118b)$$

$$\lambda \leq k_j, \quad j = 1, \dots, n. \quad (118c)$$

Dualproblemet knyttes til det opprinnelige problemet gjennom en egenskap som kalles ε -komplementært slingringsmonn (ε -complementary slackness), forkortet ε -KS. Hvis k er mengden av kostnadene $\{k_j\}$ og κ er tilsvarende for fordelene sies en tilordning S sammen med kostnad-fordel paret (k, κ) å tilfredsstillere ε -KS om

$$\kappa_i + k_j \geq a_{ij} - \varepsilon, \quad \forall (i,) \in \mathcal{A}, \quad (119a)$$

$$\kappa_i + k_j = a_{ij}, \quad \forall (i, j) \in S, \quad (119b)$$

$$k_j \leq \min_{k: \text{tilordnet under } S} p_k, \quad \forall \text{ objekter } j \text{ ikke tilordnet under } S. \quad (119c)$$

Betydningen av ε -KS blir tydelig fra følgende resultat. Om S er en brukbar tilordning som sammen med (k, κ) oppfyller ε -KS, vil S være innen $m\varepsilon$ fra å være optimal for det asymmetriske tilordningsproblemet. Det vil si at resultatet av summen i likning (117a) når x_{ij} representerer S i verste fall er $m\varepsilon$ mindre enn maksimalverdien av summen. Bevis for dette finnes på s. 184 i Bertsekas (1991).

Auksjonsalgoritmene beskrevet i Castanon (1992) består av iterasjoner som for en fiksert ε starter med og opprettholder en tilordning S , et par (k, κ) og et tall λ som oppfyller de to første kravene til ε -KS (119a, 119b), samtidig som

$$\lambda \leq k_j \quad \forall \text{ objekter } j \text{ som ikke er tilordnet under } S. \quad (120)$$

Iterasjonene, som kan være en av to typer, utføres frem til S blir brukbar og den tredje ε -KS betingelsen (119b) blir oppfylt. De to typene kalles forover og revers iterasjon, og i beskrivelsen er det bruk for mengdene av mulige tilordninger for personer eller objekter: $A(i) = \{j | (i, j) \in \mathcal{A}\}$ og $B(j) = \{i | (i, j) \in \mathcal{A}\}$. For å kunne bruke forover iterasjon må det finnes en person som ikke er tilordnet et objekt, revers iterasjonene krever at det finnes et objekt j som ikke er tilordnet en person og har $k_j > \lambda$. Iterasjonene utføres på følgende måter.

Forover iterasjon: Finn en person i som ikke er tilordnet et objekt, personens

beste objekt j_i og verdiene v_i og w_i definert ved:

$$j_i = \arg \max_{j \in A(i)} (b_{ij} - k_j) \quad (121a)$$

$$v_i = \max_{j \in A(i)} (b_{ij} - k_j) \quad (121b)$$

$$w_i = \max_{j \in A(i), j \neq j_i} (b_{ij} - k_j) \quad (121c)$$

Om det kun er ett objekt i $A(i)$, sett w_i til $-\infty$ (i implementering et veldig stort, negativt tall). Sett så

$$k_{j_i} = \max\{\lambda, b_{ij} - w_i + \varepsilon\} \quad (121d)$$

$$\kappa_i = w_i - \varepsilon. \quad (121e)$$

Om $\lambda \leq b_{ij} - w_i + \varepsilon$ legg (i, j_i) til S og fjern (i', j_i) fra S om objektet j_i var tilordnet en person ved starten av iterasjonen.

Revers iterasjon: Finn et objekt j med $k_j > \lambda$, objektets beste person i_j og verdiene β_i og γ_i definert ved:

$$i_j = \arg \max_{i \in B(j)} (b_{ij} - \kappa_i) \quad (122)$$

$$v_j = \max_{i \in B(j)} (b_{ij} - \kappa_i) \quad (123)$$

$$w_j = \max_{i \in B(j), i \neq i_j} (b_{ij} - \kappa_i) \quad (124)$$

$$(125)$$

Om det kun er en person i $B(j)$ sett γ_j til $-\infty$ (i implementering et veldig stort, negativt tall). Fortsett med et av de to følgende tilfellene:

1. Om $\beta_j \geq \lambda + \varepsilon$ sett

$$k_j = \max\{\lambda, \gamma_j - \varepsilon\} \quad (126)$$

$$\kappa_{i_j} = b_{i_j j} - \max\{\lambda, \gamma_j - \varepsilon\} \quad (127)$$

og legg (i_j, j) til i S . Hvis person i hadde et objekt j' tilordnet ved starten av iterasjonen, fjern (i_j, j') fra S .

2. Om $\beta_j < \lambda + \varepsilon$ sett $k_j = \beta_j + \varepsilon$.

Forover og revers iterasjonene kan kombineres på mange forskjellige måter. Sammensetningen fra kapittel 3.6 i Castanon (1992) brukes videre. Algoritmen settes da sammen av runder som starter med S, λ, k og κ som oppfyller (119a, 119b) og en verdi for ε .

Steg 1 Utfør forover-iterasjoner frem til en ny person har blitt tilordnet et objekt. Hvis det fremdeles finnes personer uten tilordnet objekt, gå til steg 2. Hvis ikke gå til steg 3.

Steg 2 Utfør bakover-iterasjoner til et nytt objekt blir tilordnet, eller $k_j \leq \lambda$ for alle objekter som ikke er tilordnet en person. Hvis det fremdeles finnes personer uten tilordnet objekt, gå til steg 1. Hvis ikke gå til steg 3.

Steg 3 Utfør bakover-iterasjoner frem til $k_j \leq \lambda$ for alle objekter som ikke er tilordnet en person. Avslutt runden med ε -KS oppfylt.

Algoritmen implementerer såkalt ε skalering ved å kjøre flere slike runder med stadig synkende ε . For den første runden kan $S = \emptyset$ og $\lambda = k_j = \kappa_i = 0 \forall i, j$, i påfølgende runder settes igjen $S = \emptyset$ og $\kappa_i = 0$, men kostnadene k_j beholdes og λ settes til den minste kostnaden blant objektene som ble tilordnet forrige runde

$$\lambda = \min_{j|j\text{var tilordnet i forrige runde}} k_j. \quad (128)$$

B.1 Tilpasning av algoritmen om ikke alle personer må tilordnes

I anvendelsene av GNN i oppgaven tillates det at noen personer ikke får tilordnet noen objekter. Den samme algoritmen kan brukes til dette ved å legge til et fiktivt objekt pr person som representerer “ikke tilordnet”. De fiktive objektene vil også måtte få en passende gevinst knyttet til seg, og \mathcal{A} må få elementer som tillater tilordning til de nye elementene. Etter innføring av disse kan algoritmen over benyttes direkte.

B.2 Implementering av GNN assosiasjonen fra kapittel 4.2.3

I dette kapittelet beskrives det hvordan størrelsene som er nødvendige for å bruke auksjonsalgoritmen beregnes ut fra peilingene V , W , vinduet L og forslaget for stillingen ϕ . En anvendelse av algoritmen gir da A^{GNN} og $J^{GNN}(\phi)$.

Mengden av mulige tilordninger (\mathcal{A}) settes til å bestå av par (i, j) , $1 \leq i \leq N_v$, $1 \leq j \leq N_w$ hvor avstanden $\|w_j - r(v_i, \phi)\|^2$ er mindre enn vinduet L , i tillegg til parene $(i, i + N_w)$ som representerer ingen kobling. Gevinstene b_{ij} for $(i, j) \in \mathcal{A}$ blir satt til

$$b_{ij} = \begin{cases} -\|w_j - r(v_i, \phi)\|^2, & j \leq N_w \\ -L, & j > N_w, \end{cases} \quad (i, j) \in \mathcal{A} \quad (129)$$

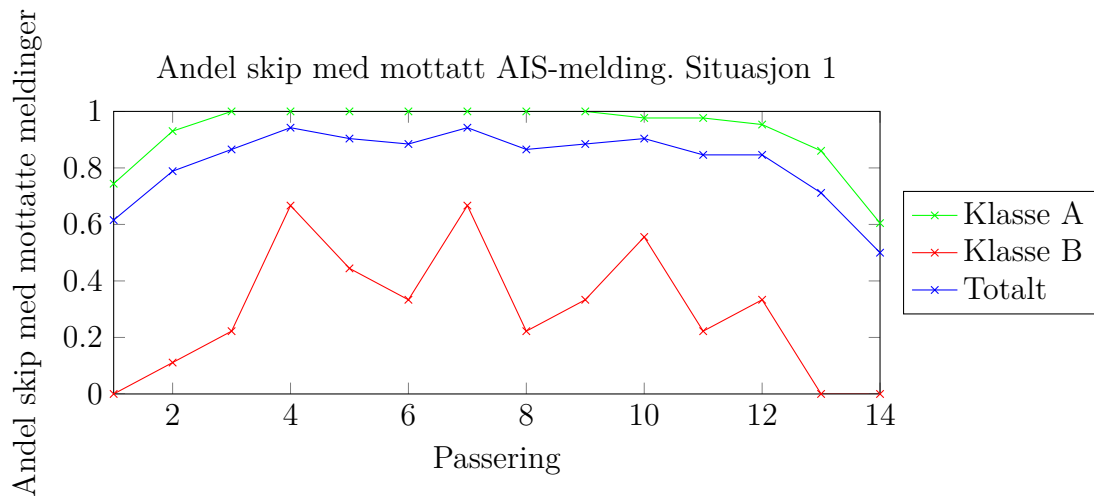
Gevinstene b_{ij} settes negative siden algoritmen søker etter maksimum, mens det ønskes at avstandene er minst mulig. Feilmålet $J^{GNN}(\phi)$ kan ved hjelp av dette beregnes i tre steg:

1. Beregn mulige tilordninger \mathcal{A} og gevinstene b_{ij} basert på ϕ .
2. Finn GNN assosiasjoner x_{ij} basert på dette.
3. Konverter assosiasjonene til vektorformen A^{GNN} ved å la $a_i^{GNN} = j$ om $x_{ij} = 1$ for $j \leq N_W$ og $a_i^{GNN} = 0$ om $x_{ij} = 1$ for $j > N_W$. Beregn $J^{GNN}(\phi) = J_{A^{GNN}}(\phi)$.

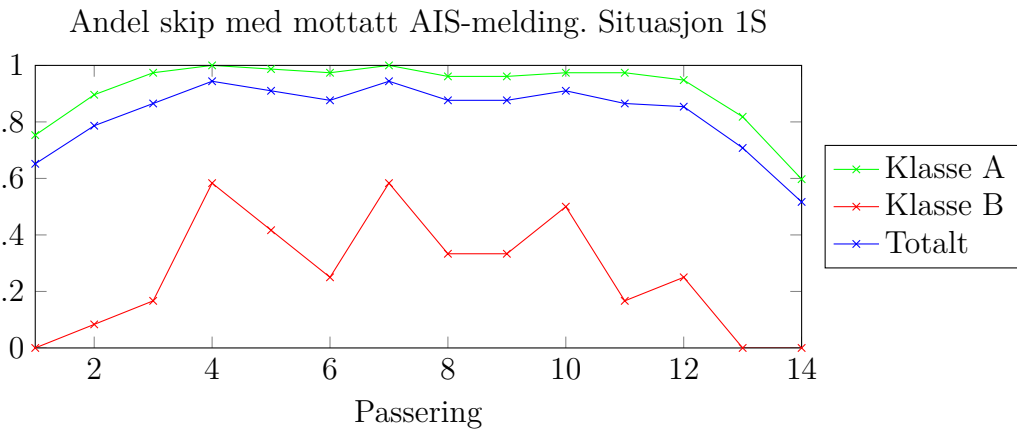
Beregningsmessig er det mer effektivt å finne J^{GNN} ved en matematisk ekvivalent metode: La $J^{GNN}(\phi) = -\sum_{(i,j) \in \mathcal{A}} x_{ij} b_{ij}$.

C Deteksjonsansynligheter for alle situasjoner

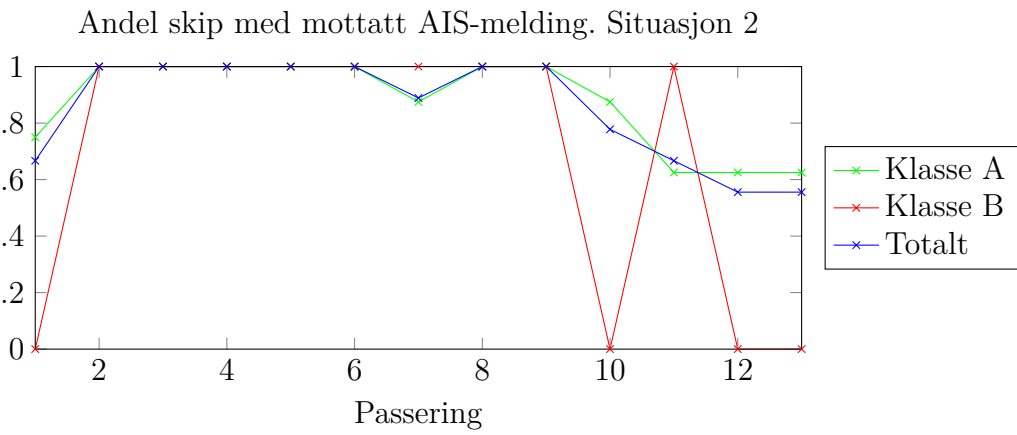
S, som i “Situasjon 1S”, står for “Stor åpningsvinkel”, det vil si 10×40 grader.



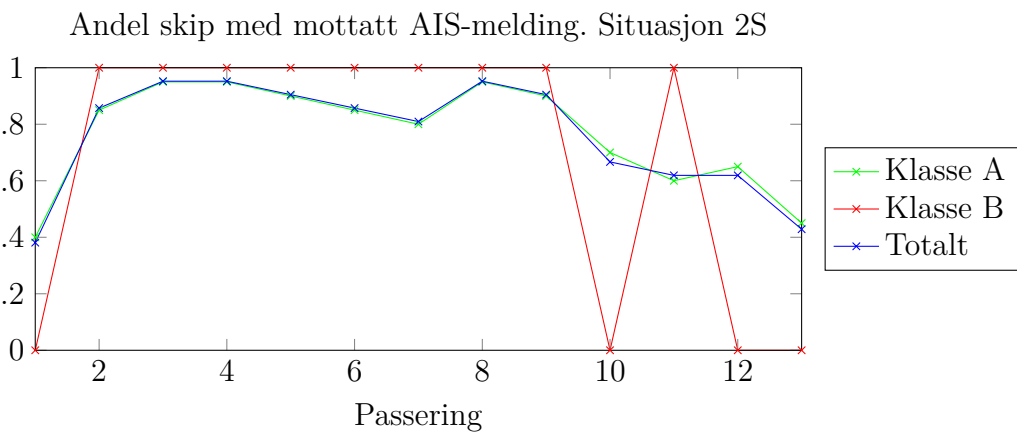
Andel skip med mottatte meldinger

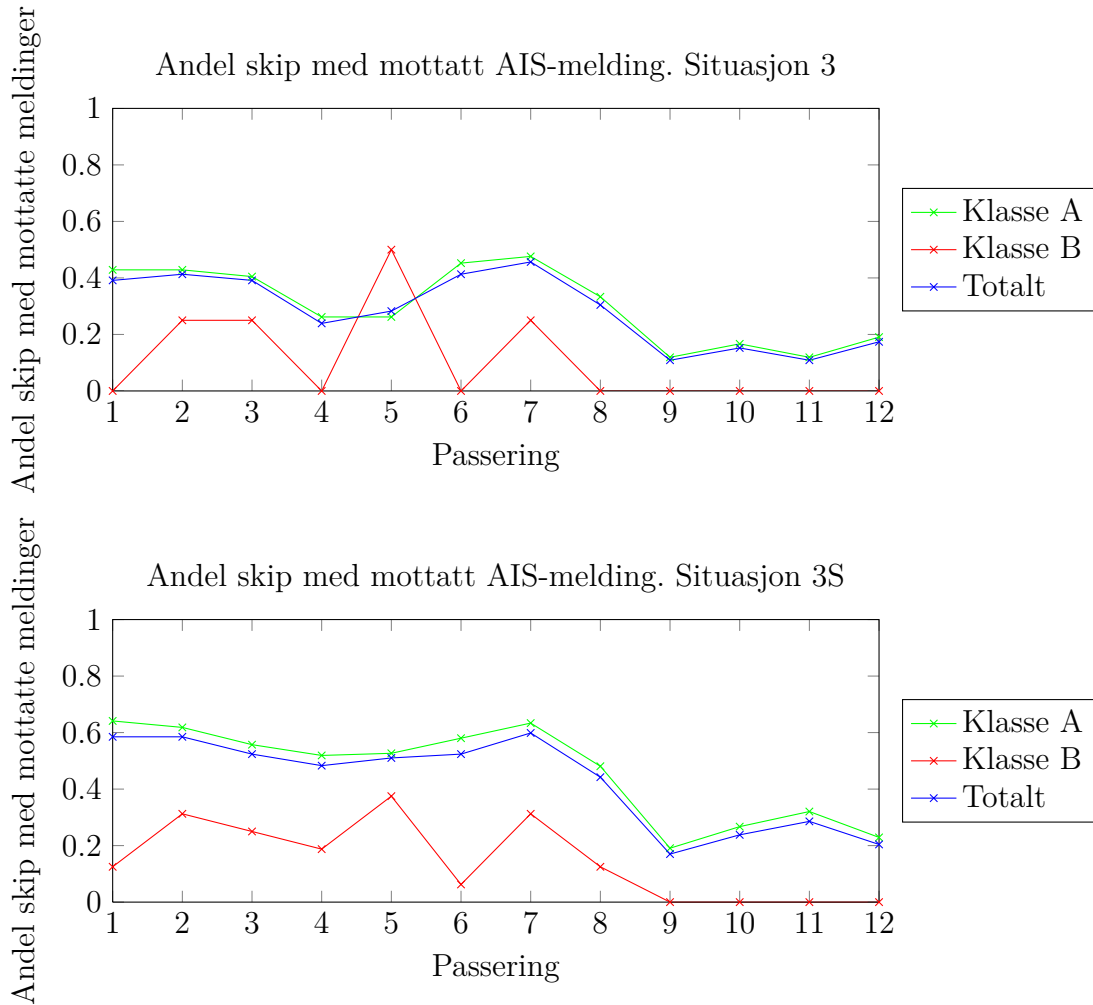


Andel skip med mottatte meldinger



Andel skip med mottatte meldinger





D Taylorutvikling av $r(\nu; \phi)$

Ved å la $x = [\phi^T, v^T]^T = [\phi_1, \phi_2, \phi_3, v_{1i}, v_{2i}]^T$ være sammensetningen av ϕ og v kan utviklingen skrives som

$$r_j(x) = r_j(0) + \sum_{k=1}^5 \left. \frac{\partial r_j(x)}{\partial x_k} \right|_{\substack{v=0 \\ \phi=0}} x_k + \frac{1}{2!} \sum_{k=1}^5 \sum_{l=1}^5 \left. \frac{\partial^2 r_j(x)}{\partial x_k \partial x_l} \right|_{\substack{v=0 \\ \phi=0}} x_k x_l, \quad j = 1, 2. \quad (130)$$

Funksjonen $r_j(x)$ er naturligvis komponent j til $r(v, \phi)$ skrevet med den nye parameteren. Beregningen av utviklingen forenkles drastisk ved å observere at begge komponentene til r er på formen $r(x) = \arctan(h_j(x)/g(x))$ hvor $h_j(0) = 0$

og $g(0) = 1$. Dette gir at den førstederiverte kan skrives som

$$\begin{aligned} \left. \frac{\partial r_j(x)}{\partial x_k} \right|_{\substack{v=0 \\ \phi=0}} &= \left(1 + \left(\frac{h(x)}{g(x)} \right)^2 \right)^{-1} \left. \frac{\frac{\partial h_j(x)}{\partial x_k} g(x) - h_j(x) \frac{\partial g(x)}{\partial x_k}}{g(x)^2} \right|_{\substack{v=0 \\ \phi=0}} \\ &= \left. \frac{\partial h_j(x)}{\partial x_k} \right|_{\substack{v=0 \\ \phi=0}}. \end{aligned} \quad (131)$$

Beregningen av de andrederiverte forenkles betraktelig når en legger merke til at $\left(1 + (h(x)/g(x))^2 \right)^{-1}$ blir 1 ved innsetting, og den deriverte av dette gir en faktor $h(x)/g(x)$ utenfor som blir 0 ved innsetting. Dermed vil de andrederiverte være på formen

$$\left. \frac{\partial^2 r_j(x)}{\partial x_k \partial x_l} \right|_{\substack{v=0 \\ \phi=0}} = \left. \frac{\partial}{\partial x_l} \frac{\frac{\partial h_j(x)}{\partial x_k} g(x) - h_j(x) \frac{\partial g(x)}{\partial x_k}}{g(x)^2} \right|_{\substack{v=0 \\ \phi=0}}. \quad (132)$$

Denne derivasjonen vil gi en brøk med $g(x)^4$ i nevner, som er 1 ved innsetting. Telleren er alt som står igjen, og den blir

$$\begin{aligned} &\left(\frac{\partial^2 h_j(x)}{\partial x_k \partial x_l} g(x) + \frac{\partial h_j(x)}{\partial x_k} \frac{\partial g(x)}{\partial x_l} - \frac{\partial h_j(x)}{\partial x_l} \frac{\partial g(x)}{\partial x_k} - h_j(x) \frac{\partial^2 g(x)}{\partial x_k \partial x_l} \right) g(x)^2 \Big|_{\substack{v=0 \\ \phi=0}} \\ &\quad - \left(\frac{\partial h_j(x)}{\partial x_k} g(x) - h_j(x) \frac{\partial g(x)}{\partial x_k} \right) 2g(x) \frac{\partial g(x)}{\partial x_l} \Big|_{\substack{v=0 \\ \phi=0}} \end{aligned} \quad (133)$$

Innsetting av $h_j(x) = 0$ og $g(x) = 1$ gir

$$\left(\frac{\partial^2 h_j(x)}{\partial x_k \partial x_l} + \frac{\partial h_j(x)}{\partial x_k} \frac{\partial g(x)}{\partial x_l} - \frac{\partial h_j(x)}{\partial x_l} \frac{\partial g(x)}{\partial x_k} - 2 \frac{\partial h_j(x)}{\partial x_k} \frac{\partial g(x)}{\partial x_l} \right) \Big|_{\substack{v=0 \\ \phi=0}}, \quad (134)$$

som forkortes til

$$\left. \frac{\partial^2 r_j(x)}{\partial x_k \partial x_l} \right|_{\substack{v=0 \\ \phi=0}} = \left(\frac{\partial^2 h_j(x)}{\partial x_k \partial x_l} - \frac{\partial h_j(x)}{\partial x_l} \frac{\partial g(x)}{\partial x_k} - \frac{\partial h_j(x)}{\partial x_k} \frac{\partial g(x)}{\partial x_l} \right) \Big|_{\substack{v=0 \\ \phi=0}}. \quad (135)$$

Etter å ha lest av uttrykkene for h_1 , h_2 og g fra likning (52) kan de nødvendige deriverte beregnes. Resultatene av førsteordens derivasjoner etter innsetting er vist i tabell 5. Veldig mange av de andrederiverte er null, resten er vist i tabell 6.

Innsetting i likning (130) gir

$$r(v; \phi) \approx 0 + \begin{bmatrix} \phi_x + v_{1i} \\ -\phi_y + v_{2i} \end{bmatrix} + \frac{1}{2} \begin{bmatrix} -2\phi_z v_{2i} + 2\phi_y \phi_z \\ 2\phi_z v_{1i} + 2\phi_x \phi_z \end{bmatrix}, \quad (136)$$

som er det samme som likning (54). Faktoren 2 i siste leddene kommer fra at både $\partial^2 h_j / \partial x_k \partial x_l$ og $\partial^2 h_j / \partial x_l \partial x_k$ er med i summen.

$a \backslash b$	$x_1 = \phi_x$	$x_2 = \phi_y$	$x_3 = \phi_z$	$x_4 = v_{1i}$	$x_5 = v_{2i}$
h_1	1	0	0	1	0
h_2	0	-1	0	0	1
g	0	0	0	0	0

Tabell 5: Oversikt over $\partial a / \partial b$ innsatt $\phi = 0$ og $v = 0$.

a, b	$x_3, x_4 = \phi_z, v_{2i}$	$x_3, x_5 = \phi_z, v_{1i}$	$x_1, x_3 = \phi_x, \phi_z$	$x_2, x_3 = \phi_y, \phi_z$
$\partial^2 h_1 / \partial a \partial b$	-1	0	0	1
$\partial^2 h_2 / \partial a \partial b$	0	1	1	0

Tabell 6: Oversikt over andrederiverte som ikke er null.

E Korrelasjonskoeffisienter for situasjonene

$\text{Kov}(X, Y) = E(X - E X)(Y - E Y)$ er i tabell 7 estimert ved

$$\text{Kov}(X, Y) = \frac{1}{n-1} \sum_i (x_i - \bar{x})(y_i - \bar{y}), \quad \bar{x} = \frac{1}{n} \sum_i x_i, \quad \bar{y} = \frac{1}{n} \sum_i y_i. \quad (137)$$

Parene $(x_i, y_i), i = 1, \dots, n$ er observasjoner av de tilfeldige variablene X og Y .

Situasjon	xy SR	xz SR	yz SR	xy FS	xz FS	yz FS
Sit1	-0.41	0.0018	-0.67	0.035	-0.23	0.16
Sit1S	-0.026	-0.097	0.35	-0.027	-0.097	0.092
Sit2	-0.065	-0.073	0.18	-0.039	0.04	0.22
Sit2S	-0.1	0.21	0.0024	-0.1	0.28	0.0075
Sit3	0.064	0.14	0.61	-0.21	-0.27	0.52
Sit3S	0.21	0.75	0.32	-0.066	0.23	0.06

Tabell 7: Korrelasjonskoeffisienter for situasjonene. De to koordinatene angir hvilken korrelasjonskoeffisient det gjelder. For eksempel betyr xy SR at $\text{Kov}(\hat{\phi}_x^{GNN}, \hat{\phi}_y^{GNN}) / ((\text{Std } \hat{\phi}_x^{GNN}) (\text{Std } \hat{\phi}_y^{GNN}))$ hvor kovariansen og standardavvikene er beregnet ut fra resultatene fra SR-metoden.