



FFI-rapport 2014/01093

Vurdering av måloppnåelse i nasjonale operasjoner – metode og teoretisk grunnlag



Stein Malerud



Vurdering av måloppnåelse i nasjonale operasjoner – metode og teoretisk grunnlag

Stein Malerud

Forsvarets forskningsinstitutt (FFI)

22. september 2014

FFI-rapport 2014/01093

3849

P: ISBN 978-82-464-2436-1

E: ISBN 978-82-464-2437-8

Emneord

Måloppnåelse

Virksomhetsstyring

Output

Outcome

Effektivitet

Godkjent av

Frode Rutledal

Prosjektleder

Espen Skjelland

Avdelingsjef

Sammendrag

Denne rapporten beskriver en metode for vurdering av måloppnåelse. Metoden gir en strukturert fremgangsmåte basert på en resultatkjede med fem nivåer: *input*, *aktiviteter*, *output*, *direkte outcome* og *indrekte outcome*. De tre første nivåene er knyttet til virksomheten, mens de to siste beskriver hvordan virksomheten påvirker sine omgivelser. *Outcome* er i henhold til definisjonen benyttet i denne rapporten, en tilstand som verdsettes av beslutningstakere og andre interessenter.

En metode for vurdering av måloppnåelse må kunne støtte de ulike trinnene i en evalueringsprosess, og være anvendbar for daglige operasjoner så vel som for operasjoner under krise og krig. Evalueringsprosessen beskrevet i *Comprehensive Operations Planning Directive (COPD)* [1] er først og fremst utviklet for å støtte evalueringer av militære operasjoner under krise og krig, og i mindre grad for daglige operasjoner. Vi har derfor valgt å basere vår metode på en resultatkjedemodell som anvendes i (daglig) virksomhetsstyring, men hvor begreper og fremgangsmåte i så stor grad som mulig er konsistent med COPD. Et annet viktig aspekt ved metoden er at den skiller klart mellom den operative virksomhetens leveranser (*output*) og de målene virksomheten skal oppnå (*outcome*). Metoden hjelper til med å identifisere mål på ulike resultatnivå, og bidrar til å gi en sporbar sammenheng mellom mål og oppgaver på høyere nivå (politisk og strategisk) og mål og resultater på lavere nivå (operasjonelt og taktisk).

Det er viktig at beslutningstaker benytter denne type vurderinger i sin beslutningsprosess. En forutsetning for dette er at resultatene er relevante og troverdige. Metoden som beskrives i denne rapporten bidrar til å underbygge relevans og troverdighet ved at man fra starten av legger vekt på å finne de spørsmålene og problemene som er relevante for beslutningsprosessen. Derneft gir den en strukturert og sporbar prosess som tar utgangspunkt i overordnede mål og bryter disse ned i hensiktsmessige resultatmål som kan måles/observeres. I tillegg beskriver rapporten en metode for å synliggjøre og håndtere usikkerhet som en viktig del av beslutningsgrunnlaget.

English summary

This report describes a method for assessing attainment of goals and objectives of military operations. The method gives a structured approach based on a result chain with five levels: *input*, *activities*, *output*, *direct outcome* and *indirect outcome*. The three first levels are related to own, controllable activities, while the two last levels describe how produced actions/services influence the environment in accordance to the goals and objectives.

For an assessment method to be relevant, it needs to support all the levels of a typical assessment process. In addition it must be applicable to daily, routine operations as well as to operations in crisis and at war. The evaluation process described in *Comprehensive Operations Planning Directive* (COPD) [1] is first and foremost designed to support assessments of operations in crisis and at war, and to a lesser degree, assessment of daily operations. Thus, we have decided to base our method on a result chain model applied in (daily) result-based management, however, the method and notions are as far as possible consistent with the prescribed approach in COPD. Another important aspect of the method is that it distinguishes between the output of actions and the attainment of goals and objectives (*outcome*). The method supports the process of identifying and organizing *outputs* and *outcomes* on different levels in the organization.

It is important that the decision maker applies this kind of assessments in his decision process. A premise for this is that the results are perceived as useful and reliable. The method described in this report contributes to ensure relevance and to build confidence to the results. The assessment process is initiated by problem structuring to find and set focus on relevant goals and objectives. Further, the method offers a structured and transparent approach for developing result metrics and parameters together with a method for describing and managing uncertainties.

Innhold

	Forord	6
1	Innledning	7
2	Teoretisk grunnlag	8
2.1	Prosessmodell for vurdering av måloppnåelse	8
2.2	Resultatkjedemodell	9
2.3	Plan- og evalueringsprosesser i NATO-publikasjoner	12
2.4	Resultatmålene <i>output</i> og <i>outcome</i>	13
2.5	Måling av ytelse og effektivitet	16
2.6	Modellering av verdi og nytte	17
3	Metode for vurdering av måloppnåelse	20
3.1	Begreper og definisjoner	21
3.2	Anbefalt metode	23
3.2.1	Utlede oppgaver og mål	23
3.2.2	Utlede måleparametere og etablere en datainnsamlingsplan	24
3.2.3	Datainnsamling og analyse	24
4	Eksempel på anvendelse av metoden	25
4.1	Oppgaver og mål	26
4.2	Måleparametere og datainnsamlingsplan	27
4.3	Datainnsamling og analyse	29
4.3.1	Tilstedeværelse	30
4.3.2	Deteksjon og identifisering av tilfeldig avvik	31
4.3.3	Følge avvik	33
4.3.4	Beregning av resultatmål	33
4.3.5	Usikkerhetsvurderinger	33
5	Oppsummering og avslutning	34
	Litteratur	36
	Forkortelser	39
	Vedlegg A Om usikkerhet og validitet	40
	Vedlegg B Modell for beregning av deteksjonssannsynlighet	45

Forord

Denne rapporten sammen med rapporten “Vurdering av måloppnåelse i nasjonale operasjoner – forslag til mål og måleparametere for daglige operasjoner” [2] utgjør dokumentasjonen fra CD&E-aktiviteten EP 1243 Vurdering av måloppnåelse i nasjonale operasjoner. Arbeidet er utført av Forsvarets forskningsinstitutt (FFI) for Forsvarets operative hovedkvarter (FOH).

Dokumentasjon og resultater fra denne aktiviteten er et innspill til arbeidet med å utvikle og oppdatere Forsvarets operative planverk. Det overordnede målet har vært å bidra til å etablere et konsept for vurdering av måloppnåelse i nasjonale operasjoner. Arbeidet har blant annet bestått av å utvikle en metode for vurdering av måloppnåelse, og å gi et innspill til mål og måleparametere for daglige operasjoner.

Denne rapporten beskriver det metodiske grunnlaget for utvikling av mål og måleparametere, mens i [2] anvendes metoden for å utvikle mål og måleparametere for daglige operasjoner.

Rapporten bidrar også til det videre arbeidet med vurdering av måloppnåelse i militære operasjoner ved FFI.

1 Innledning

En viktig del av en beslutningsprosess er å vurdere i hvilken grad aktiviteter og utførte handlinger bidrar til at man når målene for virksomheten. I en militær kontekst innebærer dette å vurdere hvordan egne operasjoner og handlinger bidrar til progresjon og måloppnåelse i forhold til planlagte mål. Typiske spørsmål man ønsker svar på er hvor godt operasjonene gjennomføres, om man gjør de riktige tingene og om dette er godt nok i forhold til mål og ambisjoner.

Vurderinger av måloppnåelse skal bidra til et oppdatert og relevant beslutningsgrunnlag som støtter opp under god situasjonsbevissthet. Gjennomføres dette på en strukturert måte kan resultatene også bidra til systematisk læring i form av å opparbeide kunnskap og erfaring som kan benyttes i fremtidige beslutningsprosesser.

Hensikten med denne rapporten er å gi et innspill til prosessen med å etablere et konsept for vurdering av måloppnåelse i nasjonale operasjoner. Rapporten beskriver en metode som gir en strukturert fremgangsmåte for å utarbeide relevante mål og måleparametere, og benytter en resultatkjede for å sette disse i sammenheng. Den gir også et utfyllende eksempel på bruk av metoden i forbindelse med maritim overvåking.

En vurdering av måloppnåelse er vellykket i den grad resultatene er riktige og relevante, slik at beslutningstaker benytter resultatene i sin beslutningsprosess. Metoden som beskrives i denne rapporten bidrar til å underbygge relevans og troverdighet. For det første legger metoden fra starten av vekt å finne de spørsmålene og problemene som er relevante for beslutningsprosessen. For det andre gir den en strukturert og sporbar prosess som tar utgangspunkt i overordnede mål og bryter disse ned i hensiktsmessige resultatmål som kan måles/observeres. Metoden er i prinsippet en multi-metode¹ hvor en kombinerer ulike metoder – både kvantitative og kvalitative – for å analysere måleresultatene opp mot beslutningstakers informasjonsbehov. En viktig del av metoden er å ha et bevisst forhold til usikkerheten i datagrunnlaget og i analysene, og å kunne svare på hva som er gjort for å validere metoder, modeller og inputdata.

En kjent utfordring innen vurderinger av måloppnåelse, er å kunne relatere oppnådde resultater på forskjellige nivå til hverandre [1;3]. I dette ligger det blant annet å kunne sette egne aktiviteter og leveranser i sammenheng med oppnåelse av kortsiktige, direkte mål og mer langsiktige, overordnede mål. Spesielt er det utfordrende å underbygge kausale relasjoner mellom egne handlinger og oppnåelse av langsiktige mål. Her vil det ofte være hensiktsmessig å benytte et bredt spekter av informasjonskilder som omfatter både kvantitative data og mer kvalitative vurderinger.

Comprehensive Operations Planning Directive (COPD) [1] er under innføring i NATO og i Forsvaret. COPD beskriver først og fremst hvordan man skal gjennomføre en planprosess på strategisk og operasjonelt nivå, men inneholder også et kapittel som omhandler evaluering av

¹ Innen operasjonsanalyse benyttes begrepet multi-metode om en gitt sammensetning av metoder som fungerer godt sammen og som brukes for å løse et problem.

måloppnåelse (*assessment*) opp mot mål i planen. Rapporten diskuterer hvor godt egnet denne prosessen er til å støtte vurderinger av måloppnåelse i daglige operasjoner. Rapporten ser også på hvordan prosessen eventuelt kan gjøres enklere og tilpasses de behovene man har for vurdering av måloppnåelse i ulike typer operasjoner.

Rapporten er organisert i følgende kapitler:

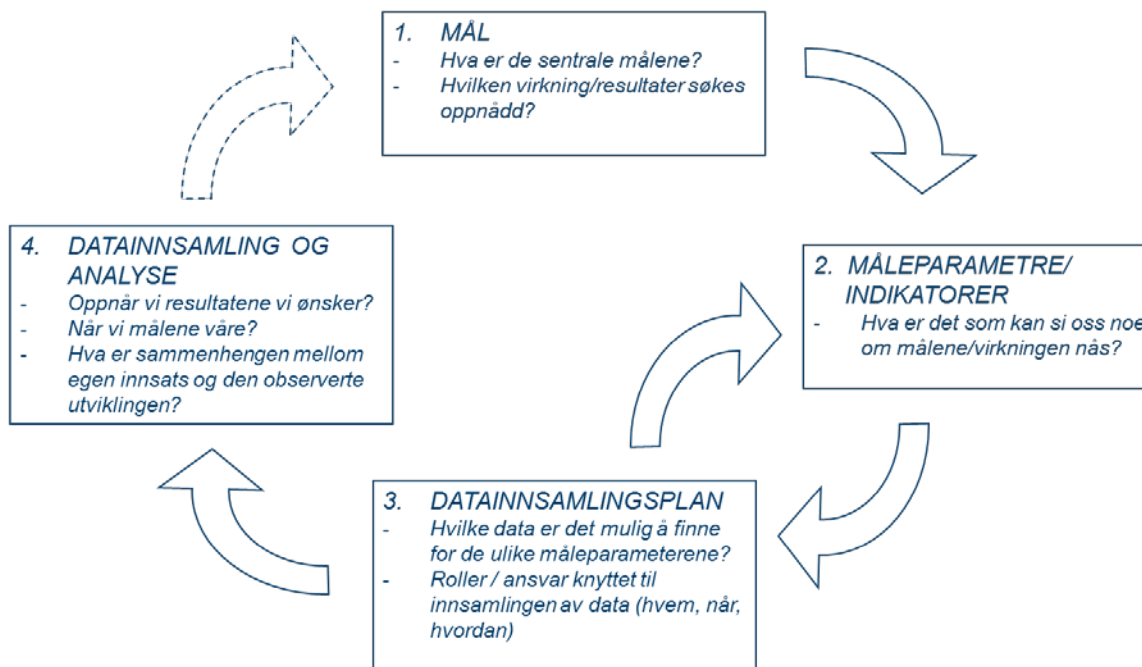
- Kapitel 2 gir det teoretiske grunnlaget for metode- og modellutviklingen
- Kapitel 3 beskriver metoden for vurdering av måloppnåelse
- Kapitel 4 gir et eksempel på anvendelse av metoden med resultater for maritim overvåkning
- Kapitel 5 oppsummer innholdet i rapporten

2 Teoretisk grunnlag

I dette kapitlet presenteres det teoretiske grunnlaget for utviklingen av en metode for vurdering av måloppnåelse. Det tas utgangspunkt i en generell prosessmodell for vurdering av måloppnåelse. Deretter presenteres og diskuteres forskjellige begreper, metoder og modeller som kan bidra til å støtte opp under gjennomføringen av prosessen.

2.1 Prosessmodell for vurdering av måloppnåelse

Figur 2.1 viser hovedtrinnene i en prosessmodell for vurdering av måloppnåelse [4].



Figur 2.1 Prosessmodell for vurdering av måloppnåelse.

Mål: Før man går i gang med å utlede måleparametere må det fremskaffes en oversikt over hvilke mål for virksomheten som det skal evalueres opp mot. Et mål er forbundet med en ønsket tilstand som man vil nå gjennom å utføre aktiviteter. Som regel tas det utgangspunkt i formålet med virksomheten og overordnede mål som brytes videre ned i hensiktsmessige mål og delmål. Eksempelvis kan Forsvarets oppgaver, og forsvarspolitiske og militærstrategiske mål benyttes som overordnede mål, mens operasjonelle mål, konsepter og planer gir mer detaljerte mål for den operative virksomheten.

Måleparametere: Måleparametere benyttes til å måle i hvilken grad virksomheten når sine mål. Måleparameterne måler tilstander relatert til omgivelsene og systemene² man ønsker å påvirke gjennom virksomhetens aktiviteter. Det utledes også måleparametere for å måle ytelsen til egen virksomhet.

Datainnsamlingsplan: Hensikten er å lage en plan for innsamling av data til måleparameterne utledet i forrige trinn. Planen bør inneholde en entydig definisjon av parameterne, en beskrivelse av hvordan de skal måles (målemetode, kilde, når og hvor ofte) og hvem som har ansvaret. Det kan være nødvendig å gå tilbake for å justere på settet av måleparametere fordi det kan være vanskelig eller lite kosteffektivt å fremskaffe det nødvendige datagrunnlaget. Det finnes flere aktuelle datakilder som for eksempel observasjoner/målinger, bruk av ekspertvurderinger (*subject matter experts* – SME) og spørreundersøkelser.

Datainnsamling og analyse: Innebærer å samle inn, tolke og analysere data som grunnlag for vurderinger av måloppnåelse. Når vi målene? Hva er sammenhengen mellom innsats, leveranser av produkter og tjenester og måloppnåelse? Resultatene av analysen kan føre til behov for å endre planen og/eller justeringer av mål. Aktiviteten omfatter også rapportering og presentasjon av relevante resultater for beslutningstakere.

2.2 Resultatkjedemodell

I vurderingen av måloppnåelse er hensikten først og fremst å kunne si noe kvalifisert om hvorvidt virksomheten evner å nå sine overordnede mål. Eksempelvis er hensikten med den operative virksomheten i Forsvaret relatert til å løse Forsvarets oppgaver slik disse er beskrevet i Prop.73S [5].

Direktoratet for økonomistyring (DFØ) har utviklet en resultatkjedemodell for statlig virksomhet [3]. En versjon av denne modellen benyttes av Forsvaret i virksomhetsstyringen, se direktiv for virksomhets- og økonomistyring (DIVØ) [6]. Modellen vist i figur 2.2 er basert på DFØ sin modell.

² Det finnes mange definisjoner av begrepet *system*. En hensiktsmessig definisjon for vår bruk er: Et *system* er en samling av komponenter (materiell, mennesker,..) som gjensidig påvirker hverandre og som danner en integrert helhet (oversatt fra Merriam-Webster: www.britannica.com).



Figur 2.2 Resultatkjedemodell.

Modellen består av fem trinn og viser årsakssammenhengen mellom innsatsfaktorer, leveranser og effekter på ulike nivå. Den tar utgangspunkt i de innsatsfaktorene som virksomheten må fremskaffe (materiell, personell, kunnskap og kompetanse etc.) for å kunne produsere leveranser i form av produkter, tjenester og handlinger. Leveransene har en effekt/virkning på omgivelsene til virksomheten – både tilsiktede og utilsiktede. De tilsiktede effektene/virkningene er planlagte tilstandsendringer som skal bidra til at virksomheten når sitt formål. I veilederen for resultatmåling defineres en *effekt* som en forandring i tilstand hos brukere eller i samfunnet som har oppstått som følge av virksomhetens tiltak.

Differensieringen mellom direkte effekter og samfunns-effekter synes å være spesielt nyttig for statlig, ikke-kommersiell virksomhet hvor formålet ofte er knyttet til de langsiktige samfunns-messige effektene. Leveransene er normalt kontrollerbare av virksomheten, mens effektene, og da spesielt de langsiktige, i mindre grad er kontrollerbare (påvirkes, men ikke kontrolleres). Både leveransene og effektene er resultater i resultatkjeden. Leveransene er resultater knyttet til hvor godt man utnytter egne ressurser (ytelse og produktivitet), mens effekter er resultater knyttet til omgivelsene som man ønsker å påvirke (effektivitet).

I [3] understrekes viktigheten av å synliggjøre kausalitet mellom observerte effekter og virksomhetens tiltak. Modellen i figur 2.2 skiller mellom direkte effekter (brukereffekter i DFØ sin modell) og mer langsiktige samfunns-effekter. For de direkte effektene bør det være klare kausale sammenhenger mellom virksomhetens leveranser og observerte effekter. I praksis betyr dette at man bør underbygge kausale sammenhenger ved hjelp av statistiske modeller som beskriver usikkerheten i relasjonene (se diskusjon rundt usikkerhet i vedlegg A).

For Forsvarets operative virksomhet er det nyttig å skille mellom daglige operasjoner rettet mot overvåkning, myndighetsutøvelse og suverenitetshevdelse, og krise og krig. Under krise og krig utvikles planer for hvordan den uønskede situasjonen skal håndteres samtidig som Forsvaret normalt har en sentral rolle. I denne type situasjoner vil det sannsynligvis være enklere å relatere egne aktiviteter og handlinger til måloppnåelse, i alle fall til de mer direkte målene/effektene. Under daglig drift (vedvarende operasjoner) kan det være mer utfordrende å sannsynliggjøre kausale sammenhenger.

For de mer langsiktige, samfunnsmessige effektene er det vanligvis større usikkerhet knyttet til sammenhengen mellom egne leveranser og grad av måloppnåelse. De samfunnsmessige effektene er normalt ikke relatert til et isolert, veldefinert system som kun påvirkes av virksomhetens aktiviteter og leveranser, men også av en mengde eksterne forhold som det kan være vanskelig å

vurdere betydningen av. Med eksterne forhold mener vi forhold som ligger utenfor virksomhetens kontroll. Eksempelvis er det utfordrende å vurdere hvilken betydning Forsvarers daglige operative virksomhet har for å sikre en langsiktig, bærekraftig utvikling av naturressursene i norske havområder. For å kunne analysere dette problemet må man anta at det er en kausal sammenheng mellom KVs tilstedeværelse, kontrollvirksomhet og håndhevelse av lover, og etterlevelse av (respekt for) lover og regler blant fiskere og videre til en bærekraftig utvikling av fiskeressursene. Selv om denne hypotesen kan begrunnes fornuftsmessig, er det svært vanskelig å underbygge den ved hjelp av kvantitative, statistiske metoder.

Forsvaret driver en virksomhet som i liten grad påvirker befolkningen direkte i motsetning til for eksempel skole- og helsesektoren hvor de direkte effektene kan relateres til brukergrupper som elever og pasienter. Forsvarets virksomhet er mer rettet mot samfunnet som helhet ved å være en sikkerhetsgaranti mot ytre trusler mot landets territorium og interesser. Eksempelvis er fravær av konflikt, og evne til å kunne håndtere potensielle trusler, viktige suksesskriterier for Forsvarets virksomhet. Dette er i liten grad målbart og introduserer en betydelig utfordring hva gjelder sentrale årsak-virkningsforhold mellom virksomhetens aktiviteter og overordnede mål. Usikkerheten i de kausale sammenhengene lar seg vanskelig håndtere ved bruk av formelle, kvantitative metoder. Ikke desto mindre bør man kunne underbygge og sannsynliggjøre sentrale kausale sammenhenger kvalitativt for i hele tatt å kunne vurdere samfunnsmessige effekter. I DFØ sin veileder [3] nevnes tre tiltak som støtter dette:

- Se på samvariasjon (kovarians) mellom tiltak og observerte resultater.
- Kontrollere at det er riktig rekkefølge (kausalt) mellom tiltak og effekt/virkning.
- Ha kontroll med andre, eksterne forhold som påvirker resultatet.

Det første tiltaket krever at man har tidsserier med relevante data, men samvariasjon er i seg selv ikke tilstrekkelig for å kunne påstå at det er en kausal sammenheng. I tillegg må man kreve at det er riktig rekkefølge mellom tiltak og observerte effekter. Det siste kulepunktet er ofte vanskelig å gjennomføre i praksis, spesielt hvis man har som mål å påvirke et komplekst system i et mer langsiktig perspektiv. Ideelt sett ønsker vi å kunne isolere den påvirkningen/effekten som Forsvarts virksomhet har på resultatoppnåelsen [6].³ Men det er nok en mer realistisk ambisjon å forsøke å få oversikt over noen av de viktigste eksterne forholdene, og på bakgrunn av dette identifisere de viktigste driverne for utviklingen. Ved å studere disse sammenhengene over tid, eventuelt kombinert med andre analyser som for eksempel scenarioanalyse [7], er det mulig å si noe om hvor sterke de ulike kausale sammenhengene er.

³ For en diskusjon rundt dette, se for eksempel Atkinson, 2005.

2.3 Plan- og evalueringsprosesser i NATO-publikasjoner

I dette kapitlet gis en kortfattet gjennomgang av hvordan vurdering av måloppnåelse (*assessment*) er beskrevet i NATO-dokumentene COPD [1] og NATOs *Operations Assessment Handbook* (NOAH) [8]⁴. COPD beskriver plan- og evalueringsprosesser på strategisk og operasjonelt nivå.

COPD er først og fremst utviklet for å støtte planlegging og evaluering under krise og krig, det vil si ved et større avvik fra normalsituasjonen, og ikke vurderinger av resultatoppnåelse under daglig drift (normaldrift). Planutviklingsprosessen beskrevet i COPD er egnet for å utvikle planer for militære operasjoner, hvor en ønsker gå fra en uakseptabel tilstand til en ønsket slutttilstand. Det er usikkert hvor god denne prosessen er med hensyn til vurderinger av måloppnåelse under daglige operasjoner, hvor målet normalt vil være å opprettholde normalsituasjonen og håndtere mindre avvik fra denne.

I COPD benyttes følgende begreper og definisjoner:

- *Action*: “The process of doing something to achieve an aim”.
- *Effect*: “A change in the state of a system (or system element), that results from one or more actions, or other causes”.⁵
- *Decisive condition* (DC): “A combination of circumstances, effects, or a specific key event, critical factor, or function that when achieved allows commanders to gain a marked advantage over an opponent or contribute materially to achieving an operational objective” (COPD/AJP-01⁶).
- *Objective*: “A clearly defined and attainable goal to be achieved”.
- *End state*: “The North Atlantic Council (NAC) state of conditions that defines an acceptable concluding situation for NATO’s involvement”.⁷

Operasjonelle og militærstrategiske mål (*operational and military strategic objectives*) er tilstander som skal oppnås på veien mot ønsket slutttilstand. Disse målene er i planprosessen brutt ned og konkretisert i delmål (DC-er). Et mål kan ha én eller flere assosierte DC-er som skal oppnås ved å skape ønskede effekter gjennom handlinger (*actions*). En effekt kan i henhold til definisjonen forstås som en forandring i en tilstand relevant for oppnåelse av én eller flere DC-er.

Vurdering av progresjon og resultatoppnåelse på strategisk og operasjonelt nivå refereres til i COPD som *operations assessment*. På operasjonelt nivå gjøres det vurderinger opp mot mål nedfelt i en operasjonsplan (OPLAN). Vurderingene gjøres rundt to aspekter. Det første dreier seg om vurderinger av om man når målene som er satt for de militære operasjonene (løser vi opp-

⁴ NOAH er utarbeidet for å gi en mer detaljert innføring i vurdering av måloppnåelse i operasjoner (*Operations Assessment*). Beskrivelsen i COPD er basert på denne.

⁵ Begrepet *effekt* er definert i Forsvarets fellesoperative doktrine (FFOD) [9] som: “Resultat eller forandring, ofte i form av en atferdsendring hos motstanderen eller andre aktører”.

⁶ *Allied Joint Doctrine* (AJP-01(D)), desember 2010.

⁷ FFOD: *Slutttilstand* er en beskrivelse av den politiske og/eller militære situasjonen som ønskes etablert når en operasjon er avsluttet [9]

draget?). Dette innebærer kontinuerlig monitorering og vurdering av effekter, DC-er og mål (*objectives*) spesifisert i operasjonsplanen. I tillegg kan også mål relatert til eksterne forhold vurderes.⁸ Samlet bidrar disse vurderingene til å støtte opp under anbefalinger til beslutningstaker om å utarbeide ny *direction and guidance* for operasjonen. Det andre aspektet fokuserer på gjennomføringen av operasjonene og dekker kort til middels tidshorisont. Her gjøres det vurderinger av status på aktuelle effekter som videre leder til DC-er som er lagt ut i tid og rom langs definerte operasjonslinjer. Vurderingene skal støtte en dag-til-dag beslutningssyklus og kan føre til anbefalinger om å forandre operasjonsplanen gjennom operasjonsordre som *Fragmentation Order* (FRAGO) og *Joint Coordination Order* (JCO).

Vurdering av måloppnåelse krever at det utarbeides relevante måleparametere for aktiviteter og resultater. I COPD benyttes *Measure of effectiveness* (MOE) til å måle resultatoppnåelse og *Measure of performance* (MOP) til å måle ytelse. Definisjonene er:

- MOE: “A metric used to measure a current system state”.
- MOP: “A metric used to determine the accomplishment of actions”.

På operasjonelt nivå benyttes MOE-ene til å måle progresjon og resultatoppnåelse opp mot ønskede effekter, DC-er og til dels operasjonelle mål, slik de fremkommer i operasjonsplanen. MOP-ene benyttes til å måle på resultater relatert til gjennomføringen av aktiviteter og planlagte handlinger.

Hvis vi sammenlikner resultatkjedemodellen presentert i kapittel 2.2 med begrepsbruken i COPD, vil handlinger (*actions*) tilsvare både aktiviteter og leveranser (produkter/tjenester). Effekter, delmål (DC-er) og til en viss grad mål (*objectives*) vil typisk være relatert til direkte effekter, mens mål på høyere nivå og slutttilstand normalt vil være forbundet med langsiktige mål (samfunnseffekter).

2.4 Resultatmålene *output* og *outcome*

Begrepet effekt er i henhold til definisjonene gitt i [1;3] forbundet med en tilstandsending i systemet man ønsker å påvirke. For å kunne vurdere måloppnåelse er det nødvendig å bestemme i hvilken tilstand systemet faktisk befinner seg i på et gitt tidspunkt, og vurdere dette resultatet opp mot fastsatte mål.

⁸ For eksempel politiske, militære, økonomiske, sosiale, infrastrukturmessige og informasjonsmessige forhold (*Political, Military, Economic, Social, Infrastructure and Information – PMESII*).

I [4;10-12] benyttes begrepene *input*, *output* og *outcome* for å måle på tilstander og resultater relatert til innsatsmidler (*input*), leveranser (*output*) og måloppnåelse (*outcome*). Vi vil benytte følgende definisjoner basert på DFØ sin veileder og OECD DAC⁹ [12]:

- *Input*: utgjør de finansielle, materielle og menneskelige ressurser som en organisasjon må fremskaffe for å utføre sine prosesser og aktiviteter.
- *Output*: er de tjenestene, produktene eller handlingene som produseres av virksomheten.

Input tilsvarer innsatsfaktorene i resultatkjedemodellen og *output* er de produktene og tjenestene som virksomheten leverer, og som er resultat av virksomhetens aktiviteter. Målinger av *output* sier noe om virksomhetens ytelse, og forholdet mellom *input* og *output* er et mål på produktiviteten til virksomheten (se for eksempel [13;14]). *Output* er et resultat, men sier i utgangspunktet ikke noe om virksomheten når sine overordnede mål.

Resultatmålet *outcome* benyttes for å vurdere i hvilken grad virksomhetens aktiviteter og leveranser bidrar til at virksomheten når sine overordnede mål. Med andre ord, om man gjør de riktige tingene for å realisere virksomhetens formål.

Outcome kan defineres på forskjellige måter.¹⁰ I henhold til [10] er det to hovedretninger i litteraturen. *Outcome* kan ses på som en endring i tilstand, det vil si en effekt. Eller, vi kan se på *outcome* som en tilstand som verdsettes av interessenter. Begge definisjonene er i bruk. Vi har valgt å benytte den sistnevnte definisjonen for å kunne skille mellom en tilstand og en endring i tilstand:

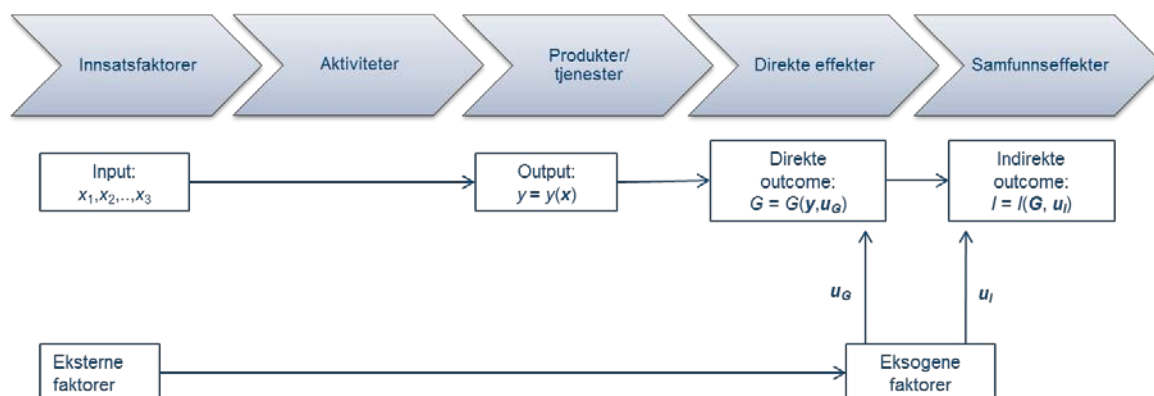
- *Outcome*: er en tilstand (relatert til omgivelsene og virksomhetens formål) som verdsettes av beslutningstakere og andre interessenter [10].
- *Effekt*: er en forandring i tilstanden til omgivelsene (*outcome*) som følge av virksomhetens tiltak eller andre årsaker.

Sammenhengen mellom *output* og *outcome*, og resultatkjedemodellen presentert i kapittel 2.2, er vist i figur 2.3. *Outcome* deles opp i to ulike resultatmål: *direkte outcome* hvor det er en mer direkte, observerbar kausal sammenheng mellom *output* og *outcome*, og *indirekte outcome* som dekker de mer langsiktige og indirekte resultatene av virksomhetens leveranser. Det kan, som nevnt i kapittel 2.2, være utfordrende å relatere en observert endring i *outcome* til én eller flere bestemte handlinger. Likeså er det vanskelig å forutsi virkningen på *outcome* av en bestemt handling. Spesielt for *indirekte outcome* vil eksterne faktorer kunne ha stor betydning samtidig som det kan være store tidsforsinkelser mellom handlinger og observerbare *outcome*.

⁹ *The Organization for Economic Cooperation and Development – Development Assistance Committee.*

¹⁰ Noen alternative definisjoner av *outcome*: Merriam Webster: "Outcome is something that follows as a result or consequence". OECD DAC: "The likely or achieved short-term or medium-term effects of an intervention's output". Schreyer: "Outcome is a state that is valued by the customers".

Jo mer indirekte og langsiktige målene er, jo større blir normalt innslaget av kvalitative vurderinger av virksomhetens bidrag til resultatoppnåelsen.



Figur 2.3 Input, output og outcome (basert på [10]).

Sammenhengene i modellen vist i figur 2.3 kan formuleres matematisk [10;15;16] hvor *direkte outcome* G , er en funksjon av leveranser (output) y og en vektor med eksogene faktorer, \mathbf{u}_G . En *output* y vil være en funksjon av ulike input, x_1, x_2, \dots, x_n :

$$y = y(x_1, x_2, \dots, x_n) \quad (2.1)$$

I løpet av en tid T er det N ulike leveranser $\mathbf{y} = (y_1, y_2, \dots, y_N)$. I samme periode vil G også påvirkes av de eksogene faktorer/variabler, \mathbf{u}_G . Uttrykket for *direkte outcome* ved $t = T$ blir da:

$$G(t = T) = G(\mathbf{y}, \mathbf{u}_G) \quad (2.2)$$

I modellen fremstilt i figur 2.3 er *indirekte outcome*, I , en funksjon av én eller flere *direkte outcome*, \mathbf{G} og en vektor med eksogene faktorer, \mathbf{u}_I :

$$I(t = T^*) = I(\mathbf{G}, \mathbf{u}_I) \quad (2.3)$$

I er knyttet til mer langsiktige, overordnede mål og kan også involvere bruk av mer enn rene militære virkemidler. Tidsperioden T^* vil normalt være lengre enn T , fordi resultatene ofte først viser seg over lengre tid. Betydningen av de eksogene faktorene kan være vanskeligere å estimere noe som introduserer en betydelig usikkerhet med hensyn til å isolere bidragene fra den militære virksomheten.

I COPD vil *input* tilsvare de ressursene man benytter i forbindelse gjennomføring av operasjoner, det vil si innsatsfaktorene i resultatkjeden. Output er et sett av handlinger som utføres av militære styrker i den hensikt å skape resultater for å oppnå delmål (DC-er) og mål (*objectives*). I planprosessen beskrevet i COPD, utledes et hierarki av mål og delmål som skal nås på veien frem mot ønsket slutttilstand. *Outcome* vil her kunne benyttes til å si noe om i hvilken grad man når fastsatte mål.

I noen situasjoner vil man med relativt stor sikkerhet kunne etablere en kausal sammenheng mellom egne handlinger (*output*) og grad av måloppnåelse (*outcome*), slik som beskrevet i kapittel 2.2. Det antas at disse situasjonene er mer fremtredende under krise og krig enn under fredsdrift. I førstnevnte tilfelle har man en uønsket situasjon som man ønsker å bringe til en ønsket situasjon/slutttilstand (se kapittel 2.3). Det sistnevnte tilfelle dreier seg om daglige operasjoner hvor det overordnede målet er å opprettholde normaltstanden og håndtere mindre avvik fra denne.

En viktig forskjell mellom fremgangsmåten beskrevet i COPD og metoden beskrevet i dette kapitlet, er at begrepene *output* og *outcome* skiller klarere mellom resultater knyttet til egen ytelse og resultater relatert til hvordan man evner å påvirke omgivelsene (planlagte/ønskede resultater).

2.5 Måling av ytelse og effektivitet

Ytelse måles ved MOP-er (*Measures of Performance*) og effektivitet med MOE-er (*Measures of Effectiveness*). Den prinsipielle forskjellen er at MOP-ene skal si oss noe om vi gjør tingene riktig, mens MOE-ene skal si oss noen om vi gjør de riktige tingene i forhold til overordnede målsettinger, krav og føringer. MOP-ene er med andre ord knyttet til *output*, mens MOE-ene er knyttet til *outcome*.

En MOP sier noe om hvor gode vi er til å produsere *output* (leveranser). I en militær, operativ sammenheng er som regel ytelsen knyttet til hvor godt man utfører planlagte aktiviteter og handlinger. COPD definerer MOP som ”a metric used to determine the accomplishment of actions” (se kapittel 2.3) og som benyttes til *measurement of activity*.¹¹ En mer generell definisjon av MOP er:

- MOP måler ytelse relatert til *output*.

Vi vil benytte denne i det videre, fordi den er mer generell. COPD sin definisjon kan ses på som en presisering av denne. Ytelsen kan videre si noe om virksomhetens produktivitet ved å se på forholdet mellom ytelse og bruken av innsatsfaktorer. En MOP vil være knyttet til en observerbar størrelse.

En MOE skal fortelle oss noe om virksomhetens leveranser bidrar til at de overordnede målene for virksomheten nås, det vil si *outcome*.¹² Gjør vi de riktige tingene i forhold til overordnede mål

¹¹ Det er vanlig å skille mellom målinger av aktivitet og målinger av *output*. I henhold til definisjonene så er *output* knyttet til resultatene/leveransene som produseres ved virksomhetens aktiviteter.

¹² Noen alternative definisjoner av MOE: “MOE is a metric used to measure a current system state” (COPD/NATO operations assessment handbook [1;8]). “MOE is a criterion used to assess changes in system behavior, capability, or operational environment that is tied to measuring the attainment of an end state, achievement of an objective, or creation of an effect” (Commander’s handbook for assessment planning and execution, version 1.0, Joint staff, J-7 Suffolk, Virginia). “MOE assist in making the right choice by indicating ‘how well’ a solution meets the stakeholders need,.....” [17].

og føringer? Er vi på rett spor, og er progresjonen god nok i forhold til en ønsket fremtidig tilstand?

Hva er så en hensiktsmessig definisjon av MOE tilpasset vårt formål? Vi ønsker å benytte MOE-ene som måleparametere for *outcome*. *Outcome* har vi definert som en tilstand som verdsettes av interessenter. Det er derfor naturlig å knytte MOE-ene til målinger av tilstander relatert til omgivelsene som påvirkes av virksomhetens aktiviteter og tiltak. Vi velger av denne grunn å benytte COPD sin definisjon av MOE med en presisering om at tilstandsendingene gjelder omgivelsene som man ønsker å påvirke ved virksomhetens tiltak:

- MOE er en parameter som måler tilstanden til omgivelsene som påvirkes av virksomhetens aktiviteter og tiltak.

Et gitt *outcome* vil i denne sammenheng være en tilstand målt/observert ved én eller flere MOE-er, og en effekt vil være en endring i *outcome* fra to suksessive målinger.

En MOE kan måles direkte, det vil si MOE-en er i seg selv en måleparameter, eller den kan måles indirekte ved hjelp av én eller flere parametere. I det sistnevnte tilfelle må man kunne beskrive sammenhengen mellom MOE (avhengig variabel) og de observerbare parametere. Det finnes flere aktuelle metoder som kan benyttes til dette. Noen eksempler er *Structural Equation Modelling* (SEM) [18] eller Bayesianske nettverk (BN) [19].

I definisjonen understrekes det at en effekt er en endring i tilstand som kan relateres til virksomhetens leveranser. Et *outcome* eller en effekt er normalt ikke et resultat av en enkelt aktivitet/leveranse alene, men resultatet av flere leveranser (ofte spredt ut i tid) og andre, eksogene faktorer. En utfordring, som nevnt tidligere, er derfor å sannsynliggjøre viktige årsak-virkningssammenhenger mellom aktiviteter og *output* og observerte *outcomes* og effekter.

I eksemplet med myndighetsutøvelse (se kapittel 2.2) hvor et *outcome* er at lover og regler for fiskerivirksomhet etterleves, er antall observerte lovbrudd i et bestemt havområde en aktuell MOE. Denne MOE-en sier i utgangspunktet lite om det reelle antallet lovbrudd, men er betinget av ytelsesparametere (MOP-er) som tilstedeværelse i området, sensordekning (evne til overvåkning) og antall inspeksjoner av fiskefartøy. Ved å gjennomføre inspeksjoner i området over tid, og se dette i sammenheng med egen tilstedeværelse og aktivitetsnivå for fiske (som er sesongbetont), kan man danne seg et bilde av graden av etterlevelse av lover og regler.

2.6 Modellering av verdi og nytte

Operativ virksomhet dreier seg om å omsette ressurser (innsatsfaktorer) til operativ verdi/nytte. Verdien/nyttens av den operative virksomheten under fred, krise og krig gjenspeiler hva som er hensikten med operasjonene, og hvordan beslutningstakere og interessenter verdsetter ulike resultater. En viktig del av metoden er derfor å modellere beslutningstakers preferanser med

hensyn til ulike resultater (*output* og *outcome*).¹³ Det er viktig at vurderingene av operativ verdi ikke reduseres til en vurdering av ytelse og produktivitet alene, men at man knytter preferansene til *outcome* som per definisjon er en tilstand som vedsettes av beslutningstaker. I praksis er det ofte hensiktsmessig å knytte preferanser til MOE-ene for å kunne knytte verdier til observerbare størrelser.

Vi vil i det videre benytte begrepene verdi og verdifunksjon, fordi disse er mer dekkende for det vi mener med verdi/nytte i denne sammenheng [7;20]. Prosessen med å avdekke og modellere hvordan beslutningstakerne verdsetter forskjellige *outcomes* refereres til som preferansemodellering. Verdivurderingene brukes blant annet til å si noe om når noe er godt nok (når har vi oppnådd et tilfredsstillende resultat?).¹⁴

Verdifunksjonen, $V(\textit{outcome})$, uttrykker beslutningstakers preferanser for ulike resultater. For å kunne beregne verdien av forskjellige *outcomes* må man definere en preferansefunksjon som avbilder oppnådde resultater på en verdiskala. Formen på verdifunksjonen avgjør verdien av oppnådde resultater [7;23]. I sin enkleste form er den lineær (lineær sammenheng mellom resultater og verdi). I praksis vil ofte funksjonen være konkav hvor stigningen er størst for lave og/eller middels resultater og avtagende for høye. *Outcome* kan i prinsippet angis på ulike skalaer (ordinal, intervall), men må avbildes som en verdi på en forholdstallskala med et veldefinert nullpunkt [24].

Verdifunksjonene for *direkte* og *indirekte outcome* betegnes henholdsvis $V_G(G(y, U_G))$ og $V_I(I(G, U_I))$. Et *outcome* kan måles ved en eller flere MOE-er, slik at verdifunksjonene kan uttrykke verdien av en enkelt MOE, eller den aggregerte verdien av flere MOE-er.

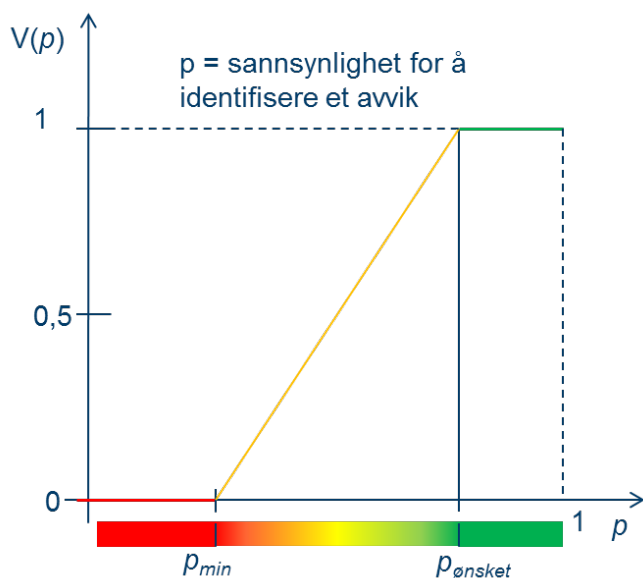
Det å tolke resultater inn i en kontekst vil involvere subjektive vurderinger. Det er derfor viktig at preferansemodelleringen i så stor grad som mulig gjøres uavhengig av observasjoner og målinger for å hindre at funksjon og nivåer påvirkes av at man kjenner resultatene. Et annet moment er at preferansefunksjoner kan endre seg over tid, for eksempel ved at beslutningstaker endrer preferanser som følge av ny informasjon, at beslutningstaker endrer verdier eller at beslutningstaker som person byttes ut.

I en operativ sammenheng er det sjelden behov for en detaljert beskrivelse av preferansefunksjonen. Ofte vil en forenklet representasjon være mer hensiktsmessig. En slik funksjon kan være binær, det vil si at den kun har ett nivå, hvor resultater over dette referansenivået er akseptable, mens resultater under er uakseptable. Ved å innføre ett ekstra nivå, for eksempel i form av et ønsket nivå og et minimumsnivå, vil man oppnå en mer nøyaktig beskrivelse av preferansene. Her vil typisk resultater som ligger på og over ønsket nivå være meget bra, resultater som ligger mellom nivåene være mer eller mindre akseptable (tiltak bør vurderes), mens resultater under minimumsnivå vil være uakseptable (tiltak må iverksettes). Det er også

¹³ I dette kapitlet refereres det til verdi/nytte-modellering relatert til *outcome*. Ofte kan det også være hensiktsmessig å knytte verdi/nytte til *output* (MOP).

¹⁴ Se [16;21;22] for eksempler på anvendelse av preferansemodellering.

mulig å tenke seg at preferansefunksjonen er lineær mellom minimum og ønsket nivå. Denne tonivåmodellen er enkel å forholde seg til for beslutningstakere ved at man bare trenger å bestemme to nivåer. Figur 2.4 viser et eksempel på en verdifunksjon hvor *outcome* er sannsynligheten for å identifisere et avvik i forbindelse med overvåkning av et havområde, mens $V(p)$ uttrykker verdien av forskjellig grad av resultatoppnåelse.



Figur 2.4 Eksempel på en verdifunksjon med to nivåer.

Sammenlikner vi disse verddivurderingene med begrepene fra COPD, vil det å fastsette mål, suksesskriteria og ønsket slutttilstand innebære en form for verddivurdering. Begrepet *Acceptable Condition* (AC), innebærer et nivå for *outcome* hvor en ønskelig/planlagt situasjon er oppnådd. Andre relevante begreper fra COPD er *Threshold of Success* (TOS) og *Threshold of Failure* (TOF), som betyr henholdsvis et vippepunkt / kritisk punkt fra hvor en positiv trend ikke vil stoppe og mest sannsynlig er irreversibel og et kritisk punkt hvor en uopprettelig (negativ) situasjon oppstår. Et annet begrep som kan benyttes for å knytte preferanser til effekter (tilstandforandringer), er *Rate of Change* (ROC) som er definert som forandringen i en måleparameter over et gitt tidsrom i løpet av operasjonen. Preferanser kan her knyttes til ulike forandringsrater.

Diskusjonen av forskjellige begreper, metoder og modeller i dette kapitlet oppsummeres og anvendes i neste kapittel i utviklingen av en metode for vurdering av måloppnåelse.

3 Metode for vurdering av måloppnåelse

En metode for vurdering av måloppnåelse må kunne støtte de ulike trinnene i evalueringsprosessen som vist i figur 2.1, og være anvendbar for daglige operasjoner så vel som for operasjoner under krise og krig. Evalueringsprosessen beskrevet i COPD er først og fremst utviklet for å støtte evalueringer av militære operasjoner under krise og krig, og i mindre grad for daglige operasjoner. Vi har derfor valgt å utvikle en metode basert på resultatkjeden til DFØ [3] hvor begrepsbruken i så stor grad som mulig er konsistent med COPD, men hvor vi gjør et klarere skille mellom den operative virksomhetens leveranser (*output*) og de målene virksomheten skal oppnå. Metoden skal hjelpe til med å identifisere mål på ulike resultatnivå og gi en sporbar sammenheng mellom mål og oppgaver på høyere nivå (politisk og strategisk) og mål og resultater på lavere nivå (operasjonelt og taktisk).

I den foreslåtte modellen skilles det mellom to nivåer, *direkte* og *indirekte outcome* som beskrevet i kapittel 2.2. I dette ligger det å skille mellom resultater som er direkte observerbare, og resultater relatert til langsiktige mål som først viser seg over tid og hvor det er mange potensielle bidragsyttere til resultatene. Et viktig poeng er at metoden skal hjelpe til med å identifisere de målene og resultatene som det er naturlig for en beslutningstaker å benytte i sin beslutningsprosess. En militær plan (operasjonsdesign) er ofte meget detaljert hva gjelder mål, delmål og effekter. Det er ikke sikkert at involverte beslutningstakere har behov for eller ønske om alle detaljene, slik at metoden må kunne hjelpe til med å tilpasse vurderingene etter behov og nivå (relevans).

Metoden krever at man gjør eksplisitte vurderinger av kausale sammenhenger mellom egen innsats og hvordan denne påvirker oppnåelsen av kortsiktige og langsiktige mål.¹⁵ I vurdering av måloppnåelse er det svært viktig å gjøre denne type analyser for i det hele tatt å kunne si noe om i hvilken grad egne handlinger har bidratt til oppnådde mål. Det er ofte slik at det er enklere å samle inn kvantitative data om egne leveranser, det vil si ytelse. Kvantitative data for måloppnåelse (effektivitet) er ofte vanskeligere tilgjengelig. Her bør man vurdere å supplere de kvantitative dataene med for eksempel ekspertvurderinger og spørreundersøkelser der dette er egnet. Vurderingen av måloppnåelse vil på denne måten kunne sammenfatte objektiv informasjon med mer subjektiv informasjon i en vurderingsprosess hvor man på en strukturert måte gjør bruk av kvantitative og kvalitative metoder (multi-metode perspektiv).

Metoden må legge til rette for å kunne samle inn data om normaltilstanden som kan utgjøre en baseline (referanse) for senere vurderinger rundt hva som er avvik fra normalsituasjonen og når man går over i en krise. En fordel med dette er at man har et sett av mål og måleparametere, og et datagrunnlag som benyttes i det daglige, og som kan tilpasses til en krisesituasjon.

Som nevnt tidligere er en evaluering vellykket i den grad resultatene er riktige og relevante, slik at beslutningstakere og interessenter benytter resultatene i styringen av virksomheten og til

¹⁵ COPD advarer mot å anta kausalitet mellom handlinger og resultater, det vil si å anta at å gjennomføre planlagte handlinger vil føre til ønskede resultater.

erfaringslæring. En forutsetning for dette er at de har tillit til resultatene. Under listes det opp noen tiltak som bidrar til å skape tillit og troverdighet for resultater, metoder og modeller:

- Sørge for relevans ved at man svarer på de ”riktige” problemene. Det vil si at man måler resultater som er relevante for de beslutningene som skal tas på ulike nivå.
- Fremstille resultatene på en enkel, lettfattelig og transparent måte tilpasset beslutningstakers behov. Men samtidig kunne vise til at resultatene er bygget på vitenskapelige baserte, valide metoder og modeller.
- Vurderinger av usikkerhetsbidrag, og hvilken betydning disse har for resultatene (se vedlegg A, A.1).
- Verifisere og validere metoder, modeller og inputdata (se vedlegg A, A.2).

Alle disse punktene bør ivaretas i en evalueringsprosess og fremheves i dokumentasjon og presentasjoner av resultatene.

3.1 Begreper og definisjoner

De viktigste begrepene og definisjonene fra kapittel 2, er gjengitt i det følgende:

- *Input*: utgjør de finansielle, materielle og menneskelige ressurser som en organisasjon må fremskaffe for å utføre sine prosesser og aktiviteter.
- *Output*: er de tjenestene, produktene eller handlingene som produseres av virksomheten.
- *Outcome*: er en tilstand (relatert til omgivelsene og virksomhetens formål) som verdsettes av beslutningstakere og andre interessenter.
- *Effekt*: er en forandring i tilstanden til omgivelsene (*outcome*) som følge av virksomhetens tiltak eller av andre årsaker (basert på COPD).
- *MOP* er en parameter som måler ytelse (relatert til *output*).
- *MOE* er en parameter som måler tilstanden til omgivelsene som påvirkes av virksomhetens aktiviteter og tiltak.
- *Måleparameter* er en kvantitativ eller kvalitativ (avhengig eller uavhengig) variabel.

I tabell 3.1 sammenstilles begreper fra punktlisten over med begreper som benyttes i COPD.

		Resultatkjede	Målhierarki (COPD-definisjon)	
Resultater/mål	Indirekte outcome	Tilstander (relatert til langsiktige og overordnede mål) som verdsettes av beslutningstakere og andre interessenter	<i>End state</i>	<i>The NAC state of conditions that defines an acceptable concluding situation for NATO's involvement</i>
			<i>Objectives</i>	<i>A clearly defined and attainable goal to be achieved</i>
		Tilstander (relatert til mer		<i>A combination of circumstances, effects, or a</i>

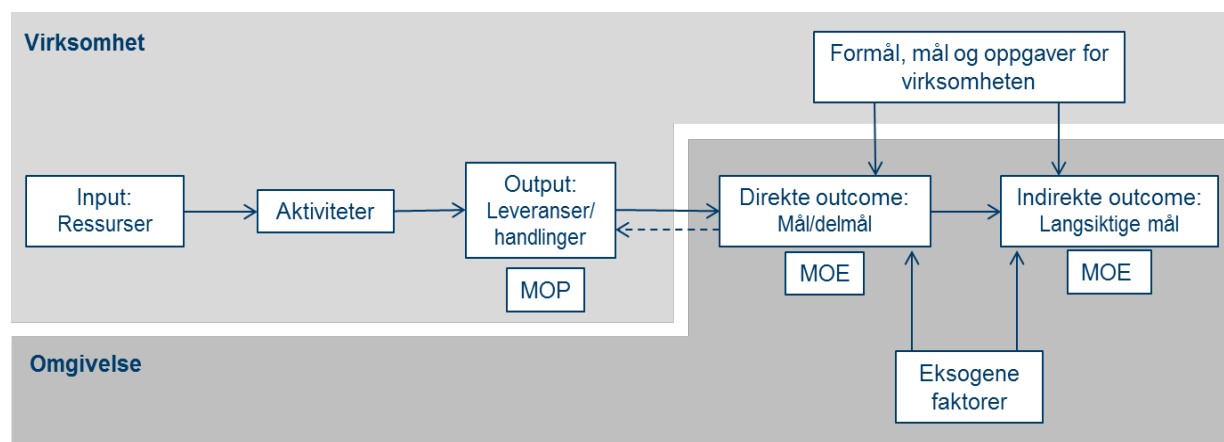
	Direkte outcome	kortsiktige mål og delmål) som verdsettes av beslutningstakere og andre interessenter	<i>Decisive condition (DC)</i>	<i>specific key event, critical factor, or function that when achieved allows commanders to gain a marked advantage over an opponent or contribute materially to achieving an operational objective</i>
	-	-	<i>Effect</i>	<i>A change in the state of a system (or system element), that results from one or more actions, or other causes</i>
	Output	Er de tjenestene, produktene eller handlingene som produseres av virksomheten	<i>Action</i>	<i>The process of doing something to achieve an aim</i>
	Aktiviteter	Virksomheten gjennomfører aktiviteter for å omsette input/ innsatsfaktorer til <i>output</i>		
Måleparametere	MOE	En parameter som måler tilstanden til omgivelsene eller til et system som påvirkes av virksomhetens aktiviteter og tiltak	<i>MOE</i>	<i>A metric used to measure a current system state</i>
	MOP	En parameter som måler ytelse (relatert til <i>output</i>)	<i>MOP</i>	<i>A metric used to determine the accomplishment of actions</i>

Tabell 3.1 Sammenstilling av begreper fra resultatkjedemodell og COPD.

Begrepet *effect* brukes i planprosessen i praksis som et nivå av delmål under DC-er, mens *actions* er aktivitetene og handlingene man utfører for og nå mål og delmål. Begrepet *effekt* er ikke en del av resultatkjeden, men er definert som en forandring i tilstand (*outcome*).

3.2 Anbefalt metode

Figur 3.1 viser metoden for vurdering av måloppnåelse.



Figur 3.1 Metode for å fastsette resultatmål på ulike nivå.

Metoden tar utgangspunktet i formålet til virksomheten og de målene som utledes for å realisere formålet. Målene bearbeides videre til resultatmål som så organiseres i henhold til resultatkjeden i direkte og indirekte outcome. Output er resultater knyttet til egen ytelse, det vil si planlagte aktiviteter og handlinger som utføres for å nå fastsatte mål, og som er kontrollerbare av virksomheten. Den stiplede pilen i figur 3.1 fra direkte outcome til output indikerer at man kan stille krav til output (ytelse og produktivitet) basert på analyse av outcome og overordnede mål [10].

Normalt vil virksomheten operere i en omgivelse hvor det er flere eksterne forhold som påvirker resultatoppnåelsen. I metoden er dette reflektert gjennom et sett med eksogene faktorer som ikke er kontrollerbare av virksomheten. Disse faktorene kan potensielt ha stor betydning for resultatoppnåelsen, slik at det er viktig å gjøre en vurdering av bidraget fra disse.

Fra figur 3.1 ser vi at output måles ved MOP-er, mens direkte og indirekte outcome måles ved MOE-er.

I det følgende beskrives hvordan metoden presentert i figur 3.1 kan støtte de ulike trinnene i prosessmodellen for vurdering av måloppnåelse gjengitt i figur 2.1.

3.2.1 Utlede oppgaver og mål

Første trinn av evalueringprosessen er å utlede og få oversikt over målene man ønsker å evaluere opp mot. Som oftest vil denne prosessen ta utgangspunkt i overordnede mål og føringer for virksomheten, og bryte disse ned til oppgaver og delmål (målhierarki) som er hensiktsmessige for de vurderingene som skal gjøres. En viktig forutsetning for dette arbeidet er at man har en klar forståelse for hvilke overordnede mål som man ønsker å nå, og hvilket nivå (strategisk, operasjonelt, taktisk) og prosesser vurderingene skal støtte. For å sikre relevante resultatmål og måleparametere er det viktig at beslutningstakere (brukere) involveres i prosessen. Involvering

bidrar til forankring og til å styrke troverdigheten til modell og resultater. Denne prosessen kan støttes ved bruk av mer formelle problemstrukturerende metoder (se for eksempel [25;26]).

I evalueringen må man kunne bestemme verdien/nytten av forskjellige resultater. En utfordring er at målene ofte er vagt formulert, slik at det er nødvendig å presisere disse gjennom modellering av beslutningstakers (og eventuelt andre interessenters) preferanser. Det anbefales å benytte en relativt enkel modell hvor en modellerer preferansene i to nivåer, et nivå som representerer en ønsket tilstand (COPD: *Acceptable Condition* og *Threshold of Success*) og et nivå hvor verdier under dette er uakseptable (COPD: *Threshold of Failure*). I tillegg til å sette verdier for *outcome*, kan man også modellere preferanser på effekter og på ytelse (se kapittel 2.6). Preferanser knyttet til ytelse bør begrunnes i analyse av hvilke krav man må stille til ytelsen for å kunne nå overordnede mål.

3.2.2 Utlede måleparametere og etablere en datainnsamlingsplan

Neste trinn i prosessen er å utlede måleparametere for målene. *Outcome* er i henhold til definisjonen, en tilstand som verdsettes av beslutningstaker og måles ved hjelp av MOE-er. Et *outcome* kan måles ved én eller flere MOE-er, hvor MOE-en kan være en avhengig eller uavhengig parameter. I de tilfellene hvor MOE-en er en avhengig parameter, vil verdien på denne baseres på én eller flere uavhengige målevariabler. Det er da nødvendig med en modell som beskriver sammenhengene mellom de ulike målevariablene. De uavhengige målevariablene benevnes enkelte ganger som indikatorer. Noen eksempler på aktuelle modeller er: lineær vektet sum, Bayesianske nettverk (BN) [19] og *Structural Equation Modell* (SEM) [27]. Den førstnevnte modellen er en første-ordens tilnærming, som benyttes mye. De andre modellene er statistiske og involverer betingede sannsynligheter, korrelasjon og regresjon. Som nevnt i kapittel 2.2, kan det være utfordrende å modellere kausale sammenhenger mellom måleparametere, MOE-er, direkte *outcome* og *indirekte outcome*. Spesielt gjelder dette for indirekte *outcome*, hvor det ofte er mange eksogene forhold som påvirker resultatet, og hvor virkningen først viser seg over tid. Her kan det være nødvendig med mer kvalitative vurderinger av kausale forhold. Se diskusjonen rundt kausalitet i kapittel 2.2, og diskusjonen om usikkerhet og validitet i vedlegg A.

Etter å ha utarbeidet et sett med relevante måleparametere må det etableres en datainnsamlingsplan som beskriver hvilke data som skal samles inn, hvordan, hvor ofte, hvilke datakilder som skal benyttes og hvem som har ansvaret. I prosessen med å utarbeide planen vil det ofte finnes måleparametere som det ikke er mulig å få tak i brukbare data for. Hvis det er tilfelle, bør man vurdere å fjerne måleparameteren, eventuelt kan man prøve å finne andre måleparametere eller indikatorer som kan benyttes til å estimere MOE og *outcome*.

3.2.3 Datainnsamling og analyse

Det siste trinnet i prosessen består i å samle inn data i henhold til datainnsamlingsplanen og beregne MOP-er og MOE-er. Dette danner videre grunnlaget for å vurdere måloppnåelse opp mot definerte *output* og *outcome*.

Analysene av innsamlede data skal bidra til å gi et godt og troverdig beslutningsgrunnlag ved å beskrive nåværende tilstand i forhold til de mål man skal nå på kort og lengre sikt. I analysen vil

man benytte historiske data (tidsserier og trender) som grunnlag for å vurdere fremdrift og resultatoppnåelse i forhold til planlagte mål og effekter. Vurderingene vil også ofte inneholde en fremskriving av situasjonen basert på kunnskap om nåsituasjonen og trender. Her kan det være nyttig å benytte scenarioer for å beskrive noen plausible fremtidige tilstander knyttet til målene (*outcome*) som man ønsker å oppnå [28]. En god hjelp i analysene er å ha god kunnskap om normalsituasjonen. Under daglig operasjoner (vedvarende operasjoner) kan denne referansen (*baseline*) bidra til at man lettere oppdager og forstår konsekvensene av ulike avvik. I en krisesituasjon vil en slik referanse være nyttig for å forstå den tilstanden (slutttilstand) som ønskes oppnådd. Resultatene må tolkes inn i den konteksten hvor operasjonene finner sted. Dette er viktig for å vurdere hvor gode resultatene er, og for å få oversikt over bidraget fra andre, eksogene faktorer. Denne delen av analysen vil ofte være mer kvalitativ.

For å støtte analysene, kan det være behov for å utvikle egne analysemodeller som sammenstiller informasjonen som samles inn om de forskjellige målene. Disse modellene kan eksempelvis benyttes til å beregne måloppnåelse og effekter, og til å gjøre “hva-hvis”-analyser.

Resultatene av analysen skal bidra til å gi beslutningstakere god situasjonsforståelse, det vil si at de skal oppfatte og forstå nåværende situasjon, og kunne predikere fremtidig tilstand (se [29]). For å oppnå dette bør beslutningsgrunnlaget inneholde vurderinger av usikkerhet i resultatene og kunne vise til at anvendte metoder, modeller og inputdata er tilstrekkelig kvalitetssikret (verifisert og validert). Dette bidrar til å skape tillit til og øke troverdigheten til beslutningsgrunnlaget. Mer om dette i Vedlegg A.

4 Eksempel på anvendelse av metoden

I dette kapitlet gis et eksempel på anvendelse av metoden for vurdering av måloppnåelse som er beskrevet i kapittel 0. Vi har valgt å se nærmere på en case innen maritim overvåkning under daglige operasjoner (vedvarende operasjoner).¹⁶ Maritim overvåkning er en viktig del av den daglige operative virksomheten, og bidrar til å løse Forsvarets oppgaver innen suverenitetshevdelse, myndighetsutøvelse, å sikre et nasjonalt beslutningsgrunnlag og å bidra til ivaretagelse av samfunnsikkerhet [31;32].¹⁷ I tillegg til dette har Norge forpliktelser med hensyn til trygghet og sikkerhet til sjøs med et stort SAR-ansvarsområde (SAR – *Search and Rescue*). I denne casen benyttes ingen reelle data for å holde innholdet ugradert.

I henhold til [32] har overvåkningen av sivil skipstrafikk, herunder også fiske, blitt langt bedre de senere årene etter innføring av systemer som AIS (*Automatic Identification System*), SafeSeaNet, satellittsporing ved VMS (*Vessel Monitoring System*), Paris MoU og egenrapportering ved kontrollpunkter og ved inn- og utpassering av indre farvann. Disse avtalene og systemene er i stor

¹⁶ Det er gjennomført en egen maritim overvåkningsstudie ved FFI [30]. Metoden benyttet i dette eksemplet er en forenklet utgave av metoden benyttet i studien. Hensikten her er å vise hvordan man ved hjelp av en relativt enkel modell kan knytte output (*MOP*-er) til outcome (*MOE*-er).

¹⁷ For flere av disse oppgavene er det ikke Forsvaret som har hovedansvaret, men bidrar i henhold til avtaler og anmodninger om bistand.

grad basert på at fartøyene følger gjeldende forskrifter for hvilke systemer de skal ha om bord og bruken av disse.

For å holde oversikt over fartøyer som er unntatt disse avtalene (militære og andre statlige fartøyer) og fartøyer som forsøker å unngå å bli oppdaget (ved ikke å bruke eller manipulere med systemer som AIS og VMS), har man behov for aktiv og passiv overvåkning av havområdene. Overvåkning ved bruk av marinefartøy, fly, landbaserte installasjoner og satellittbilder kan fylle gap i dekningen til sivile systemer. Den kan brukes til å verifisere informasjon fra sivile systemer, holde oversikt over militær trafikk og til å oppdage avvik av fartøy som ikke fanges opp av de sivile systemene. Et viktig mål er derfor å ha evne til å oppdage, identifisere og følge avvik og andre interessante fartøy i våre havområder. Dette er utfordrende da det daglig er i overkant av 4000 fartøyer innom FOH sitt ansvarsområde.¹⁸

De sivile systemene sammen med overvåkning og etterretning bidrar med informasjon som samles og bearbeides til et gjenkjent maritimt situasjonsbilde (*Recognized Maritime Picture – RMP*). I denne casen vil vi se på mål og resultater for den militære maritime overvåkingen alene, og ikke vurdere effekten av de sivile systemene og etterretning. Vi ønsker svar på følgende spørsmål:

- Hvor god er den maritime overvåkingen under daglige operasjoner?
- Er dette godt nok i forhold til overordnede mål og ambisjoner?

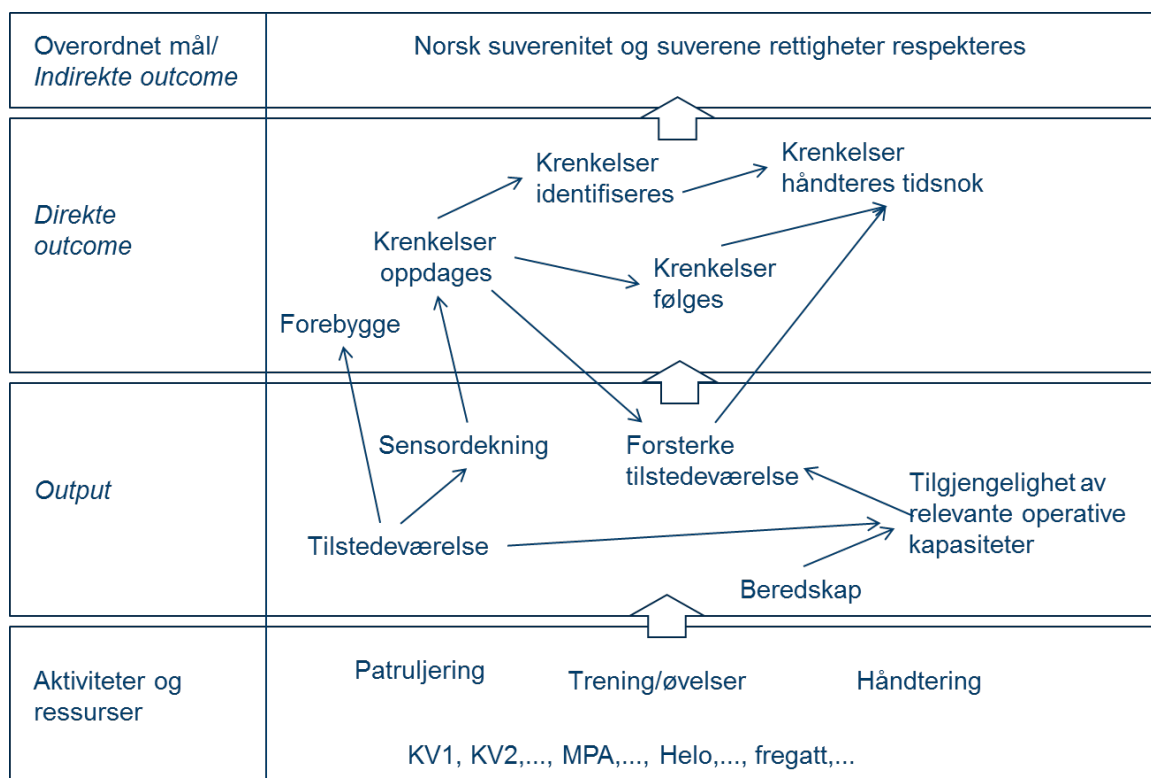
Kapittelet følger prosessmodellen illustrert i figur 2.1.

4.1 Oppgaver og mål

Resultatmålene knyttes til formålet med den maritime overvåkingen. Informasjonen som samles inn, benyttes sammen med informasjon fra sivile systemer og etterretning til å bygge og vedlikeholde et RMP for å støtte situasjonsbevissthet, og å gi en realistisk mulighet til å oppdage og identifisere avvik.

Målene organiseres i henhold til resultatkjedemodellen beskrevet i kapittel 3.2 og skal uttrykke ønskede resultater (*output* og *outcome*) fra den operative virksomheten. Figur 4.1 viser et eksempel på resultatmål for maritim overvåking til støtte for suverenitetshevdelse til havs. Det overordnede målet (*indirekte outcome*) er at norsk suverenitet og suverene rettigheter respekteres.

¹⁸ Informasjon fra overvåkingscenteret ved FOH.



Figur 4.1 Resultatmål for suverenitetshevdelse.

Direkte outcome for den maritime overvåkingen er forbundet med evnen til å oppdage, identifisere og følge et avvik innenfor gitte geografiske områder. Dette kan uttrykkes som sannsynligheter estimert på bakgrunn av observasjoner/målinger. Sannsynlighetene relaterer måleparameterne til det skisserte, litt upresise ambisjonsnivået for maritim overvåking som er å ha en realistisk mulighet til å oppdage et avvik. Vi kan også benytte verdifunksjonsmetoden beskrevet i kapittel 2.6 for å knytte preferanser til ulike resultater, for eksempel definere et ønsket nivå og et minimumsnivå for deteksjonssannsynligheten i et område.

Aktuelle innsatsfaktorer for maritim overvåking er kystvaktfartøy, maritime patruljefly (MPA), helikoptre, satellittovervåking, kystradarer etc. Disse gjennomfører aktiviteter som patruljering, øvelser og håndtering av hendelser, og gir som *output* (leveranser) antall seilingsdøgn, utseilt distanse per døgn, antall flytimer og tilstedeværelse i ulike områder. Det er den samlede innsatsen til disse strukturelementene som danner grunnlaget for vurderingen av Forsvarets bidrag til å håndheve norsk suverenitet og suverene rettigheter.

Når et avvik oppdages vil det, avhengig av avvikets karakter og egen tilstedeværelse i området, kunne være behov for å drive situasjonsbasert overvåking, det vil si å øke tilstedeværelsen av overvåkingsressurser i et område for å klassifisere og følge avviket.

4.2 Måleparametere og datainnsamlingsplan

I tabell 4.1 og tabell 4.2 vises eksempler på aktuelle *outcome* og *output* for maritim overvåking sammen med utledede MOE-er og MOP-er. Hvis en MOE kan observeres direkte vil måle-

parameteren være MOE-en, mens i de tilfellene hvor MOE-en ikke kan observeres direkte vil MOE-en være en avhengig parameter som må beregnes/vurderes på bakgrunn av én eller flere uavhengige måleparametere. I dette eksemplet vil noen MOE-er kunne observeres direkte, mens andre beregnes på bakgrunn av målinger/observasjoner.

Utledet mål	Direkte outcome	MOE	MOP og parametere
Avvik oppdages, identifiseres og følges.	Avvik oppdages.	Antall oppdagede avvik.	Antall registrerte avvik.
		Sannsynlighet for å oppdage avvik.	Tilstedeværelse. Dekningsområde.
	Avvik identifiseres.	Antall oppdagede avvik som identifiseres.	Antall avvik som oppdages og som klassifiseres entydig.
		Sannsynlighet for å identifisere et avvik.	Sannsynlighet for å identifisere et avvik gitt at det oppdages.
	Avvik følges (evne til å følge avvik).	Antall oppdagede avvik som følges.	Antall avvik som man evner å følge. Antall man ønsker å følge som man ikke kapasitet til. Trackoppdateringsrate.
		Sannsynlighet for å kunne følge avvik.	Trackoppdateringsrate.
Et relevant beslutningsgrunnlag som støttes av et godt og tidsriktig situasjonsbilde.	Et situasjonsbilde som støtter opp under god situasjonsbevissthet.	Korrekthet, kompletthet og tidsriktighet av informasjonen i bildet. Korrekthet og kompletthet av bildet.	Alder på track (oppdateringsrate). Nøyaktigheten til kinematiske variable. Korrekthet med hensyn til klassifikasjon. Antall falske track i bildet.

Tabell 4.1 Eksempler på outcome og MOE-er for maritim overvåkning.

Output	MOP	Parametere
Tilstedeværelse	Antall døgn/timer med plattformer tilstede i patruljeområde.	Sammenhengende tid i området. Gjenvisittid. Antall tokt med MPA. Utseilt distanse i område.
Sensordekning	Sensordekningsgrad i patruljeområde midlet over en gitt periode.	Tekniske parametere for plattform og sensorer. Tilstedeværelse i periode. Oppetid for stasjonære sensorer i periode. Antall satellittfoto i periode.

Tabell 4.2 Eksempler på output og MOP-er for maritim overvåkning.

For sikre at de riktige dataene samles inn til rett tid og på riktig måte er det viktig å utarbeide en datainnsamlingsplan. Planen må som et minimum fortelle hvilke data som skal samles inn, hvor ofte, hvem som har ansvaret og eventuelt hvordan dette skal gjøres. Eksempelvis vil KV rapportere inn antall seilingsdøgn med KV-skip i ulike områder i løpet av en uke. På tilsvarende måte vil man kunne samle inn data fra andre aktuelle sensorplattformer, for eksempel flytokt med MPA og antall satellittbilder.

4.3 Datainnsamling og analyse

Innledningsvis i dette kapitlet ble det stilt to spørsmål som vi ønsker svar på gjennom vurdering og analyse av innsamlede data: (i) Hvor god er overvåkingen av havområdene, og (ii) er dette godt nok i forhold til overordnede mål og krav? Formålet med analysen er å gi fornuftige svar på disse spørsmålene ved å vurdere grad av måloppnåelse opp mot etablerte resultatmål (*outcome*).

Hvor gode er vi til å oppdage, identifisere og følge tilfeldige avvik i et gitt havområde? Sannsynligheten for å oppdage et avvik i et gitt område avhenger av tilstedeværelsen av overvåkingsressurser. Dette kan være statiske sensorer som kystradarer, eller bevegelige sensorplattformer som kystvaktskip og MPA, hvor tilgjengelighet og prioriteringer av plattformer til ulike oppdag, vil gi seg utslag i varierende sensordekning.

For å kunne ta gode og rettidige beslutninger har beslutningstaker behov for oppdatert informasjon om situasjonen. En viktig del av beslutningsgrunnlaget er klassifisering og identifisering av kontakter. Klassifiseringen er nivådelte fra ukjent kontakt til at kontakten er entydig bestemt (identifisert). En måte å klassifisere kontakter på er beskrevet i [33].

Her benyttes følgende nivåer (eksempler er vist i parentes):

- ukjent
- plattform (om dette er et overflatefartøy eller fly)
- rolle (om dette er et fiskefartøy, handelsfartøy eller marinefartøy)
- type (om dette er en fregatt, kystkorvett eller destroyer)
- klasse (om er en fregatt av Nansen-klasse eller Udaloy-klasse)
- entydig (om dette er KNM Roald Amundsen)

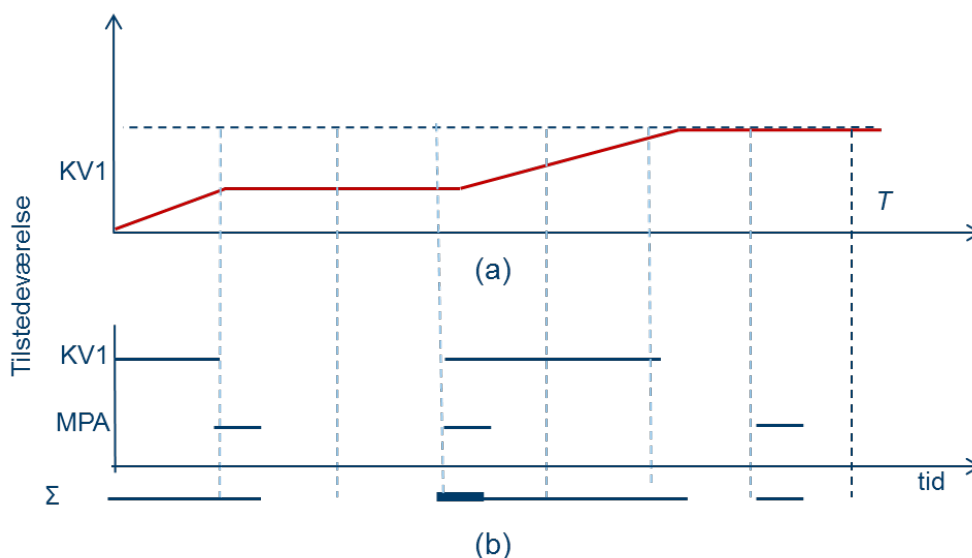
En sikker identifisering krever at kontakten er entydig bestemt. I en studie gjennomført av kanadiske DRDC (*Defence Research and Development Canada*) [34] kreves det at fartøyet må observeres visuelt for å kunne identifiseres.

I tillegg til klassifiseringen av kontakten kommer kinematisk informasjon som kurs og fart. I en operativ beslutningsprosess er det ønskelig å samle inn så mye informasjon som mulig om et avvik for å kunne vurdere om dette er en potensiell trussel. I denne sammenheng er det viktig å ha evne til å oppdage og identifisere en potensiell trussel, og å ha kapasitet til å kunne følge (*track*) denne over tid. Her er oppdateringsraten av informasjonen om kontakten en sentral parameter (*track update rate*). Hvilke deler av informasjonen i kontakten som oppdateres og oppdateringsraten, vil avhenge av hvilke sensorer som følger kontakten. Hvis oppdateringsintervallene blir lange, vil kvaliteten på informasjonen degraderes ved at usikkerheten i posisjon, kurs og fart blir større. Man kan også risikere å miste kontakten.

4.3.1 Tilstedeværelse

Forsvaret er tilstede med overvåkingsressurser i norsk territorium og i havområder hvor vi hevder suverenitet. I den operative beslutningsprosessen allokeres tildelte ressurser til forskjellige oppdrag i ulike områder for blant annet å vise tilstedeværelse (være synlig) og drive overvåkning.

Gjennom vurdering av måloppnåelse vil man kunne si noe om tilstedeværelsen av overvåkingsressurser i disse områdene over en hensiktsmessig tidsperiode, T . I område A vil det i perioden T kunne være forskjellige sensorplattformer tilstede, for eksempel KV-fartøy og maritime patruljefly (MPA). Figur 4.2 (a) viser den akkumulerte tilstedeværelsen for én type plattform i løpet av perioden T (for eksempel over én uke). Både den totale tilstedeværelsen i området i løpet av T og sammenhengende tilstedeværelse er viktige parametere.



Figur 4.2 (a) Akkumulert tilstedeværelse av én plattform (KV1), og (b) oversikt over total tilstedeværelse i område A i løpet av perioden T .

Den totale tilstedeværelsen i området kan enkelt beregnes ved å inkludere tilsvarende tilstedeværelsesprofiler for andre kapasiteter i en total sum over T som vist i figur 4.2 (b). Her må man ta hensyn til om det er overlapp mellom tidsperiodene de ulike plattformene er tilstede eller ikke. Sammenhengende tilstedeværelse har betydning for evnen til å detektere et avvik. Det er eksempelvis forskjell på om en plattform er tilstede i et område fire ganger seks timer i løpet av T (for eksempel én uke), eller om den er tilstede én gang sammenhengende i 24 timer.

4.3.2 Deteksjon og identifisering av tilfeldig avvik

På bakgrunn av en oversikt over tilstedeværelse av sensorplattformer i et område A , er det mulig å si noe om evnen til å oppdage og identifisere et tilfeldig avvik i A . Sensorplattformene har ulike sensorer som bidrar til overvåkningen. Disse sensorene kan grovt deles inn i kategoriene, radar, EO-IR (elektrooptisk og infrarød) og ESM. Ved å ha kjennskap til noen egenskaper ved sensorene, er det mulig å estimere sensordekningsgrad per tidsenhet.

Dekningsgraden danner videre grunnlaget for å estimere sannsynligheten for å oppdage et tilfeldig avvik i område A . Modellen som benyttes er relativt grovkornet og inneholder en rekke forutsetninger som påvirker usikkerheten og gyldighetsområdet til resultatene, se vedlegg B. En sensor på en plattform har en effektiv søkebredde w , hvor alle fartøyer som er innenfor denne oppdages, det vil si at $P_d = 1$ innenfor w . Det antas at sensoren beveger seg med gjennomsnittshastigheten til sensorplattformen v i område A . Det antas videre at sensorplattformen driver tilfeldig søk, det vil si at den patruljerer og ikke søker etter et bestemt mål. Gitt disse forutsetningene kan likning B.1 i vedlegg B benyttes for å gi et grovt estimat av sannsynligheten for å oppdage et tilfeldig avvik. I tabell 5.3 vises den totale deteksjonssannsynligheten for radarene på sensorplattformene KV1 og MPA som er tilstede i område A over en syvdagersperiode. Det antas at området som skal overvåkes har en lengde på 275 km og en bredde på 370 km. Arealet som skal overvåkes blir da ca. 102,000 km². Her er det antatt at plattformene patruljerer uavhengig av

hverandre. Dette er en rimelig antakelse i dette tilfellet hvor et KV-fartøy patruljerer i området og hvor en MPA flyr over som en del av et planlagt tokt.

Plattform	Deteksjonssannsynlighet (P_d) for radar per døgn						
	1	2	3	4	5	6	7
KV1	0,17	0	0	0,17	0,3	0	0
MPA	0	0,03	0	0,03	0	0	0,03
Sum P_d	0,17	0,03	0	0,19	0,3	0	0,03

Tabell 4.3 Estimat av deteksjonssannsynlighet for et tilfeldig avvik i område A gitt tilstedeværelsen som vist i figur 4.2.

Resultatene i dette eksemplet indikerer at man midlet over én uke oppnår en deteksjonssannsynlighet på litt over 0,1, det vil si ca. 10 % sjansje for å oppdage et tilfeldig avvik i A.

Et viktig spørsmål for den videre vurderingsprosessen er hvorvidt det er godt nok med en 10 % deteksjonssannsynlighet i patruljeområdet? Det skisserte ambisjonsnivået for maritim overvåkning i Forsvarets operative planverk (FOPV) er å ha en realistisk mulighet for å oppdage avvik. For å kunne vurdere om det oppnådde resultatet er godt nok, må man konkretisere ambisjonsnivået. Hva er så et ønsket nivå for P_d ? Her er det mulig å gjøre verdi/nyttebetraktninger, som beskrevet i kapittel 2.6 ved for eksempel å definere et ønsket nivå og et minimumsnivå. I fastsettelsen av disse preferansenivåene må man vurdere hva som er hensikten med den maritime overvåkingen og hva som er en akseptabel risiko for at en trussel ikke blir oppdaget.

Innledningsvis i dette kapitlet poengteres det at når et avvik oppdages, er det viktig å få klassifisert og identifisert dette for å få et så godt beslutningsgrunnlag som mulig. I [34] benyttes nettopp evnen til å identifisere et avvik som det overordnede effektivitetsmålet for den maritime overvåkingen. Ulike sensorer har forskjellig kapasitet med hensyn til å klassifisere kontakter. I denne analysen forutsettes det at oppdagede avvik må klassifiseres entydig for at de skal være identifisert. For å oppnå dette klassifikasjonsnivået antas det at kontakten må klassifiseres visuelt. Det vil si at alle oppdagede avvik klassifiseres visuelt, slik at sannsynligheten for å identifisere et avvik settes lik sannsynligheten for å oppdage avviket gitt at man har kapasitet for visuell klassifikasjon.

Sannsynligheten for å oppdage et tilfeldig avvik brukes som et effektivitetsmål. En annen mulighet er å se på evnen til å oppdage og identifisere et avvik når et fartøys kurs avviker fra godkjente trafikkruter langs kysten ved at det for eksempel beveger fra sjøterritoriet inn mot indre farvann eller mot bestemte offshore installasjoner [34].¹⁹ Her er det avgjørende at det er overvåkingsressurser til stede i området når avviket skjer, noe som betyr at tilstedeværelse og gjenvisitid for sensorplattformer er viktige parametere. I dette tilfellet må man også gjøre antakelser

¹⁹ Et mulig scenario kan være et fiskefartøy som slår av eller manipulerer med AIS og VMS.

rundt en eventuell trusselaktørs evne til å utnytte perioder hvor tilstedeværelsen i området er liten. Dette vil igjen avhenge av hvorvidt patruljeringen skjer ved faste tidspunkt eller om patruljeringen skjer mer tilfeldig.

4.3.3 Følge avvik

Når et avvik er oppdaget har man behov for å følge (*tracke*) dette. Evnen til å følge en kontakt vil avhenge av tilgjengeligheten på mobile sensorplattformer, gitt at området ikke er dekket av stasjonære sensorer. Et vanlig mål på evnen til å følge mål er *track* oppdateringsrate, det vil si hvor ofte informasjonen om *track*-et oppdateres. En måte å fremstille dette på er å beregne den kumulative "ventetiden" på en oppdatering. DRDC-rapporten [34] skiller mellom oppdateringsrate for sensorer (rådata) og situasjonsbildet.

4.3.4 Beregning av resultatmål

I eksemplet som gis i dette kapitlet, beregnes sannsynligheten for å oppdage et tilfeldig avvik innfor et gitt patruljeområde over en tidsperiode. I tillegg samles det inn data for antall og type avvik som er observert og håndtert over samme periode. I analysen kan man sammenholde tilstedeværelse og oppdagelsessannsynlighet med observerte avvik. Når man har gjort dette over tid, og fått en tidsserie med data, kan man benytte ulike statistiske metoder og modeller for å trekke slutninger med hensyn til resultatoppnåelse. I tillegg kan man gjøre enkle trendanalyser.

4.3.5 Usikkerhetsvurderinger

Usikkerhetsvurderinger er en viktig del av et beslutningsgrunnlag. I vedlegg A presenteres en metode som gir en strukturert gjennomgang av ulike usikkerhetsfaktorer. En viktig dimensjon i usikkerhetsvurderingene er lokalisering av usikkerhet. Modellen trekker frem fem lokaliteter: kontekstuell usikkerhet, modellusikkerhet, inputusikkerhet, parameterusikkerhet og usikkerhet relatert til modellresultatene.

I dette eksemplet er den kontekstuelle usikkerheten blant annet knyttet til hva man legger i begrepet suverenitetshevdelse, hva man ønsker å oppnå (mål, ambisjoner og resultater), hvilke andre, eksterne faktorer/forhold som påvirker måloppnåelsen og avgrensing av konteksten (hva modellen skal omfatte). En god innledende problemstrukturering kan bidra til å redusere den kontekstuelle usikkerheten.

Modellusikkerheten er relatert til mangelfull kunnskap og forståelse av systemet som skal modelleres og dets oppførsel. En modell kan defineres som en forenklet representasjon av et system, som har de samme relevante egenskapene som systemet. Modellusikkerheten knyttes til hvilke elementer av det virkelige systemet som tas med i modellen og relasjonene mellom dem. I vårt eksempel vil det eksempelvis være modellusikkerhet knyttet til de kausale sammenhengene mellom egne leveranser (*output*) og resultatoppnåelse (*outcome*), og til beregningene av deteksjonssannsynlighet.

Inputusikkerheten knyttes til data som beskriver egenskaper og oppførsel til systemet som modelleres, og eksterne faktorer som påvirker systemet og dets ytelse. Noen eksempler er tilstedeværelse av KV-fartøy i et område, patruljemønster og vær- og vindforhold.

Parameterusikkerheten er relatert til de konstante parameterverdiene i modellen. Parameterne kan gis en eksakt verdi (middelverdi) eller man kan representere verdien med en sannsynlighetsfordeling.

Usikkerheten i modellresultater er den totale usikkerheten i resultatene fra modellen, det vil si den aggregerte usikkerheten fra de andre bidragene.

5 Oppsummering og avslutning

Rapporten beskriver en metode for vurdering av måloppnåelse som er basert på en resultatkjede med fem nivåer: *input*, *aktiviteter*, *output*, *direkte outcome* og *indirekte outcome*. De tre første nivåene er knyttet til virksomheten, mens de to siste beskriver hvordan virksomheten påvirker sine omgivelser. *Outcome* er i henhold til definisjonen benyttet i denne rapporten, en tilstand som verdsettes av interessenter.

Vurdering av måloppnåelse er en sentral del av en beslutningsprosess som støtter opp under beslutningstakers situasjonsbevissthet. Vurderingene baseres på innsamlede data supplert med ekspertvurderinger og andre informasjonskilder ved behov.

Vurderinger av måloppnåelse sier noe om progresjon i forhold til planlagte mål og i hvilken grad egne (kontrollerbare) aktiviteter og handlinger bidrar til at målene nås. Det er viktig med et klart skille mellom leveranser/handlinger som utføres, og som er kontrollerbare av virksomheten, og resultatene man oppnår. Resultater er normalt knyttet til ønskede tilstandsendringer i omgivelsene og relateres i metoden til to nivåer, direkte observerbare resultater (*direkte outcome*) og lang-siktige, indirekte resultater (*indirekte outcome*). På det sist nevnte nivået er ofte egne leveranser en av flere bidragsyttere til resultatoppnåelsen. Det er derfor viktig å synliggjøre kausale forhold for at vurderingene skal kunne brukes til å si noe om egne bidrag. Vurderinger opp mot lang-siktige mål vil ofte være mer kvalitative enn vurderingene rundt mer direkte og observerbare resultater.

Det er viktig at beslutningstaker benytter denne type vurderinger i sin beslutningsprosess. En forutsetning for dette er at resultatene er relevante og troverdige. Det førstnevnte krever at man har forståelse for hvilke behov og ønsker beslutningstaker har, mens det andre fordrer at man har en strukturert og kvalitetssikret prosess for å samle inn og analysere data. I dette ligger det å benytte valide data, metoder og modeller, og at man har et bevisst forhold til usikkerheten i datagrunnlag og analysemodeller. I analysefasen vil ulike kombinasjoner av metoder og modeller – både kvantitative og kvalitative – kunne benyttes for å støtte vurderingene av måloppnåelse. Vi mener at den anbefalte metoden bidrar til å underbygge vurderingenes relevans og troverdighet.

Vi ønsker en metode for vurdering av måloppnåelse som er egnet for vurderinger under normaldrift så vel som under krise og krig. Vi mener at den anbefalte metoden har flere fordeler i forhold til COPD. Den er bedre egnet for vurderinger av måloppnåelse under normaldrift. I COPD fokuserer planutvikling og *assessment* på å gå fra en uønsket tilstand til en ønsket slutttilstand. I forbindelse med daglige operasjoner er ikke dette hensiktsmessig. Her er det viktig å støtte opp under læring og god situasjonsbevissthet for blant annet å identifisere mulige forbedringsområder og operative behov. I tillegg ser vi for oss synergier mellom arbeidet som gjøres for å etablere mål og måleparametere for den operative virksomheten, og rapporteringen som gjøres i forbindelse med Forsvarets virksomhetsstyring (DIVØ [35]). Her vil det være betydelige overlapp mellom informasjonen som samles inn for å brukes i *assessment* -prosessen ved FOH, og den informasjonen som rapporteres inn i forbindelse med virksomhetsstyringen i Forsvaret (resultat og kontrollrapport – RKR).

Innsamling av relevante data under daglige operasjoner bidrar til å bygge opp kunnskap om normalsituasjonen, som videre blant annet vil være til hjelp i vurderingene av risiko ved observerte avvik fra det normale. Forslag til mål og måleparametere for daglige operasjoner er dokumentert i [2].

Litteratur

- [1] NATO, *Allied command operations comprehensive operations planning directive - COPD*, Interim V2.0 ed Supreme HQ Allied Powers Europe, 2013.
- [2] Malerud S., "Vurdering av måloppnåelse i nasjonale operasjoner - forslag til mål og måleparametere for daglige operasjoner", FFI-notat 2014/01560, Begrenset, 2014.
- [3] Senter for statlig økonomistyring (SSØ), "Resultatmåling. Mål- og resultatstyring i staten", desember 2010.
- [4] Hennem A.C., Eggereide B., Rutledal F., and Marthinussen E., "Utvikling av rammeverk for fremdriftsvurderinger knyttet til norsk støtte i Faryab-provinsen i Afghanistan", FFI-rapport 2010/02519, 2010.
- [5] Forsvarsdepartementet, "Et forsvar for vår tid", Prop.73 S (2011–2012), mars 2012.
- [6] Atkinson, "Atkinson review: Final report - Measurement of government output and productivity for the national accounts", Palgrave Macmillan, 2005.
- [7] Goodwin P. and Wright G., *Decision analysis for management judgement*, John Wiley & Sons, third ed., Chichester, England, 2004.
- [8] NATO, "NATO Operations Assessment Handbook (NOAH)", version 2.0, desember 2012.
- [9] Forsvarets stabsskole, "Forsvarets fellesoperative doktrine", Forsvarsstaben, juni 2007.
- [10] Schreyer P., "Output and outcome - measuring the production of non-market services", *Paper prepared for the 30th General Conference of The International Association for Research in Income and Wealth*, juni 2008.
- [11] Mandel U., Dierx A., and Ilzkovitz F., "The effectiveness and efficiency of public spending", *European Economy. Economic Papers* 301, februar 2008.
- [12] OECD Development Assistance Committee, "Guidance on evaluating conflict prevention and peacebuilding activities", 2008.
- [13] Øhrn M., "Produktivitetstallinger i Forsvaret – erfaringer fra pilotprosjekt i 2012", FFI-rapport 2013/00064, 2013.
- [14] Hanson T., "Produktivitetstallinger i Forsvaret – metode og anvendelsesområder", FFI-rapport 2010/01495, 2010.
- [15] Førstund F.R., "Measuring efficiency in the public sector", 2012.
- [16] Hanson T., "Produktivitetstallinger i Forsvaret – eksperimentell modell for prioriteringseffektivitet", FFI-rapport 2012/02265, 2012.
- [17] Sproles N., "Measures of effectiveness - The standards for success", University of South Australia, januar 1999.
- [18] Anderson R.D. and Vastag G., "Causal modeling alternatives in operations research: Overview and application", *European Journal of Operational Research*, vol. 156 (2004) 2004.
- [19] Jensen F.V., *An introduction to Bayesian Networks*, UCL Press Limited, 1996.

- [20] Belton V. and Stewart T.J., *Multi criteria decision analysis - An integrated approach* Kluwer Academic Publishers Group, 2003.
- [21] Ljøgdott H., Kråkenes T. og Malerud S., "Metoderammeverk for analyse av kriseberedskap og lavintensitetskonflikter illustrert ved anvendelse på oljevernberedskap", FFI-rapport 2007/01813, 2007.
- [22] Johansen I., "Hvilke utfordringer skal Forsvaret dimensjoneres mot? Vekting av scenarioer med flermålsanalyse", FFI-rapport 2004/02636, 2004.
- [23] Keeney R.L. and Raiffa H., *Decisions with multiple objectives - Preferences and value tradeoffs*, John Wiley & Sons, 1976.
- [24] Malerud S. og Kråkenes T., "Metoder for flermålsanalyse – En oversiktsstudie fra GOAL", FFI-rapport 2005/03041, 2005.
- [25] Rosenhead J. and Mingers J., *Rational analysis for a problematic world revisited*, John Wiley & Sons Ltd, second ed., Chichester, England, 2001.
- [26] Gilljam M. og Ljøgdott H., "Problem structuring methods. A survey and a case study", FFI-rapport 2005/00852, 2005.
- [27] Gupta S. and Kim H.W., "Linking structural equation modeling to Bayesian networks: Decision support for customer retention in virtual communities", *European Journal of Operational Research*, vol. 190 (2008), 2008.
- [28] Malerud S. og Fridheim H., "Metode for utvikling av scenarioer til spill og øvelser", FFI-rapport 2013/00219, 2013.
- [29] Endsley M.R., "Toward a theory of situation awareness in dynamic systems", *Human Factors*, vol. 37(1), 1995.
- [30] Vatne D.F., Gisas H. og Krey S.M., "Funksjonell studie – maritim overvåkning", FFI-rapport 2014/00360, Begrenset, 2014.
- [31] Skogstad A.K. og Solheim S.O., "Kystvaktens oppgaver", FFI-rapport 2010/01312 (untatt offentlighet), 2010.
- [32] Skogstad A.K., "Maritim overvåkning i en bred kontekst – interesser og ansvarsområder", FFI-rapport 2011/01484 (untatt offentlighet), 2011.
- [33] Mevassvik O.M., Hansen B.J. og Sanden H., "Utvidelse av sensorsim", FFI-notat 2003/00006, 2003.
- [34] Horn S.A., Carson N.L., and Wind A.F., "A metric for maritime intelligence, surveillance and reconnaissance (ISR) - Probability of identification", DRDC CORA TM 2009-037, 2009.
- [35] Forsvarsstaben, "Direktiv for virksomhets- og økonomistyring", juni 2009.
- [36] Walker W.E. et al, "Defining uncertainty. A conceptual basis for uncertainty management in model-based decision support", *Integrated Assessment*, vol. Vol. 4, No: 1 (2003), 2003.
- [37] NATO RTO, "NATO Guide for Judgement-Based Operational Analysis in Defence Decision Making", RTO Technical Report AC/323(SAS-087)TP/345, juni 2012.

- [38] Uk Ministry of Defence, "Guidelines for the verification and validation of operational analysis modelling capabilities", D/DCS(S&A)/7/18/7, november 1997.
- [39] Hartley D.S., "Verification and validation in military simulations", *Proceedings of the 1997 Winter Simulation Conference*, 1997.
- [40] Robinson S., "Simulation model verification and validation: Increasing the users' confidence", *Proceedings of the 1997 Winter Simulation Conference*, 1997.
- [41] Gondran M. and Minoux M., *Graphs and algorithms*, John Wiley & Sons, 1984.
- [42] US Naval Academy, *Naval operations analysis*, US Naval Institute, 1968.

Forkortelser

AC	Acceptable Condition
AIS	Automatic Identification System
AJP	Allied Joint Publication
BN	Bayesianske nettverk
COPD	Comprehensive Operations Planning Directive
DC	Decisive Condition
DIVØ	Direktiv for virksomhets- og økonomistyring
EO-IR	Elektrooptisk-infrarød
EP	Eksperiment prosjekt
ESM	Electronic Support Measure
FOH	Forsvarets operative hovedkvarter
FOPV	Forsvarets operative planverk
FRAGO	Fragmentation Order
JCO	Joint Coordination Order
KV	Kystvakt
MOE	Measures of Effectiveness
MOP	Measures of Performance
MPA	Maritime Patrol Aircraft
NAC	North Atlantic Council
NATO	North Atlantic Treaty Organization
NOAH	NATO's Operations Assessment Handbook
OA	Operasjonsanalyse
OECD DAC	The Organization for Economic Cooperation and Development – Development Assistance Committee
PMESII	Political, Military, Economic, Social, Infrastructure, Information
RMP	Recognized Maritime Picture
ROC	Rate of Change
SAR	Search and Rescue
SME	Subject Matter Experts
SSØ	Senter for statlig økonomistyring
TOF	Threshold of Failure
TOS	Threshold of Success
VMS	Vessel Monitoring System

Vedlegg A Om usikkerhet og validitet

En metode for vurdering av måloppnåelse er vellykket i den grad resultatene er riktige og relevante, slik at beslutningstakere benytter disse i sine beslutningsprosesser. For å oppnå dette er det avgjørende at metoder, modeller, datagrunnlag og resultater oppleves som relevante og troverdige. Troverdighet henger sammen med validitet som underbygges av modellens relevans, kompletthet og korrekthet og god usikkerhetskåndtering.

I dette kapitlet diskuteres to sentrale forhold som sammen bidrar til å skape tillit og troverdighet: usikkerhetskåndtering og validering. Først diskuteres ulike usikkerhetsbidrag som kan påvirke resultatene av evalueringsprosessen, dernest diskuteres verifisering og validering av metoder, modeller og inputdata.

A.1 Om usikkerhet

En viktig del av en analyse er vurderinger av ulike usikkerhetsbidrag og den totale usikkerheten i resultatene. Vi benytter her en typologi over usikkerhetsbidrag utviklet av Walker et al. [36] for vurderinger av usikkerhet innen modellbasert beslutningsstøtte. Modellen er bygd opp rundt tre usikkerhetsdimensjoner:

1. *Lokalitet*: beskriver hvor usikkerheten gjør seg gjeldende. Det er definert fem generiske lokaliteter for usikkerhet: kontekst, modell, inputdata, parameterusikkerhet og modellresultater.
2. *Nivå*: beskriver graden av usikkerhet mellom ytterpunktene deterministisk kunnskap og total uvitenhet/ignoranse. Nivåene som benyttes i [36] er:
 - *Statistisk usikkerhet*, som er usikkerhetsbidrag som kan beskrives ved bruk av statistiske metoder.
 - *Scenariocusikkerhet*, som har sitt opphav i mangelfull kunnskap om omgivelsene (drivende krefter og viktige sammenhenger) som påvirker systemet. Denne usikkerheten betyr at det i prinsippet er mange scenarioer som kan være aktuelle, men at man i liten grad kan knytte sannsynligheter til disse på grunn av at mekanismene/sammenhengene som leder frem til disse er lite kjent.
 - *Anerkjent uvitenhet* betyr at det er stor usikkerhet knyttet til omgivelsene og systemets oppførsel. Mangelfull kunnskap gjør det vanskelig å lage relevante og troverdige scenarioer. Denne usikkerheten kan være reduserbar ved at man innhenter mer kunnskap og/eller forsker mer på aktuelle tema.
 - *Total uvitenhet* innebærer at man ikke engang vet at man ikke vet.
3. *Opphav/natur*: beskriver hvorvidt usikkerheten skyldes mangelfull kunnskap (epistemisk usikkerhet) eller iboende variabilitet i de fenomenene som studeres.

I det følgende vil vi anvende dimensjonene over til å diskutere usikkerhet relatert til metoden for vurdering av måloppnåelse. Først vil vi vurdere ulike lokaliteter for usikkerheten, dernest vil vi se på nivå og karakter av usikkerhet for identifiserte lokaliteter.

Usikkerhet relatert til kontekst:

Med kontekst menes omstendigheter for en hendelse eller tilstand. Konteksten virker inn på hvordan vi tolker denne hendelsen eller tilstanden.²⁰ I forbindelse med vurderinger av mål-oppnåelse er det viktig å forstå hvilken del av konteksten man ønsker å påvirke. I dette ligger det å foreta en fornuftig avgrensning av konteksten for å rette oppmerksomheten mot de delene som man ønsker å påvirke. Problemstrukturering kan hjelpe til med:

- Problemformulering (informasjon om beslutningsproblemene som skal støttes)
 - Hva er formålet med virksomheten og/eller operasjonen?
 - Hva ønskes oppnådd (hvilke mål skal nås)?
- Avklare hvem som er brukere av resultatene
 - Hva er deres verdier og preferanser?
 - Hvilke nivå er involvert (strategisk, operasjonelt eller taktisk nivå)?
- Avgrensning av kontekst
 - Hva skal med i modellen, og hva er eksterne forhold?
 - Detaljgrad
- Eksisterer det begrensinger/føringer med hensyn til kontekst (politiske, militære, økonomiske, sosiale)?

Når det gjelder kontekst er det i hovedsak snakk om scenariosikkerhet og (anerkjent) uvitenhet, og opphavet er i hovedsak relatert til mangelfull kunnskap (epistemisk usikkerhet) om kontekstuelle forhold. Denne usikkerheten kan reduseres ved å legge vekt på problemstrukturering og involvering av interessenter i en tidlig fase av modellutviklingen. Både COPD og *assessment handbook* understreker viktigheten av å involvere *assessment staff* i planutviklingen.

Modellusikkerhet:

En modell kan defineres som en forenklet representasjon av et system som innehar de samme relevante egenskapene som systemet. En fullstendig komplett modell av virkeligheten hvor et hvert fysisk element av virkeligheten er representert i modellen, er en urealistisk ambisjon. Det vi ønsker er en tilstrekkelig komplett og korrekt modell som er relevant for de forholdene vi vil studere.

Walker et al. skiller mellom to typer modellusikkerhet:

- Modellstrukturusikkerhet
- Modellteknisk usikkerhet

I første omgang er det usikkerhet knyttet til modellstrukturen som er interessant for oss. Denne har blant annet sitt opphav i mangelfull kunnskap og forståelse av systemet (operasjonen) som skal modelleres og dets oppførsel. Mer presist knyttes usikkerheten til hvilke elementer av det reelle systemet som skal inkluderes i modellen og deres innbyrdes relasjoner uttrykt ved parametere, variabler, funksjoner, input og output. Denne usikkerheten gir seg uttrykk i at det, i

²⁰ Fra: www.no.wikipedia.org

prinsippet, eksisterer mange mulige, plausible representasjoner av systemet. Utfordringen består i å finne den representasjonen som i størst grad stemmer overens med modellens hensikt, men med en fare for at ingen av disse er en tilstrekkelig god nok representasjon av virkeligheten.

I modellene for vurdering av måloppnåelse er det spesielt utfordrende å modellere kausale sammenhenger mellom virksomhetens aktiviteter, *output* og *outcome*. I resultatkjedemodellen skiller vi mellom *direkte outcome* og mer langsiktig samfunnsmessig *outcome*. Kravet som stilles til *direkte outcome* er at disse skal være kortsiktige, observerbare virkninger hvor det kan sannsynliggjøres en klar sammenheng mellom årsak (aktiviteter og output) og observert virkning (*outcome*). For *indirekte outcome* vil det ofte være vanskeligere å relatere en observert virkning til en bestemt årsak på grunn av at det kan være lange tidskonstanter og mange eksterne forhold som påvirker resultatet.

Modellene kan deles inn i tre kategorier.

- Hierarkiske modeller som beskriver sammenhengen mellom oppgaver, mål, delmål og resultatmål for den operative virksomheten.
- Kausale modeller som beskriver sammenhenger mellom ulike nivåer i resultatkjeden og diverse modeller for å beregne *output* og *outcome* basert på innsamlet empiri. Hensikten med disse modellene er å relatere måleparametere til mål på ulike nivå. Dette kan være enkle lineære modeller basert på funksjoner og likninger eller statistiske modeller.²¹
- Ulike modeller for å støtte analysen av måloppnåelse.

Den strukturelle usikkerheten i modellkategoriene, nevnt over, gjør seg gjeldene på alle nivåer fra statistisk usikkerhet til total uvitenhet om forhold som påvirker langsiktige samfunnsmessige *outcome*. Det meste av usikkerheten har sitt opphav i mangelfull kunnskap om eksterne forhold som påvirker tilstanden til omgivelsene.

Usikkerheten rundt de langsiktige samfunnsmessige resultatene vil ofte være så stor at man vanskelig kan påstå at observerte endringer skyldes spesielle tiltak. Det er derfor utfordrende, basert på målinger og empiri, å trekke bastante slutninger om resultatoppnåelse. Det er derfor naturlig å vurdere samfunnsmessig *outcome* mer kvalitativt, men med god støtte i innsamlede data og observasjoner.

Inputusikkerhet:

Inputusikkerhet er relatert til data som beskriver egenskaper og oppførsel til systemet som modelleres og eksterne drivende faktorer som påvirker systemet og dets ytelse. Eksterne drivende krefter som ikke er under kontroll av aktuelle beslutningstakere, er av spesiell interesse. Disse faktorene har stor påvirkning på systemets oppførsel, og ulike kombinasjoner av verdier på disse faktorene vil i prinsippet gi opphav til mange ulike scenarioer [28]. Ofte vil det være vanskelig å vite hvor stor betydning de enkelte faktorene har på resultatene, slik at denne usikkerheten kan karakteriseres som scenariouisikkerhet. Eksterne og interne forhold som man ikke kjenner til eller

²¹ Se for eksempel ****.

har kunnskap om skyldes uvitenhet om, disse forholdene. En del av denne uvitenheten kan reduseres ved ytterligere kunnskapsoppbygging.

Usikkerheten knyttet til systemdataene skyldes mangelfull kunnskap om egenskapene ved systemet som man modellerer og variabilitet i de fenomenene man studerer. Walker et al. [36] skiller mellom noen ulike kilder for usikkerhet knyttet til variabilitet. Disse er: naturlige tilfeldigheter, sosial variabilitet og variabilitet knyttet til menneskelig oppførsel. Alle disse kildene er relevante for våre modeller.

Parameterusikkerhet:

Parametere er konstanter i modellen. Walker skiller mellom:

- eksakte parametere
- bestemte (konstante) parametere
- a priori parametere (parameterverdier er estimert basert på erfaringer)
- kalibrerte parametere (parameterverdier er kalibrert basert på historiske data)

Parameterne kan gis en eksakt verdi, eller de gis verdier representert ved en sannsynlighetsfordeling.

Modellresultater:

Usikkerheten i modellresultatene er den totale usikkerheten i resultatene av analysene og vurderingene som gjøres. Dette er den aggregerte usikkerheten fra alle de enkelte bidragene. Noen av disse vil ha større betydning enn andre, og usikkerheten vil øke jo lengre ut i resultatkjeden man beveger seg. Usikkerheten uttrykker forholdet mellom modellresultater og den sanne verdien, som man i de fleste tilfeller ikke kjenner. Det er da viktig å gjennomføre en god valideringsprosess og å beskrive ulike usikkerhetsbidrag for å skape troverdighet til de beregnede resultatene.

A.2 Validering

Validering skal bidra til å øke troverdigheten til modell og resultater. Validering går ut på å underbygge at modellen er en tilstrekkelig god (nøyaktig) representasjon av den virkeligheten og de problemene man ønsker å modellere, det vil si å vise at man bygger den “riktige modellen” [37–40]. Verifisering av modellen er å sjekke at denne (implementasjonen) er i overensstemmelse med spesifikasjonene, med andre ord å svare på om vi bygger modellen riktig.

Hva bidrar til validitet?

- Relevans av modell og modellresultater
- Reproduserbarhet av modellresultater
- Sporbarhet i utviklingen av modellen
- Transparens i modellutvikling og resultater

- Kompletthet – modellen inkluderer alle relevante aspekter av systemene og problemstillingene som man vil modellere
- Korrekthet – modellen er en tilstrekkelig god representasjon av den virkeligheten man modellerer
- God usikkerhetshåndtering med hensyn til modell, input og modellresultater
- Bruk av anerkjente standarder, metoder og modeller

Hvordan validere?

- Skaffe seg en god forståelse av de problemene som skal håndteres og den virkeligheten man skal modellere.
- Klargjøre gyldighetsområdet til modellen, slik at modellen benyttes til det den er utviklet for.
- Sammenlikne modellresultater opp mot reelle observasjoner (empiri) fra forsøk, øvelser og historiske data.
- Få en gjennomgang (revisjon) av metoder, modeller og resultater av ekstern ekspertise med fokus på styrker og svakheter.
- Benytte godt validerte, anerkjente vitenskapelig metoder og modeller.
- Sikre god dokumentasjon av metode og modeller.

B.1 Representasjon av geografi og terreng

Det er mange måter å representere geografi og terreng på – fra detaljerte til overordnede beskrivelser. Beskrivelsene må tilpasses behovet. I forbindelse med vurderinger av den operative virksomheten har vi funnet det hensiktsmessig å benytte en graf med noder og kanter. Det er mange fordeler med denne fremstillingen fordi man kan benytte metoder fra grafteori [41]. En grafrepresentasjon er en aggregert modell som innehar relevante egenskaper ved geografien, og er en mellomting mellom en nøyaktig koordinatbasert representasjon og en mer kvalitativ beskrivelse av et geografisk område. Vi antar at geografien (land, sjø, luft) er delt inn et sett av hensiktsmessige områder som kan representeres ved noder i en graf, $G = [N, K]$, hvor N er settet av noder og K er settet av kanter.

En annen mye benyttet metode er å dele inn geografiske områder i ruter eller heksagoner av ulik størrelse. Disse modellene gir mulighet for en mer detaljert representasjon av geografien hvor nøyaktigheten avhenger av størrelsen på rutene.

B.2 Beregning av deteksjonssannsynlighet

Vi velger oss ut et bestemt område (node), N_i , og antar at sensorplattformene driver tilfeldig søk i dette området. Tilfeldig søk innebærer at det ikke søkes i et bestemt mønster etter et bestemt objekt (patruljering). Vi velger at N_i har en rektangulær form med lengde l_i og en bredde b_i som gir et areal $A_i = l_i * b_i$. I dette området antar vi at det er et tilfeldig avvik, det vil si at det er like stor sannsynlighet i hele A_i for at avviket kan opptre (flat fordeling over arealet). Vi er ute etter å finne sannsynligheten for at dette avviket oppdages av de sensorene som er tilstede i noden. Vi benytter likningen under (Koopmans modell) for å beregne den kumulative deteksjonssannsynligheten i området for bevegelige plattformer (se referanser [34;42]):

$$P_n(t) = 1 - e^{-\left(\frac{v \cdot w}{A_i}\right)t} \quad (\text{B.1})$$

Dette er en relativt grovkornet modell hvor v er gjennomsnittshastigheten til plattformen, w er søkebredden (se avsnitt B.3) og t er tiden fra plattformen entrer noden og begynner å søke. Alternativt kan man benytte utseilt distanse i området N_i ved sammenhengende tilstedeværelse, så vil dette erstatte $v * t$ i uttrykket over.

Vi har et område med areal A_i hvor en sensorplattform opererer. Sensorplattformen dekker et område lik $w * l$, hvor $l = L/n$ er et linjestykke av den totale patruljelengden L tilbakelagt i området i løpet av en tidsperiode T . Sannsynligheten for ikke å detektere et avvik i et dekningsområde, $l * w$ er $(1 - l * w / A_i)$. I løpet av tidsperioden T rekker plattformen å av søke en lengde $L_n = n * l = v * T$, slik at sannsynligheten for ikke å detektere et avvik i løpet av de n første segmentene blir:

$$\left(1 - \frac{l \cdot w}{A_i}\right)^n \quad (\text{B.2})$$

Sannsynligheten for å få minst én deteksjon i løpet av n segmenter blir da:

$$P_d = 1 - \left(1 - \frac{l \cdot w}{A_i}\right)^n \approx 1 - e^{-\frac{l \cdot w}{A_i}}, \quad \frac{w \cdot l}{A_i} \ll 1 \quad (\text{B.3})$$

Et grovkornet maksimumsestimert av deteksjonssannsynligheten er å se på et lineært søk uten overlapp hvor deteksjonssannsynligheten er forholdet mellom søkeareal dekket som funksjon av tiden dividert med det totale arealet av noden, A_i :²²

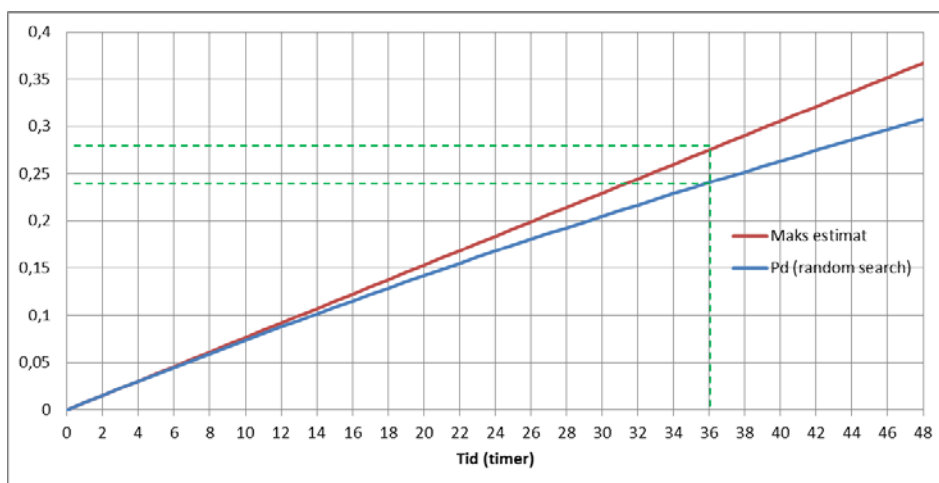
$$P_d \approx \frac{l \cdot w}{A_i} = \frac{w \cdot v}{A_i} t \quad (\text{B.4})$$

I eksemplet i kapittel 4.3 gjengis resultatene fra et estimat av sannsynligheten for å detektere et tilfeldig avvik i et område med et areal på $A_i = 102\,000 \text{ km}^2$. Tabell B.1 viser tekniske data for aktuelle sensorplattformer.

Plattform	Søkebredde (w) radar [km]	Søkebredde (w) EO-IR [km]	Gjennomsnittlig søkehastighet [km/t]
KV1	28	15	28
MPA	230	15	500

Tabell B.1 Tekniske data for ulike plattformer.

Resultatene av estimatene av P_d for tilfeldig søk (likning B.3) og maksimumsestimert (likning B.4) for plattformen KV1 er vist i figur B.1.



Figur B.1 Sannsynlighet for deteksjon av tilfeldig avvik.

²² Maksimumsestimert tilsvarende en førsteordens tilnærming til Koopmans modell (første leddet i rekkeutviklingen).

Figur B.1 viser at maksimumsestimatet vokser lineært med tiden plattformen opererer i området, mens Koopmans modell følger en negativ eksponential utvikling. De grønne stiplede linjene viser estimert deteksjonssannsynlighet etter at plattformen har operert i området i 36 timer.

Maksimumsestimatet gir $P_d \approx 0,28$, mens tilfeldigøk modellen til Koopman gir $P_d \approx 0,24$.

Tilfeldigøk-modellen anses å utgjøre et minimumsestimat fordi man ofte vil planlegge en patruljerute gjennom et område for blant annet å unngå for mye overlapp. Hvis man har indikasjoner på at det er avvik i området, vil man normalt av søke området mer systematisk. For disse tilfellene finnes det andre og bedre modeller enn tilfeldigøk modellen (se for eksempel NOA [42]).

I diskusjonen i avsnitt 4.3.2 hevdes det at det ønskede *outcome* for den maritime overvåkingen er å kunne identifisere avvik. Dette innebærer å ha evne til å fremskaffe relevant informasjon om avviket for å understøtte situasjonsbevissthet hos beslutningstakere.

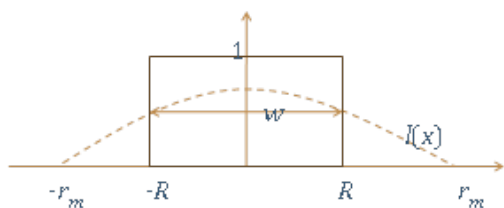
En førsteordens tilnærming til P_{id} er å anta $P(id/d) = 1$, det vil si at man antar at man evner å identifisere alle avvik som oppdages, $P_{id} = P_d$. Dette er i og for seg selv ikke en urimelig antakelse da man mest sannsynlig (under normaldrift) vil forsøke å allokere nødvendige ressurser for å klassifisere og identifisere avviket hvis det er behov for det. En mulig forbedring av denne modellen er å innføre ulike avstander for deteksjon og identifisering. Dette kan gjøres ved å introdusere ulike rekkevidder, som gir soner for deteksjon og identifisering/klassifisering. Denne metoden er blant annet beskrevet i [33].

Forenklingen, $P_{id/d} = 1$, innebærer at det ikke er noen hensikt i å skille mellom sensorklassene, fordi radaren vil normalt ha lengst rekkevidde og vil dermed dominere P_d . Det forutsettes også at man evner å følge (*tracke*) et fartøy som detekteres og vurderes som et avvik.

B.3 Beregning av søkebredde

En lateral rekkevidde (*lateral range curve*), $l(x)$, kan defineres som den kumulative sannsynligheten for å oppdage et mål når dette passerer igjennom deteksjonsområdet til sensoren med en *Closest Point of Approach* (CPA) = x . Kurven $l(x)$ vil blant annet avhenge av egenskaper ved målet og værmessige forhold som betyr at man i prinsippet må ha et stort antall kurver for å dekke alle forhold. Vi ønsker i utgangspunktet kun én $l(x)$ for hver sensor. Dette introduserer en unøyaktighet i estimatene av sensorrekkevidde og deteksjonssannsynlighet. Utfordringen er å finne en tilstrekkelig god representasjon.

Vi ønsker å uttrykke deteksjonsegenskapene ved en sensor i en fysisk meningsfull størrelse slik som for eksempel søkebredden w , se figur B.2.



Figur B.2 Beregning av søkebredde ved bruk av en “Cookie cutter”-modell.

Søkebredden kan beregnes ved å integrere $l(x)$ mellom $(-r_m, r_m)$:

$$w = \int_{-r_m}^{r_m} l(x) dx \quad (\text{B.5})$$

En mye benyttet modell for å uttrykke sensorrekkevidde er ”cookie cutter”-modellen hvor en ekvivalent $l(x)_{cc}$ defineres ved å ta arealet under $l(x)$ og omforme dette til et rektangel med høyde, $P_d = 1$ og bredde lik effektiv søkebredde $w_{cc} = 2R$. Denne modellen kan benyttes hvis vi antar at målet/avviket er uniformt distribuert (uniform sannsynlighetsfordeling) innenfor rekkevidden av sensoren, slik det er i tilfeldigsøk-modellen (*Koopmans random search*) beskrevet i B.2.

Uttrykket for $l(x)_{cc}$ blir da:

$$l(x)_{cc} = \begin{cases} 1, & x \in [-R, R] \\ 0, & \text{ellers} \end{cases} \quad (\text{B.6})$$

hvor effektiv søkebredde er:

$$w_{cc} = \int_{-R}^R 1 dx = 2R \quad (\text{B.7})$$