Nederlandse organisatie
voor toegepast
natuurwetenschappelijk
onderzoek

TNO-rapport

Fysisch en Elektronisch
Laboratorium TNO

Postbus 96864
2509 JG 's-Gravenhage
Oude Waalsdorperweg 63
's-Gravenhage

Telefax 070 - 328 09 61
Telefoon 070 - 326 42 21

rapport nr. FEL-91-A323

exemplaar nr. 8

titel

Waardering en analyse van simulatieresultaten

auteur(s):

Ir. M. van der Kaaij

Drs. J.K. Vink

datum :

december 1991

rubricering		
titel	:	ongerubriceerd
samenvatting	:	ongerubriceerd
rapporttekst	:	ongerubriceerd
bijlage A & B	:	ongerubriceerd
oplage	:	37
aantal bladzijden	:	68 (incl bijlagen, excl. RDP & distributielijst)
aantal bijlagen	:	2

2 VAN VRAAG NAAR ANTWOORD

Met behulp van gevechtssimulaties kunnen tal van vragen beantwoord worden. Voorbeelden van zulke vragen zijn:

- Wat is de invloed van een bepaalde verbetering aan een wapensysteem?
- Welke wijze van optreden is beter?
- Welke mix van wapensystemen is het beste?

In dit hoofdstuk wordt beschreven wat er gedaan moet worden om met behulp van een stochastisch simulatiemodel antwoord te krijgen op bovenstaande vragen, als al in een eerder stadium is gekozen voor een stochastisch simulatiemodel voor het beantwoorden van de vraag.
Aan het eind van het hoofdstuk is de beschreven methodiek weergegeven in twee schema's. Het eerste schema (figuur 1) geeft weer hoe tot een keuze van een statistische methode kan worden gekomen. Het tweede schema (figuur 2) geeft weer hoe onderzocht kan worden of het waarschijnlijk is dat een set waarnemingen uit een normale verdeling komt.

Om te beginnen moet de vraag duidelijk zijn. Een aantal dingen moet bij de vraagstelling bepaald worden:

- Op basis van welke maat van effectiviteit (MOE) bepaalde uitkomsten met elkaar worden vergeleken.
- Op welk moment de MOE uitgerekend moet worden.
  Dit kan het einde van de simulatie zijn of bepaald worden aan de hand van een aantal criteria.
  In een gevechtssimulatie kan bijvoorbeeld de MOE uitgerekend worden als de ene partij tot 40% gesleten is of als de andere partij tot 30% gesleten is.

Vaak kan er geen antwoord gekregen worden met behulp van de bestaande scenario's. Met een scenario wordt bedoeld de hele set van invoergegevens die nodig is om een simulatie te draaien.
Er zullen nieuwe scenario's gecreëerd moeten worden of oude aangepast. Voor het beantwoorden van de meeste vragen zijn meerdere scenario's nodig. Door het vergelijken van de uitkomsten van de scenario's kunnen antwoorden verkregen worden.

report no.	:	FEL-91-A323
title	:	Valuation and analysis of simulation results
author(s)	:	M. van der Kaaij, J.K. Vink
institute	:	TNO Physics and Electronics Laboratory
date	:	December 1991
NDRO no.	:	A89KL619
no. in pow '91	:	701.1
Research supervised by:		P.A.B. van Schagen
Research carried out by:		M. van der Kaaij, J.K. Vink

ABSTRACT (UNCLASSIFIED)

For answering questions about the deployment of weaponsystems, the stochastic combat simulation model FSM (Force Structure Model) is one of the models developed. This report describes the way to answer a question by means of the stochastic simulation model and statistical techniques.

The outcomes of combat simulations have to be valued before they can be compared to each other. This can be done with the help of one or more measures of effectiveness (MOE's). When using a stochastic model, the model has to be runned several times with the same input. In this way a set of observations is generated and something can be said about the distribution of possible outcomes.

In this report a number of statistical tests are described with which two sets of observations can be compared to each other.

Some computer programs are written for making the analysis of simulation results easier. These programs are also described.

An impulse is given to further research.

SAMENVATTING		2
ABSTRACT		3
INHOUDSOPGAVE		4
1	INLEIDING	6
2	VAN VRAAG NAAR ANTWOORD	7
3	MATEN VAN EFFECTIVITEIT	13
3.1	Overzicht	13
3.2	Nadere beschrijving van de MOE's	13
3.3	Bruikbaarheid van de MOE's	17
4	GEBRUIKTE STATISTISCHE METHODEN	19
4.1	Frequentiehistogrammen	19
4.2	Empirische verdelingsfuncties	21
4.3	Statistieken	22
4.4	Goodness of fit	25
4.4.1	Kolmogorov-Smirnov	25
4.4.2	Chi-kwadraat	27
4.5	Betrouwbaarheidsintervallen	28
4.6	Bootstrap methode	30
4.7	Datasets vergelijken	31
4.7.1	Verdelingsafhankelijke methoden	31
4.7.2	Verdelingsvrije methoden	33
5	DE ONTWIKKELDE PROGRAMMATUUR	39
5.1	Het computerprogramma WRITE_UNITS	41
5.2	Het programma CALCULATE_MOE	43
5.3	Het programma DEC_MOE_VIS	45
5.3.1	Handleiding programma DEC_MOE_VIS	45

5.3.2 Voorbeelden van het werken met DEC_MOE_VIS 46
5.3.3 De benodigde invoer 60

6 MOGELIJKHEDEN VOOR VERDER ONDERZOEK EN AANBEVELINGEN 61

LITERATUUR OVERZICHT 63

BIJLAGE A INDEX

BIJLAGE B HET GEBRUIK VAN FSM_USER_FILES EN FSM_USER_PARAMETERS

1 INLEIDING

Binnen de researchgroep wapensysteem- en beleidsstudies LAND van het Fysisch en Elektronisch Laboratorium wordt, voor het uitvoeren van studies, gebruik gemaakt van stochastische gevechtssimulatiemodellen. Het Force Structure Model (FSM) is één van die modellen.

In het kader van de opdracht FSM (A89KL619) is onderzoek gedaan naar de methodiek om met behulp van dergelijke modellen antwoorden te verkrijgen op vragen die voortvloeien uit de studies. Deze methodiek wordt in hoofdstuk 2 beschreven.

De methodiek is gebaseerd op de kwantitatieve waardering van de simulatieresultaten. Dit houdt in dat elke simulatie uitkomst in een getal wordt uitgedrukt. Zo'n getal wordt MOE (Measure Of Effectiveness) genoemd. In hoofdstuk 3 worden enkele MOE's voor gevechtssimulaties beschreven.

Met behulp van MOE's is het mogelijk om simulatieresultaten grafisch weer te geven en met elkaar te vergelijken. Hierbij wordt gebruik gemaakt van statistische methoden. Deze methoden worden in hoofdstuk 4 beschreven.

Om de analyse van simulatieresultaten te vergemakkelijken zijn drie computerprogramma's geschreven. In hoofdstuk 5 worden deze drie programma's beschreven. Een aanzet tot vervolgonderzoek wordt gedaan in hoofdstuk 6.

De hoofdstukken zijn afzonderlijk te lezen. Afhankelijk van de interesse van de lezer kan het lezen van een enkel hoofdstuk voldoende zijn.
Via de index in bijlage A kan gevonden worden waar in het rapport meer over een bepaald onderwerp staat geschreven.

Relatieve gevechtskracht blauw

De relatieve gevechtskracht blauw wordt gedefinieerd als de huidige gevechtskracht van blauw gedeeld door de initiële gevechtskracht van blauw. Deze MOE geeft alleen de eigen sterkte weer. In het vervolg wordt deze MOE aangeduid met FORCE_BLUE. FORCE_BLUE kan als volgt worden uitgedrukt:

$$\text{FORCE\_BLUE} = \frac{B\text{-}b}{B}$$

Het bereik van deze MOE is [0,1].

Relatieve verliezen rood

Relatieve verliezen rood wordt gedefinieerd als de verliezen van rood gedeeld door de initiële gevechtskracht van rood. Deze MOE geeft het verlies van de vijand weer, en kan als volgt worden uitgedrukt:

$$\text{LOSS\_RED} = \frac{r}{R}$$

Ook deze MOE ligt tussen 0 en 1.

Relatieve gevechtskracht verhouding

De relatieve gevechtskracht verhouding wordt gedefinieerd als de verhouding van de gevechtskracht van blauw tot de gevechtskracht van rood gedeeld door de initiële gevechtskracht verhouding. Deze MOE geeft de relatieve sterkte van blauw ten opzichte van rood weer. In formulevorm ziet deze MOE er als volgt uit:

$$\text{FORCE\_RATIO} = \frac{B\text{-}b/R\text{-}r}{B/R}$$

Deze MOE ligt tussen 0 en oneindig.

Met elk scenario moet een aantal runs gedraaid worden. Het aantal runs beïnvloedt de betrouwbaarheid van de in het statistisch deel gebruikte schattingen. Het draaien van runs kost echter veel (computer)tijd en veel geheugencapaciteit. Over het algemeen zullen 50 runs een goede indicatie geven over de verdeling van de uitkomsten.

Bereken nu de MOE van elke run.

Maak van de MOE's voor elk scenario een frequentiehistogram en/of een grafiek van de empirische verdelingsfunctie. Dit geeft een indicatie of er iets vreemds aan de hand is (rare uitschieters en dergelijke kunnen zo opgespoord worden).
Als er met verschillende criteria is gewerkt voor het bepalen van het moment waarop de MOE uitgerekend moest worden is het zinvol om frequentiehistogrammen te maken die bij elke run aangeven om welke reden de MOE van die run op dat moment is uitgerekend.

Nu kunnen de scenario's paarsgewijs vergeleken worden. Bij elke vergelijking hoort een nulhypothese, een alternatieve hypothese (Meestal: nulhypothese is niet waar) en een betrouwbaarheidsniveau 1 - α.
α geeft aan wat de toelaatbare kans is dat de alternatieve hypothese aanvaard wordt terwijl de nulhypothese waar is. 0.05 en 0.10 zijn gebruikelijke waarden voor α.
Er kunnen twee verschillende groepen nulhypotheses onderscheiden worden. Ten eerste de groep waarbij de hele verdeling van de uitkomsten van belang is. Dan luidt de nulhypothese bijvoorbeeld: de verdeling van de MOE's van het ene scenario is gelijk aan de verdeling van de MOE's van het andere scenario. Ten tweede de groep waarbij alleen de verwachting van belang is. Dan luidt de nulhypothese bijvoorbeeld: de verwachting van de MOE in het ene scenario is gelijk aan het verwachting van de MOE in het andere scenario.

Het vergelijken van de gehele verdelingen van de MOE's uit twee scenario's begint met het maken van een frequentiehistogram en een grafiek van de empirische verdelingsfuncties. Op het oog kan dan al iets gezegd worden over de overeenkomst tussen de verdelingen van de MOE's. Dit is echter zeer subjectief. Daarom moet er daarna de toets van Kolmogorov-Smirnov gedaan worden. In het programma DEC_MOE_VIS is een gewijzigde Kolmogorov-Smirnov toets ingebouwd. Deze toets weet raad met nulhypotheses als: de verdeling van de MOE's van het ene scenario verschilt minder dan een door de vraagsteller vastgestelde waarde van de verdeling van de MOE's uit het andere scenario.

Om alleen de verwachtingen met elkaar te vergelijken moet eerst een statistisch overzicht gemaakt worden van de MOE's uit beide scenario's. Hieruit kunnen de gemiddelde worden afgelezen. Dit zijn schatters voor de verwachting. Het verschil tussen de twee gemiddelden geeft een indicatie voor het al dan niet aanvaarden van de aanname dat er verschil is tussen beide verwachtingen.

Daarna moet bekeken worden of de waarnemingen van de MOE's afkomstig kunnen zijn uit een normale verdeling. Dit is bepalend voor het latere gebruik van toetsen die gebaseerd zijn op de normale verdeling of voor het gebruik van verdelingsvrije toetsen.

Het testen of de waarnemingen afkomstig kunnen zijn uit een normale verdeling begint aan de hand van grafische weergaven van de resultaten en aan de hand van de statistieken scheefheid (Engels: skewness) en kurtosis. De scheefheid en kurtosis zijn gelijk aan 0 voor een normale verdeling. De beide statistieken mogen niet al te veel van 0 afwijken om aan te nemen dat de waarnemingen uit een normale verdeling komen. Voor 50 waarnemingen zou als richtlijn een maximale afwijking van 1.0 genomen kunnen worden.
Als nu nog steeds getwijfeld wordt over het al dan niet afkomstig zijn van de waarnemingen uit een normale verdeling dan moet een Goodness of fit toets uitgevoerd worden. Als er veel waarnemingen zijn dan kan het beste de Chi-kwadraat toets gebruikt worden (richtlijn: meer dan 25 waarnemingen). Zijn er minder waarnemingen dan is het beter om de Kolmogorov-Smirnov toets te gebruiken. Het gebruik van deze toets is echter af te raden als de waarnemingen uit een discrete verdeling lijken te komen. In dat geval moet de aanname dat de waarnemingen uit een normale verdeling komen verworpen worden.
In figuur 2 staat de methodiek schematisch weergegeven.

Als in een eerder stadium de aannames zijn aanvaard dat de waarnemingen van MOE's uit normale verdelingen komen, kan een Student's t-toets gebruikt worden of een variant hierop, de Satterthwaite toets. Deze laatste mag gebruikt worden zonder de, op theoretische gronden gebaseerde, aanname van gelijke varianties.
Is de aanname van normale verdelingen niet gerechtvaardigd dan moet een verdelingsvrije methode gebruikt worden. Dit kan zijn de toets van Wilcoxon of de bootstrap methode.
Met behulp van de toets van Wilcoxon kan getoetst worden of de verwachtingen van elkaar verschillen en met behulp van de bootstrap methode kan een betrouwbaarheidsinterval gecreëerd

worden van het verschil tussen de verwachtingen. Dit interval geeft aan dat met de opgegeven kans de werkelijke waarde van het verschil tussen de verwachtingen in het interval ligt.

Soms kan het nodig zijn dat van één of meerdere scenario's een aantal extra runs gedraaid moeten worden om een antwoord met meer precisie te verkrijgen.

Nu moet met behulp van de verkregen statistische gegevens antwoord worden verkregen op de vraag. Het is moeilijk om in het algemeen te beschrijven hoe dit moet. In het geval van een eenduidige vraag (bijvoorbeeld: duurt het in dit ene scenario net zolang tot de vijand is gesleten tot 40% dan in dat andere scenario) dan is een antwoord vrij makkelijk te geven. Vaak is de vraagstelling complexer en roept een statistische vergelijking andere vragen op. Ook is het vaak zo dat er niet vergeleken wordt op basis van één MOE maar op basis van meerdere. In dat geval moeten technieken als multi-criteria analyse uitkomst brengen. Statistische technieken kunnen dan gebruikt worden om interval-schattingen te maken voor verwachtingen of voor verschillen tussen verwachtingen van twee scenario's.

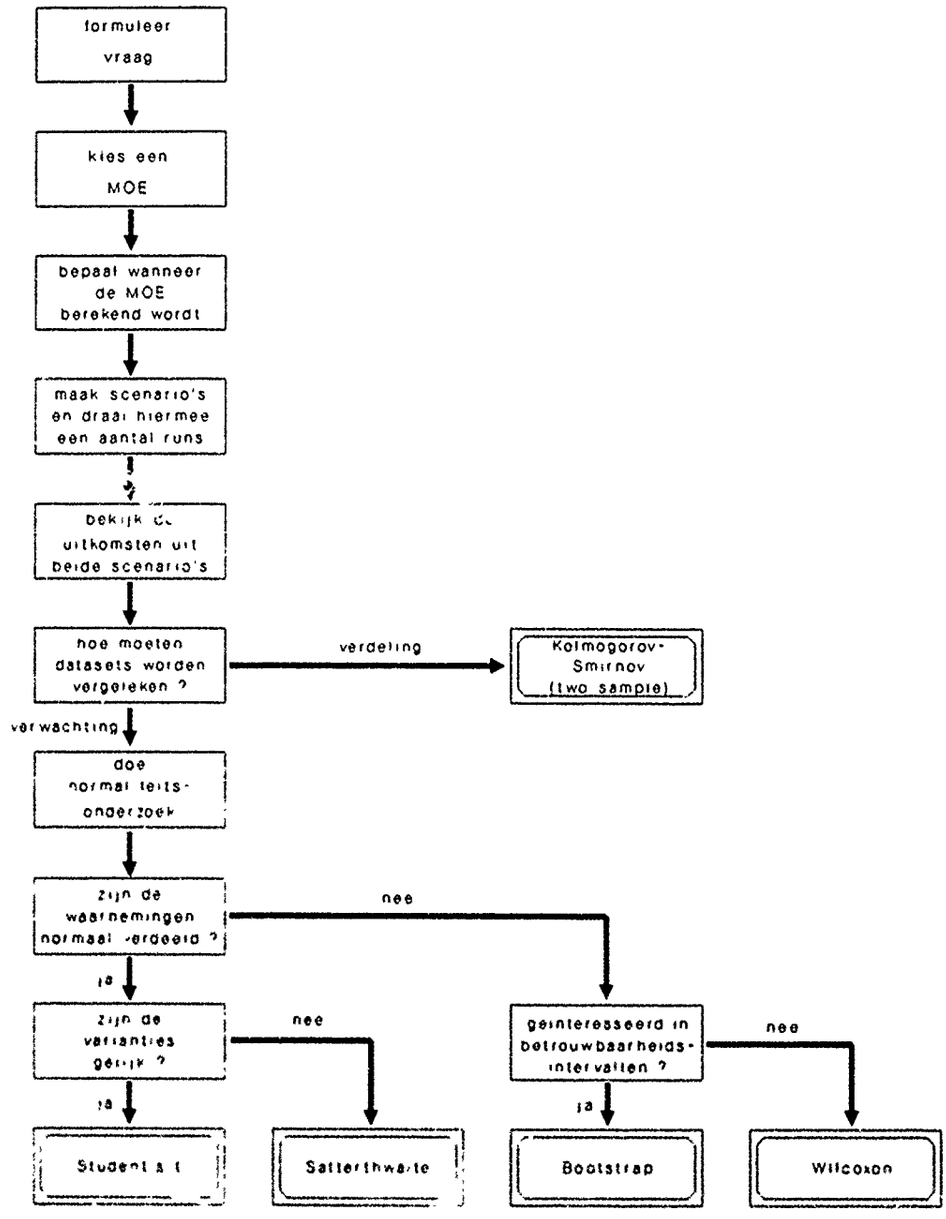

Fig. 1: Schematisch overzicht van de beschreven methodiek

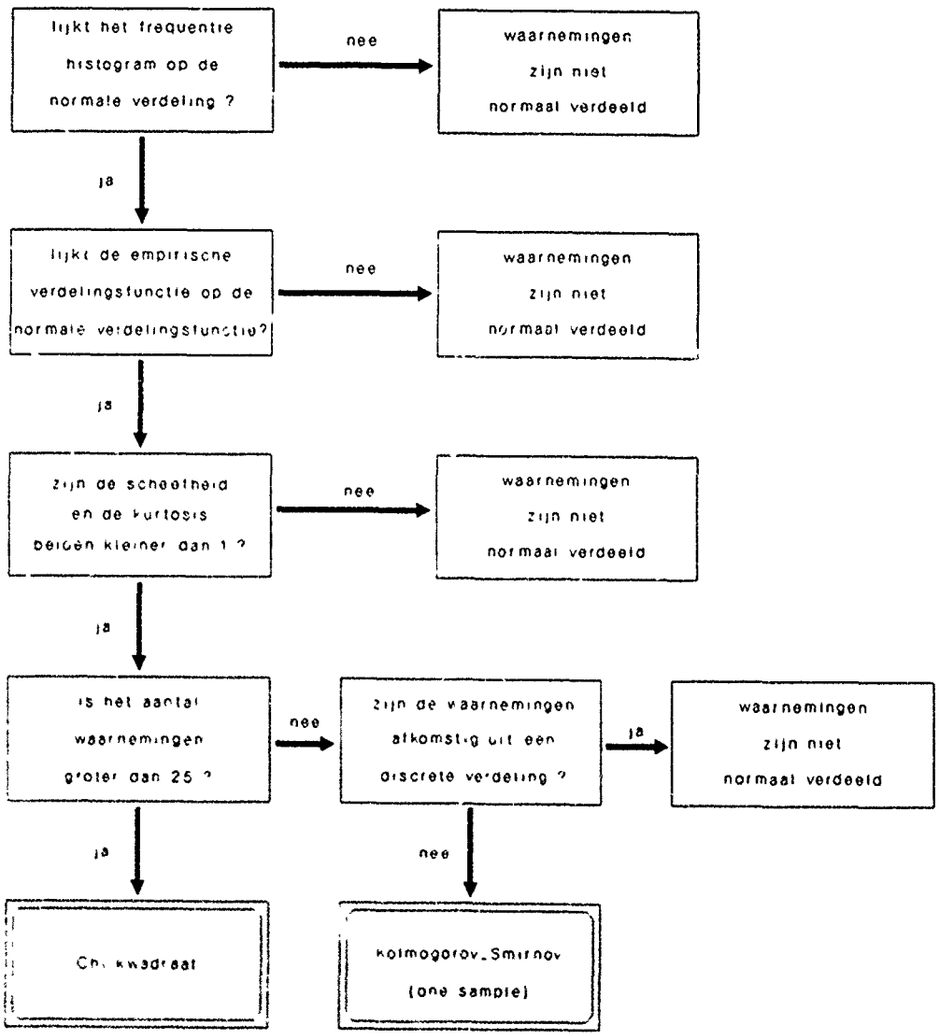

Fig. 2: Schematische overzicht van normaliteitsonderzoek

3 MATEN VAN EFFECTIVITEIT

Dit hoofdstuk geeft een beschrijving van verschillende maten van effectiviteit (MOE's) voor gevechtssimulaties. Eerst wordt een overzicht van een aantal MOE's gegeven. Vervolgens worden deze MOE's nader beschreven, en wordt er aangegeven in welk soort gevecht ze gebruikt kunnen worden. De geschiktheid van een MOE hangt namelijk af van het soort gevecht dat gevoerd wordt.

3.1 Overzicht

Enkele veelgebruikte MOE's zijn:

1. Relatieve gevechtskracht blauw
2. Relatieve verliezen rood
3. Relatieve gevechtskracht verhouding
4. Relatieve verlies verhouding
5. Win factor
6. Succes boolean
7. Terreinwinst
8. Tijdsduur

Door ons zelf werd ontwikkeld:

9. Hindernis verdedigingskracht

Deze MOE's worden in de volgende paragraaf uitgewerkt.

3.2 Nadere beschrijving van de MOE's

Om te komen tot een duidelijke beschrijving van de in 3.1 genoemde MOE's, worden de volgende definities gebruikt:

B : Initiële gevechtskracht blauw
R : Initiële gevechtskracht rood
b : Verlies blauw
r : Verlies rood

waarbij de gevechtskracht en de verliezen worden uitgedrukt in aantallen voertuigen. Met behulp van deze definities worden nu de verschillende MOE's beschreven.

Relatieve gevechtskracht blauw

De relatieve gevechtskracht blauw wordt gedefinieerd als de huidige gevechtskracht van blauw gedeeld door de initiële gevechtskracht van blauw. Deze MOE geeft alleen de eigen sterkte weer. In het vervolg wordt deze MOE aangeduid met FORCE_BLUE. FORCE_BLUE kan als volgt worden uitgedrukt:

$$\text{FORCE\_BLUE} = \frac{B\text{-}b}{B}$$

Het bereik van deze MOE is [0,1].

Relatieve verliezen rood

Relatieve verliezen rood wordt gedefinieerd als de verliezen van rood gedeeld door de initiële gevechtskracht van rood. Deze MOE geeft het verlies van de vijand weer, en kan als volgt worden uitgedrukt:

$$\text{LOSS\_RED} = \frac{r}{R}$$

Ook deze MOE ligt tussen 0 en 1.

Relatieve gevechtskracht verhouding

De relatieve gevechtskracht verhouding wordt gedefinieerd als de verhouding van de gevechtskracht van blauw tot de gevechtskracht van rood gedeeld door de initiële gevechtskracht verhouding. Deze MOE geeft de relatieve sterkte van blauw ten opzichte van rood weer. In formulevorm ziet deze MOE er als volgt uit:

$$\text{FORCE\_RATIO} = \frac{B\text{-}b/R\text{-}r}{B/R}$$

Deze MOE ligt tussen 0 en oneindig.

Relatieve verlies verhouding

De relatieve verlies verhouding wordt gedefinieerd als de verhouding van de verliezen van rood tot de verliezen van blauw gedeeld door de initiële gevechtskracht verhouding. Deze MOE geeft de relatieve verliezen van rood ten opzichte van blauw weer, en kan als volgt worden uitgedrukt:

$$\text{LOSS\_RATIO} = \frac{r\ /b}{R\ /B}$$

Deze MOE ligt tussen 0 en oneindig.

Win factor

De win factor is een combinatie van de relatieve gevechtskracht verhouding en de relatieve verlies verhouding. Deze kan als volgt geschreven worden:

$$\text{WIN\_FACTOR} = \frac{\text{FORCE\_RATIO} + \text{LOSS\_RATIO}}{2}$$

Ook de. ÷ MOE ligt tussen 0 en oneindig.

Opdracht geslaagd

Deze MOE geeft weer of een opdracht geslaagd is of niet. Wanneer de opdracht geslaagd is, krijgt deze MOE de waarde 1; wanneer de opdracht niet geslaagd is, de waarde 0.

$$\text{SUCCES} = \begin{cases} 1 & \text{als opdracht geslaagd} \\ 0 & \text{als opdracht niet geslaagd} \end{cases}$$

Deze MOE kan dus alleen de waarde 0 of 1 aannemen.

Terreinwinst

Deze MOE geeft weer hoeveel terreinwinst er door de aanvallende partij geboekt is. Ervan uitgaande dat rood de aanvallende partij is, kan er
geschreven worden:

TERRAIN = min (x-coordinaten rood)

Bij de bepaling van deze MOE wordt aangenomen dat rood uit het oosten aanvalt. Deze MOE wordt groter naarmate rood meer terreinwinst boekt.

Is blauw de aanvallende partij, dan kan de terreinwinst als volgt uitgedrukt worden

TERRAIN = max (x-coordinaten blauw)

waarbij ervan uitgegaan wordt dat blauw uit het westen aanvalt.

Tijdsduur

De tijdsduur geeft weer hoe lang blauw zijn verdediging heeft gehandhaafd. Wanneer rood zich terugtrekt heeft deze MOE een vaste waarde T, wanneer blauw zich terugtrekt geeft deze MOE het tijdstip t waarop blauw zich terugtrekt:

$$\text{DURATION} = \begin{cases} T & \text{als rood zich terugtrekt} \\ t & \text{als blauw zich terugtrekt} \end{cases}$$

waarin T constant en $T > t$.

Hindernis verdedigingskracht

De hindernis verdedigingskracht geeft een waardering aan een opstelling van eenheden, die een hindernis verdedigen. Hiertoe wordt berekend wat de gemiddelde kans is, dat een vijandelijke tank die de hindernis betreedt, wordt uitgeschakeld. Hierbij wordt in beschouwing genomen welke verdedigende eenheden zicht hebben op (delen van) de hindernis.

De berekening van deze kans geschiedt door de hindernis op te delen in N stukjes van 5 meter. Vervolgens wordt voor elk stukje h_i bepaald welke blauwe eenheden line-of-sight met dit stukje hindernis hebben. Voor elke eenheid j met line-of-sight wordt bepaald wat de kill-kans P_{ij} is tegenover een tank die zich op h_i bevindt.

Met behulp van de formule

$$P_i = 1 - \prod_j (1-P_{i,j})$$

kan nu de kill-kans tegenover een tank op h_i, worden berekend. De gemiddelde kill-kans wordt nu berekend door het gemiddelde over alle stukjes h_i te nemen:

$$OBST\_DEFENCE = \frac{1}{N} \sum_i P_i$$

Deze MOE ligt tussen 0 en 1.

3.3 Bruikbaarheid van de MOE's

De bruikbaarheid van de MOE's hangt af van het soort gevecht dat gevoerd wordt. De volgende gevechtsvormen worden onderscheiden:

1. Verdedigend gevecht:
   Gevechtsvorm die ten doel heeft de aanvaller zo mogelijk de toegang tot, maar tenminste de doorgang door het ter verdediging toegewezen vak - onder behoud van het tactisch essentieel gebied - te ontzeggen.

2. Vertragend gevecht:
   Gevechtsvorm met als doel de aanvaller over een opgedragen diepte van het toegewezen vak op te houden en hem verliezen toe te brengen, zodat het vervolgens voeren van een verdedigend of aanvallend gevecht wordt begunstigd.

3. Aanvallend gevecht:
   Gevechtsvorm waarbij de beschikbare middelen op een door de (hogere) commandant te bepalen tijd en plaats worden ingezet met als doel het vermeesteren van een tactisch essentieel gebied, waarbij de vijand de mogelijkheid en de wil wordt ontnomen verder weerstand te bieden.

Binnen het verdedigend gevecht bestaat de mogelijkheid om een tegenaanval of een tegenstoot te plaatsen:

- Tegenaanval:
  Een in het operatieplan al dan niet voorziene aanval met het doel tactisch belangrijk gebied te hernemen of verloren gaan van tactisch essentieel gebied te voorkomen.
- Tegenstoot:
  Een in het operatieplan al dan niet voorziene aanval met het doel de vijand zoveel mogelijk verliezen toe te brengen.

De eerste vijf MOE's uit het overzicht in 3.1 kunnen bij elk soort gevecht worden gebruikt. De eigen sterkte en de sterkte van de vijand zijn namelijk altijd belangrijk, ongeacht het soort gevecht dat gevoerd wordt. De zesde MOE, succes boolean, kan eveneens in elk soort gevecht worden gebruikt, aangezien het uitvoeren van een opdracht in iedere gevechtsvorm voorkomt.

Naast deze zes algemene MOE's zijn er ook MOE's die gebonden zijn aan het soort gevecht dat gevoerd wordt. Wordt er een verdedigend gevecht gevoerd, dan is de terreinwinst van de vijand belangrijk. Hier kan dus de zevende MOE, terreinwinst, worden gebruikt. Dient er in het verdedigend gevecht een hindernis te worden verdedigd, dan is de hindernis verdedigingskracht het meest geschikt als MOE.

In het vertragend gevecht dient de vijand zo lang mogelijk opgehouden te worden. De tijdsduur is hier dus een geschikte MOE. Verder dienen er zoveel mogelijk verliezen aan de vijand te worden toegebracht. Hierbij zijn de relatieve verliezen rood en de relatieve verlies verhouding geschikte MOE's.

Het aanvallend gevecht heeft als doel het bezetten van vijandelijk gebied. De terreinwinst is hier dus een geschikte MOE. Deze MOE kan ook worden gebruikt bij de tegenaanval.

De tegenstoot heeft als doel de vijand zoveel mogelijk verliezen toe te brengen, teneinde de gevechtskrachtverhouding in het eigen voordeel te wijzigen. MOE's die hierbij gebruikt kunnen worden zijn de relatieve verliezen rood, de relatieve verlies verhouding, de relatieve gevechtskracht verhouding en de win factor.

In sommige situaties kan het zinvol zijn om een combinatie van MOE's te gebruiken.

4 GEBRUIKTE STATISTISCHE METHODEN

In dit hoofdstuk worden enkele methoden beschreven, waarmee twee scenario's op basis van een gekozen MOE met elkaar vergeleken kunnen worden. Hiertoe worden voor beide scenario's uit de gedraaide simulatieruns de waarden van de gekozen MOE bepaald. Zijn de verkregen datasets normaal verdeeld, dan kunnen verdelingsafhankelijke methoden worden gebruikt. Zijn de verkregen datasets niet normaal verdeeld, dan dient er voor een verdelingsvrije methode te worden gekozen.

Om te weten te komen of de waarnemingen normaal verdeeld zijn, is het zinvol om de waarnemingen grafisch weer te geven. Dit kan door middel van frequentiehistogrammen en empirische verdelingsfuncties. Vervolgens kan er met behulp van zogenaamde "Goodness of fit" toetsen worden bepaald of het waarschijnlijk is dat de waarnemingen uit een normale verdeling afkomstig zijn.

4.1 Frequentiehistogrammen

Voor het maken van een frequentiehistogram worden de waarnemingen in een aantal klassen opgesplitst. Het aantal waarnemingen binnen een klasse noemt men frequentie. De verdeling van de frequenties over de verschillende klassen wordt frequentieverdeling genoemd. Worden de frequenties gedeeld door het totaal aantal waarnemingen, dan spreekt men van een relatieve frequentieverdeling.

Door de frequentieverdeling uit te zetten in een histogram krijgt men inzicht in de werkelijke verdeling van de waarnemingen.

Het bovenstaande wordt geïllustreerd aan de hand van het volgende voorbeeld:
Met een bepaald scenario werden 50 simulatieruns gedraaid. Uit de resultaten hiervan werd voor elke simulatierun de relatieve gevechtskracht verhouding berekend. De 50 waarnemingen van deze MOE werden in 6 klassen verdeeld waarbij de volgende relatieve frequenties werden waargenomen:

FORCE_RATIO	Frequentie (relatief)
0.5 <= 1.0 :	11 (22.00%)
1.0 <= 1.5 :	23 (46.00%)
1.5 <= 2.0 :	11 (22.00%)
2.0 <= 2.5 :	3 (6.00%)
2.5 <= 3.0 :	0 (0.00%)
3.0 <= 3.5 :	2 (4.00%)

In de volgende figuur zijn de relatieve frequenties uitgezet in een histogram, waarbij tevens de frequentieverdeling van een normale verdeling met hetzelfde gemiddelde en dezelfde standaardafwijking is weergegeven.

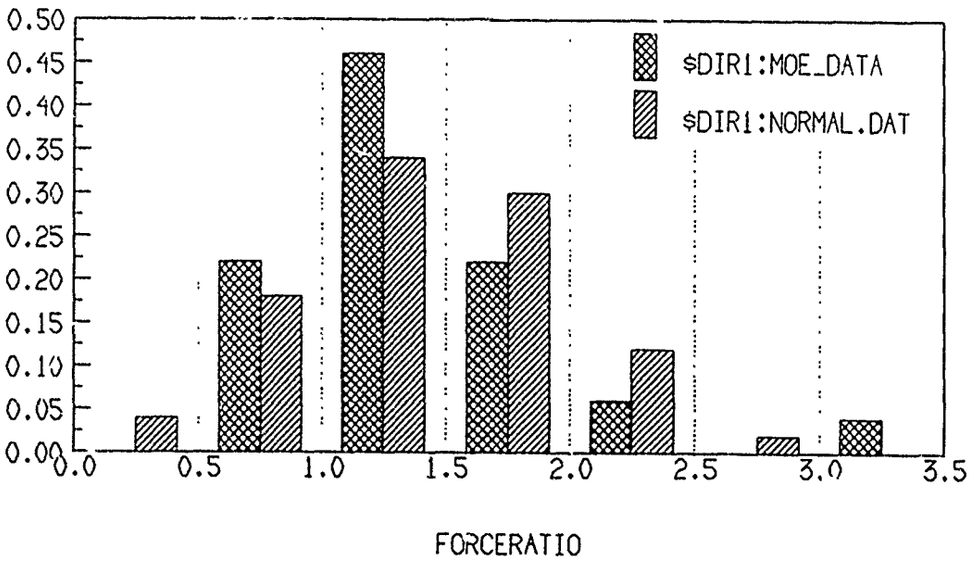

Fig. 3: Frequentiehistogram

Op basis hiervan kan geconcludeerd worden dat het twijfelachtig is of de MOE uit ons voorbeeld nonnaal verdeeld is. Het is dus nog maar de vraag of voor deze dataset een verdelingsvrije methode kan worden gebruikt.

4.2 Empirische verdelingsfuncties

De empirische verdelingsfunctie wordt gedefinieerd als

$F_n(x) = k/n$ wanneer k van de in totaal n waarnemingen kleiner zijn dan x.

$F_n(x)$ is een benadering van de werkelijke verdelingsfunctie

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

waarin f de kansdichtheidsfunctie van X is.

De volgende figuur geeft de empirische verdelingsfunctie van de MOE uit het voorbeeld in paragraaf 4.1, waarbij tevens de verdelingsfunctie van een normale verdeling met hetzelfde gemiddelde en dezelfde variantie is weergegeven.

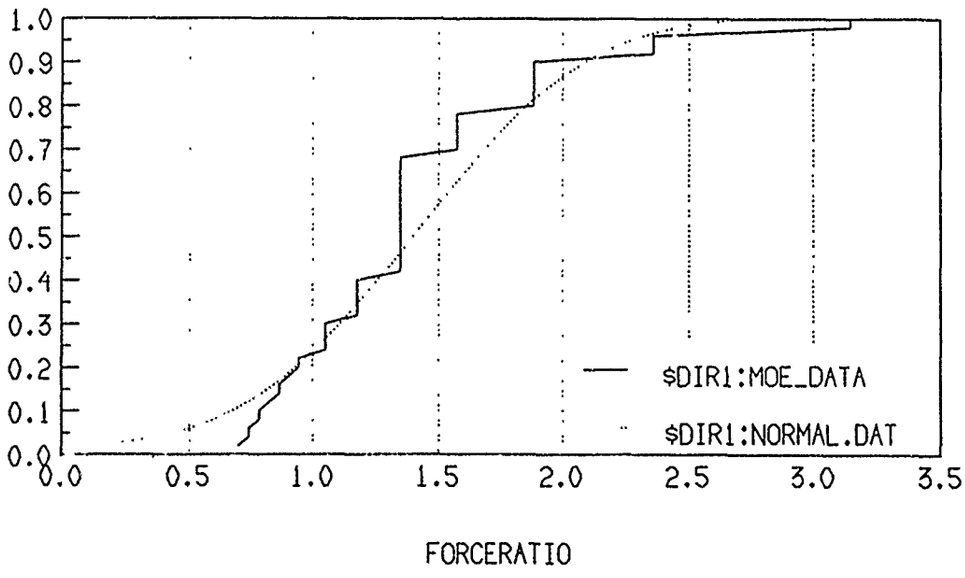

Fig. 4: Empirische verdelingsfunctie

Ook hier kan geconcludeerd worden dat het twijfelachtig is of de MOE normaal verdeeld is.

4.3 Statistieken

Een andere manier om inzicht te krijgen in de verdeling van een MOE, is door te kijken naar een aantal statistieken. Dit zijn bepaalde grootheden die uit de waarnemingen worden berekend.

Aantal waarnemingen

Het aantal waarnemingen geeft de grootte van een set waarnemingen weer. Uitgaande van een set waarnemingen $x_1..x_n$, wordt het aantal waarnemingen gegeven door n.

Minimum

Het minimum geeft de laagste waarde uit de waarnemingen weer:

$$x_{min} = \min_i (x_i)$$

Maximum

Het maximum geeft de hoogste waarde uit de waarnemingen weer:

$$x_{max} = \max_i (x_i)$$

Gemiddelde

Het gemiddelde geeft de algemene ligging van een set waarnemingen weer. Veelal wordt onder het gemiddelde het rekenkundig gemiddelde verstaan. De definitie voor het rekenkundig gemiddelde luidt:

$$\bar{x} = \frac{1}{n} \sum_i x_i$$

Standaardafwijking

De standaardafwijking is een maat voor de spreiding van de waarnemingen rondom het gemiddelde. De definitie voor de standaardafwijking luidt:

$$s = \sqrt{\frac{1}{n-1} \sum_i (x_i - \bar{x})^2}$$

Variantie

De variantie is het kwadraat van de standaardafwijking en wordt gedefinieerd door

$$s^2 = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2$$

Scheefheid

De scheefheid is een maat voor de symmetrie van een verdeling. Een symmetrische verdeling heeft scheefheid 0. De definitie voor de scheefheid luidt:

$$\alpha_3 = \frac{\sum_i (x_i - \bar{x})^3/n}{[\sum_i (x_i - \bar{x})^2/n]^{3/2}}$$

Kurtosis

De kurtosis geeft de welving van een verdeling weer. Een verdeling met een hoge top en een dikke staart heeft een positieve kurtosis, een verdeling met een lage top en een dunne staart heeft een negatieve kurtosis. Een normale verdeling heeft kurtosis 0.

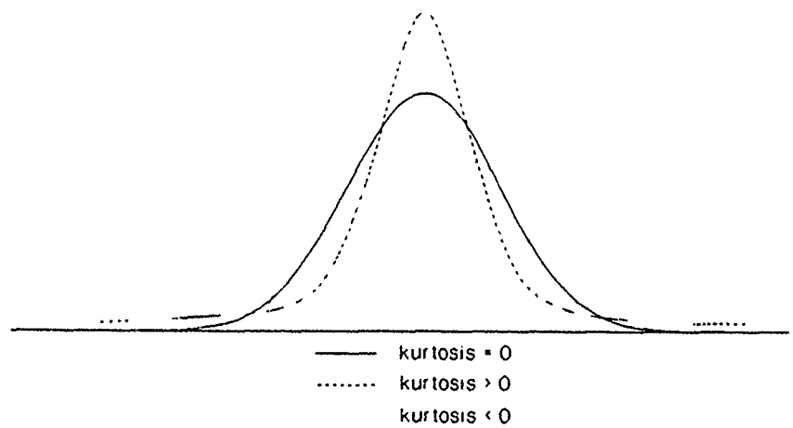

Fig. 5: Verdelingen met verschillende kurtosis.

De definitie voor kurtosis luidt:

$$\gamma_2 = \frac{\sum_i (x_i - \bar{x})^4/n}{[\sum_i (x_i - \bar{x})^2/n]^2} - 3$$

Om na te gaan of een set waarnemingen een normale verdeling heeft, is het zinvol om te kijken naar de scheefheid en de kurtosis. Deze moeten beiden ongeveer gelijk aan 0 zijn.

Om te bepalen hoeveel de scheefheid en de kurtosis van 0 af mogen wijken, is 1000 maal de scheefheid en de kurtosis van een set van 50 waarnemingen uit een standaardnormale verdeling bepaald. De 1000 waarden van de scheefheid en kurtosis zijn in de volgende frequentietabel weergegeven:

	Frequentie (relatief)	
	Scheefheid	Kurtosis
-1.5 <= -1.0 :	3 (0.30%)	14 (1.40%)
-1.0 <= -0.5 :	60 (6.00%)	264 (26.40%)
-0.5 <= 0.0 :	454 (45.40%)	366 (36.60%)
0.0 <= 0.5 :	427 (42.70%)	220 (22.00%)
0.5 <= 1.0 :	55 (5.50%)	84 (8.40%)
1.0 <= 1.5 :	1 (0.10%)	27 (2.70%)
1.5 <= 2.0 :	0 (0.00%)	16 (1.60%)
2.0 <= 2.5 :	0 (0.00%)	7 (0.70%)
2.5 <= 3.0 :	0 (0.00%)	0 (0.00%)
3.0 <= 3.5 :	0 (0.00%)	1 (0.10%)
3.5 <= 4.0 :	0 (0.00%)	1 (0.10%)

Dit betekent grofweg dat voor een set van 50 waarnemingen de absolute waarden van de scheefheid en kurtosis kleiner dan 1.0 moeten zijn. Is dit niet het geval, dan is het niet waarschijnlijk dat de waarnemingen normaal verdeeld zijn.

De waarnemingen uit het voorbeeld in paragraaf 4.1 geven de volgende statistieken:

FORCE_RATIO	
Aantal waarnemingen	50.0
Minimum	0.6974
Maximum	3.1453
Gemiddelde	1.4117
Standaardafwijking	0.5516
Variantie	0.3043
Scheefheid	1.3536
Kurtosis	2.0225

De scheefheid en de kurtosis zijn beiden groter dan 1.0, zodat het niet waarschijnlijk is dat de waarnemingen uit een normale verdeling afkomstig zijn.

4.4 Goodness of fit

In de voorgaande paragrafen zijn enkele methoden besproken om na te gaan of waarnemingen van een stochastische variabele normaal verdeeld zijn. Een andere methode om dit na te gaan is met behulp van de zogenaamde "Goodness of fit" toetsen. Deze toetsen geven weer of een set waarnemingen past bij een bepaalde verdeling (in dit geval de normale verdeling). Hiervoor kan een Kolmogorov-Smirnov of een Chi-kwadraat toets worden gebruikt. De Chi-kwadraat toets kan alleen worden gebruikt wanneer de dataset veel waarnemingen bevat. Bij een klein aantal waarnemingen kan beter de Kolmogorov-Smirnov toets worden gebruikt.

4.4.1 Kolmogorov-Smirnov

Er zijn twee versies van de Kolmogorov-Smirnov toets; de one-sample en de two-sample toets. De one-sample toets vergelijkt de verdeling van de waarnemingen met een theoretische (continue) verdelingsfunctie. De two-sample toets vergelijkt de verdelingen van twee sets waarnemingen. De one-sample toets kan voor goodness of fit worden gebruikt. De two-sample toets komt later aan de orde.

In de one-sample toets wordt de volgende nulhypothese getoetst:

H_0: $F(x) = G(x)$

waarin:

F(x): de verdelingsfunctie van de waargenomen variabele
G(x): de verdelingsfunctie van de veronderstelde verdeling

De toetsingsgrootheid van deze toets is

$$D_{max} = \max_{x} (ABS (F_n(x)-G(x)))$$

waarin:

$F_n(x)$: de empirische verdelingsfunctie van de waargenomen variabele.

Deze toetsingsgrootheid heeft onder de nulhypothese een bekende verdeling. Gegeven een onbetrouwbaarheid α is nu de kritieke waarde R te bepalen zodanig dat

$$P(D_{max} > R) = \alpha$$

Voor $n \geq 35$ geldt dat

$R \approx 1.22 / \sqrt{n}$	voor $\alpha = 10\%$
$R \approx 1.36 / \sqrt{n}$	voor $\alpha = 5\%$
$R \approx 1.63 / \sqrt{n}$	voor $\alpha = 1\%$

Wordt er nu een toetsingsgrootheid groter dan de kritieke waarde waargenomen, dan is het niet aannemelijk (behoudens een onbetrouwbaarheidsmarge α) dat de toetsingsgrootheid de veronderstelde verdeling heeft. In dit geval wordt de nulhypothese dus verworpen. Dit betekent dat er aangenomen wordt dat de verdelingsfunctie van de waargenomen variabele en de opgegeven verdelingsfunctie verschillen.

Wanneer voor G de normale verdeling wordt genomen, is het mogelijk om te toetsen of de waarnemingen afkomstig zijn uit een normale verdeling. De toets is echter minder geschikt voor gevallen waarin de verdeling van de waarnemingen discreet is. Aangezien dit in het eerder genoemde voorbeeld het geval was, kon deze toets hier niet worden gebruikt.

4.4.2 Chi-kwadraat

De chi-kwadraat toets heeft als nulhypothese:

H_0: $F(x) = G(x)$

waarin

F(x): de verdelingsfunctie van de waargenomen variabele
G(x): de verdelingsfunctie van de veronderstelde verdeling

Bij de chi-kwadraat toets worden de waarnemingen opgesplitst in een aantal klassen, aan te duiden met $E_1,E_2,..,E_k$. De waargenomen frequenties behorende bij deze klassen worden gegeven door $f_1,f_2,..,f_k$. Op grond van de in de nulhypothese gegeven verdelingsfunctie worden de frequenties $g_1,g_2,..,g_k$ verwacht. Indien de nulhypothese juist is, zullen de verschillen tussen de waargenomen frequenties f_i en de verwachte frequenties g_i klein zijn.

Aan de hand van deze verschillen kan de toetsingsgrootheid worden berekend.

$$\chi^2 = \sum_i \frac{(f_i - g_i)^2}{g_i}$$

Is de waarde van deze grootheid klein, dan is er geen reden om H_0 te verwerpen. Is de waarde van χ^2 groot, dan zal H_0 verworpen worden.

Gegeven een onbetrouwbaarheid α, kan de kritieke waarde voor χ^2 uit de tabel van de chi-kwadraat-verdeling worden afgelezen. Hierbij moet voor het aantal vrijheidsgraden k-1-m worden genomen, waarin m het aantal te schatten parameters voor de theoretische verdeling is (bij toetsing tegen een normale verdeling zijn dit er 2). Wordt er een waarde van χ^2 waargenomen, die groter is dan de kritieke waarde uit de tabel, dan zal de nulhypothese verworpen worden. Voorwaarde voor deze toets is dat de verwachte frequenties g_i allen groter dan 5 zijn.

Bij toetsing tegen een normale verdeling, zal in de nulhypothese voor G de normale verdelingsfunctie worden genomen:

$$G(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{1}{2}((t-\mu)/\sigma)^2} dt$$

Het gemiddelde μ en de standaardafwijking σ kunnen uit de waarnemingsuitkomsten worden geschat door het steekproefgemiddelde $\bar{x}$ en de steekproefstandaardafwijking s. Toetsing van de verdeling van X tegen een normale verdeling met gemiddelde $\bar{x}$ en standaardafwijking s komt overeen met toetsing van de verdeling van $Z = (X-\bar{x})/s$ tegen een standaard-normale verdeling.

Bovenstaande procedure werd uitgevoerd met de set waarnemingen uit het voorbeeld in paragraaf 4.1, hetgeen resulteerde in een verwerping van de nulhypothese. Het is dus niet waarschijnlijk dat deze set waarnemingen afkomstig is uit een normale verdeling.

4.5 Betrouwbaarheidsintervallen

De ligging van een reeks waarnemingen kan worden gerepresenteerd door het steekproefgemiddelde $\bar{x} = 1/n \Sigma x_i$. Dit steekproefgemiddelde is een schatter voor de verwachting μ van de verdeling waar de waarnemingen uit afkomstig zijn. De bedoeling is nu om aan de hand van de waarnemingen x_i een betrouwbaarheidsinterval voor μ te construeren. Hierbij ligt het voor de hand om gebruik te maken van de centrale limietstelling:

> Gegeven dat de stochastische variabelen $X_1,..,X_n$ onafhankelijk en identiek verdeeld zijn met verwachting μ en variantie σ^2. Dan geldt dat wanneer n groot is of wanneer de X_i normaal verdeeld zijn, dat $\bar{x}$ normaal verdeeld is met verwachting μ en variantie σ^2/n.

Wanneer er aan de voorwaarden voor de centrale limietstelling is voldaan, geldt dat

$$u = \frac{\bar{x}-\mu}{\sigma\sqrt{n}}$$

standaard normaal verdeeld is. Hierin is σ echter nog onbekend. σ kan geschat worden door de steekproefstandaardafwijking s. In dat geval heeft

$$t = \frac{\bar{x}-\mu}{s\sqrt{n}}$$

een Student's t-verdeling met n-1 vrijheidsgraden.
De rechter kritieke waarde van deze verdeling is uit de tabel van de Student's T-verdeling te halen. Deze wordt aangeduid met $t_\alpha(\nu)$, waarin α de overschrijdingskans en ν het aantal vrijheidsgraden.
Er geldt dus

$$P(-t_{½\alpha}(n-1) \leq t \leq t_{½\alpha}(n-1)) = 1 - \alpha$$

Dus met kans $1-\alpha$ geldt

$$\bar{x} - t_{½\alpha}(n-1)\, s/\sqrt{n} \leq \mu \leq \bar{x} + t_{½\alpha}(n-1)\, s/\sqrt{n}$$

Dit wordt het $1-\alpha$ betrouwbaarheidsinterval voor μ genoemd.
Bij het construeren van dit interval is er vanuit gegaan dat er aan de voorwaarden van de centrale limietstelling is voldaan. Dus

n is groot
óf X is normaal verdeeld

Het is gebleken dat de waarnemingen uit het voorbeeld in paragraaf 4.1 niet normaal verdeeld waren. Verder is het twijfelachtig of het aantal waarnemingen van 50 voldoende is om de centrale limietstelling te mogen gebruiken. Het is dus nog maar de vraag of de zojuist afgeleide formule voor een betrouwbaarheidsinterval voor μ in dit geval gebruikt mag worden.

Om nu toch een betrouwbaarheidsinterval voor de verwachting van deze set waarnemingen te bepalen, wordt gebruik gemaakt van een methode die geen eisen stelt ten aanzien van het aantal waarnemingen of normaliteit: de bootstrap methode.

4.6 Bootstrap methode

De bootstrap methode biedt de mogelijkheid om voor een onbekende parameter θ met behulp van een schatter $\hat{\theta}$ een betrouwbaarheidsinterval te berekenen, zonder dat daarbij aannamen worden gedaan over normaliteit en zonder dat er eisen worden gesteld aan het aantal waarnemingen. Hierbij wordt gebruik gemaakt van Monte Carlo simulatie.

Uitgaande van een steekproef met n waarnemingen, worden er uit deze steekproef n trekkingen met teruglegging gedaan. De aldus verkregen set van waarnemingen wordt een bootstrap steekproef genoemd. Deze procedure wordt 1000 maal herhaald. Vervolgens wordt voor elke bootstrap steekproef de bootstrap schatter $\hat{\theta}^*$ bepaald (het sterretje geeft aan dat de schatter betrekking heeft op een bootstrap steekproef).

De verkregen bootstrap schattingen kunnen in een frequentiehistogram worden uitgezet. Dit geeft een beeld van de verdeling van $\hat{\theta}$. Om nu een 1 - α betrouwbaarheidsinterval voor θ te berekenen, wordt gebruik gemaakt van de percentile methode (zie [Efron 2]).
Zij G(s) de verdelingsfunctie van $\hat{\theta}^*$,

$$G(s) = P(\hat{\theta}^* \leq s)$$

Dan is $[G^{-1}(\frac{1}{2}\alpha), G^{-1}(1-\frac{1}{2}\alpha)]$ een benadering voor het 1 - α betrouwbaarheidsinterval voor θ. Om de percentielen $G^{-1}(\frac{1}{2}\alpha)$ en $G^{-1}(1-\frac{1}{2}\alpha)$ te bepalen, wordt G benaderd door de empirische verdelingsfunctie G_{1000}. Om het $\frac{1}{2}\alpha$ percentiel van G_{1000} te bepalen, worden de waargenomen bootstrap schattingen geordend. De geordende bootstrap schattingen worden $\hat{\theta}_1^*, \hat{\theta}_2^*, .. , \hat{\theta}_{1000}^*$ genoemd. Wanneer voor de k^e bootstrap schatting $\hat{\theta}_k^*$ geldt dat

$$G_{1000}(\hat{\theta}_k^*) \leq \tfrac{1}{2}\alpha < G_{1000}(\hat{\theta}_{k+1}^*)$$

dan is $\hat{\theta}_k^*$ het $\frac{1}{2}\alpha$ percentiel van G_{1000}. Op dezelfde manier kan het 1 - $\frac{1}{2}\alpha$ percentiel van G_{1000} worden bepaald. Het aldus gevonden interval geeft een benadering van het 1 - α betrouwbaarheidsinterval voor θ.

Een betrouwbaarheidsinterval voor de verwachting μ kan nu verkregen worden door het bovenstaande toe te passen met als schatter het gemiddelde.

4.7 Datasets vergelijken

Er zijn verschillende statistische methoden om sets waarnemingen met elkaar te vergelijken. Deze methoden worden onderscheiden in verdelingsafhankelijke en verdelingsvrije methoden.

4.7.1 Verdelingsafhankelijke methoden

Enkele verdelingsafhankelijke methoden zijn de Student's t-toets, de toets van Satterthwaite en de F-toets voor varianties. Bij deze methoden wordt er aangenomen dat de waarnemingen normaal verdeeld zijn. De Student's t-toets en de toets van Satterthwaite toetsen of twee sets waarnemingen dezelfde verwachting hebben, de F-toets voor varianties toetst of twee sets waarnemingen dezelfde verwachting hebben.

4.7.1.1 Student's t-toets

De Student's t-toets kan worden gebruikt om te toetsen of er verschil in de verwachtingen van twee sets waarnemingen is. De toets heeft als voorwaarde dat de waarnemingen uit beide sets normaal verdeeld zijn, en dat de varianties van beide verdelingen gelijk zijn.

Gegeven zijn 2 sets waarnemingen:

$x_1, x_2, \ldots, x_m$

$y_1, y_2, \ldots, y_n$

Wanneer $x_1 .. x_m$ afkomstig zijn uit een verdeling met verwachting μ_1 en variantie σ^2 en $y_1 .. y_n$ uit een verdeling met verwachting μ_2 en variantie σ^2, dan is de nulhypothese voor de t-toets:

$$H_0 : \mu_1 = \mu_2$$

De toetsingsgrootheid is

$$t = \frac{\bar{x} - \bar{y}}{s \sqrt{\frac{1}{m} + \frac{1}{n}}}$$

waarin s de wortel is van het gewogen gemiddelde van de steekproefvarianties s_1^2 en s_2^2 :

$$s = \sqrt{\frac{(m-1)\,s_1^2 + (n-1)\,s_2^2}{m+n-2}}$$

De toetsingsgrootheid heeft onder de nulhypothese een Student's t-verdeling met m+n-2 vrijheidsgraden. Uit de tabel van de Student's t-verdeling zijn nu de kritieke waarden R_1 en R_2 te halen zodanig dat

$$P(t < R_1) = \tfrac{1}{2}\alpha$$
$$P(t > R_2) = \tfrac{1}{2}\alpha$$

Wordt er een toetsingsgrootheid kleiner dan R_1 of groter dan R_2 waargenomen, dan wordt de nulhypothese verworpen.

Ook is het mogelijk om een 1-α betrouwbaarheidsinterval voor μ_1-μ_2 te construeren :

$$\bar{x}-\bar{y} - t_{\frac{1}{2}\alpha}(m+n-2)\,s\,\sqrt{\frac{1}{m}+\frac{1}{n}} < \mu_1-\mu_2 < \bar{x}-\bar{y} + t_{\frac{1}{2}\alpha}(m+n-2)\,s\,\sqrt{\frac{1}{m}+\frac{1}{n}}$$

Ligt de waarde 0 binnen dit interval, dan is er geen reden om de nulhypothese te verwerpen.

4.7.1.2 Toets van Satterthwaite

De t-toets mag alleen toegepast worden, wanneer er aan de voorwaarde van gelijke varianties is voldaan. Is dit niet het geval, dan kan er een aangepaste versie van de t-toets worden gebruikt die deze voorwaarde niet heeft. Deze toets wordt de toets van Satterthwaite genoemd. Toetsingsgrootheid voor deze toets is

$$t' = \frac{\bar{x}-\bar{y}}{s_d}$$

waarin

$$s_d = \sqrt{\frac{s_1^2}{m}+\frac{s_2^2}{n}}$$

Onder de nulhypothese benadert de verdeling van deze toetsingsgrootheid de t-verdeling met f vrijheidsgraden, waarbij f wordt gegeven door

$$f = \frac{s_d^4}{\frac{(s_1^2/m)^2}{m-1} + \frac{(s_2^2/n)^2}{n-1}}$$

Op dezelfde manier als bij de t-toets kunnen er nu kritieke waarden worden bepaald en betrouwbaarheidsintervallen worden berekend.

4.7.1.3 F-toets voor varianties

De F-toets voor varianties kan gebruikt worden om te toetsen of twee sets waarnemingen dezelfde variantie hebben. Wanneer de eerste set bestaat uit m waarnemingen uit een normale verdeling met verwachting μ_1 en variantie σ_1^2, en de tweede set uit n waarnemingen uit een normale verdeling met verwachting μ_2 en variantie σ_2^2, dan is de nulhypothese voor deze toets:

$$H_0 : \sigma_1^2 = \sigma_2^2$$

De toetsingsgrootheid voor deze toets is

$$F = \frac{s_1^2}{s_2^2}$$

waarin $s_1^2 \geq s_2^2$ (is dit niet het geval, verander dan de nummering van beide sets). Onder de nulhypothese heeft F een F-verdeling met m-1 en n-1 vrijheidsgraden. De kritieke waarde voor F is nu uit de tabel van de F-verdeling af te lezen, en bij overschrijding van deze waarde wordt de nulhypothese verworpen.

4.7.2 Verdelingsvrije methoden

Enkele verdelingsvrije methoden zijn de Kolmogorov-Smirnov two-sample toets, de toets van Wilcoxon en een variant van de bootstrap methode. Deze methoden doen geen aannamen over de verdeling van de waarnemingen. De Kolmogorov-Smirnov two-sample toets gaat na of twee sets waarnemingen uit dezelfde verdeling afkomstig zijn. De toets van Wilcoxon en de variant van de bootstrap methode toetsen of twee sets waarnemingen dezelfde verwachting hebben.

4.7.2.1 Kolmogorov-Smirnov (two-sample)

De Kolmogorov-Smirnov two-sample toets kan gebruikt worden om te toetsen of het waarschijnlijk is dat twee sets waarnemingen uit dezelfde verdeling afkomstig zijn. Hierbij is het niet noodzakelijk dat beide sets evenveel waarnemingen hebben. De two-sample toets heeft als nulhypothese

$$H_0 : F(x) = G(x)$$

waarin

F(x): de verdelingsfunctie van de eerste set waarnemingen
G(x): de verdelingsfunctie van de tweede set waarnemingen

De nulhypothese is dus dat beide datasets dezelfde verdelingsfunctie hebben.

De toetsingsgrootheid van deze toets is

$$D_{max} = \max_x (\text{abs} (F_n(x) - G_m(x)))$$

waarin $F_n(x)$ de empirische verdelingsfunctie van de eerste dataset en $G_m(x)$ de empirische verdelingsfunctie van de tweede dataset.
Deze toetsingsgrootheid heeft onder de nulhypothese een bekende verdeling. Gegeven een onbetrouwbaarheid α is nu de kritieke waarde R te bepalen waarvoor geldt dat

$$P(D_{max} > R) = \alpha$$

Wanneer er een toetsingsgrootheid groter dan de kritieke waarde wordt waargenomen, wordt de nulhypothese verworpen. Het is dan niet aannemelijk (behoudens een onbetrouwbaarheidsmarge α) dat beide sets waarnemingen uit dezelfde verdeling afkomstig zijn.

In plaats van te toetsen of de verdelingsfuncties gelijk zijn, is het ook mogelijk om te toetsen of de verdelingsfuncties niet meer dan een bepaalde tolerantie δ van elkaar verschillen. Deze toets heeft als nulhypothese

$H_0 : F(x-\delta) \leq G(x) \leq F(x+\delta)$

De naar rechts verschoven functie $F(x-\delta)$ wordt verkregen door bij de oorspronkelijke waarnemingen δ op te tellen, de naar links verschoven functie $F(x+\delta)$ wordt verkregen door van de oorspronkelijke waarnemingen δ af te trekken. De volgende figuur geeft een voorbeeld van twee verdelingsfuncties die niet voldoen aan de hypothese $F(x) = G(x)$, maar voor een bepaalde δ wel voldoen aan H_0:

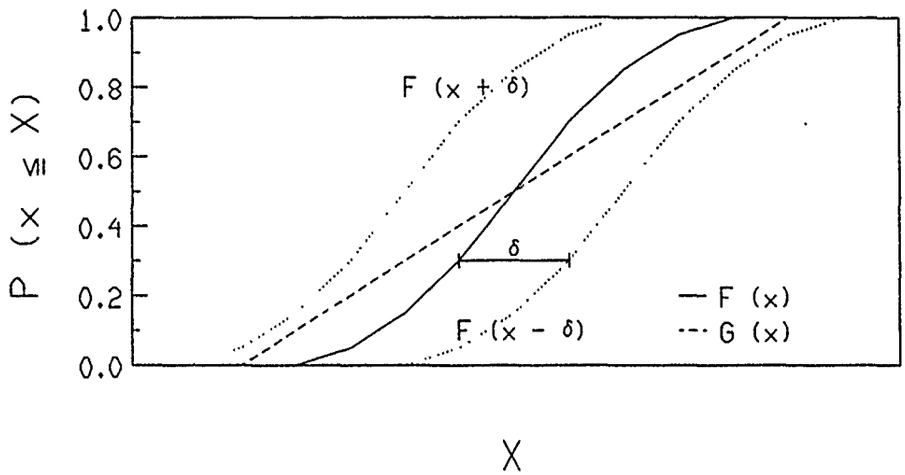

Fig. 6: Voorbeeld

Eenvoudig is aan te tonen dat de nulhypothese H_0 overeenkomt met

$H_0 : G(x-\delta) \leq F(x) \leq G(x+\delta)$

zodat het niet uitmaakt welke verdelingsfunctie $F(x)$ wordt genoemd, en welke verdelingsfunctie $G(x)$.

Is de tolerantie δ gelijk aan 0, dan geeft dit de al eerder beschreven two-sample toets.

De toetsingsgrootheid voor de toets met tolerantie is

$$D_{max} = \max (D^+, D^-)$$

waarin: $D^+ = \max_x (F_n(x-\delta) - G_m(x))$

$$D^- = \max_x (G_m(x) - F_n(x+\delta))$$

Voor deze toetsingsgrootheid kan weer de kritieke waarde worden bepaald. Wordt er een toetsingsgrootheid groter dan de kritieke waarde waargenomen, dan wordt de nulhypothese verworpen. Dit houdt in dat het niet aannemelijk is dat de functie G(x) tussen F(x-δ) en F(x+δ) in ligt.

Opmerking: voor δ ongelijk aan 0 is niet bewezen dat D_{max} dezelfde verdeling heeft als voor δ gelijk aan 0. Dit is echter wel aannemelijk.

4.7.2.2 Toets van Wilcoxon

De toets van Wilcoxon kan gebruikt worden om te toetsen of er verschil is tussen de ligging van twee sets waarnemingen. Stel de eerste set waarnemingen is afkomstig uit een verdeling met verwachting μ_1, en de tweede set is afkomstig uit een verdeling met verwachting μ_2. Dan is de nulhypothese voor deze toets

$$H_0 : \mu_1 = \mu_2$$

Voor het bepalen van de toetsingsgrootheid van deze toets worden de waarnemingen van beide sets in een geordende rij gezet (beide sets door elkaar). Vervolgens worden aan de elementen van deze rij rangnummers $Q_1, Q_2, .., Q_{m+n}$ toegekend. De toetsingsgrootheid is nu de som van de rangnummers van de waarnemingen uit de eerste dataset:

$$W = \sum_{i \in I} Q_i$$

Is de waarde van deze toetsingsgrootheid klein, dan zijn de waarnemingen uit de eerste dataset kleiner dan de waarnemingen uit de tweede dataset. Is de waarde van deze toetsingsgrootheid groot, dan zijn de waarnemingen uit de eerste set groter dan die uit de tweede. In beide gevallen zal de nulhypothese verworpen worden. Er zal dus tweezijdig getoetst moeten worden.

De toetsingsgrootheid heeft onder de nulhypothese een bekende verdeling. Gegeven een onbetrouwbaarheid α kunnen nu de kritieke waarden R_1 en R_2 worden bepaald zodanig dat

$$P(W < R_1) = \tfrac{1}{2}\alpha$$
$$P(W > R_2) = \tfrac{1}{2}\alpha$$

Zijn m en n groot genoeg, dan is W normaal verdeeld met verwachting

$$\mu_W = \frac{1}{4}(m+n) * (m+n+1)$$

en variantie

$$\sigma_W^2 = \frac{1}{12} m * n * (m+n+1)$$

De kritieke waarden R_1 en R_2 worden dan gegeven door

$$R_1 = \mu_W - z_{\frac{1}{2}\alpha} * \sigma_W$$
$$R_2 = \mu_W + z_{\frac{1}{2}\alpha} * \sigma_W$$

Voor kleine m en n moeten deze waarden in een tabel worden opgezocht. Wordt er een toetsingsgrootheid kleiner dan R_1 of groter dan R_2 waargenomen, dan wordt de nulhypothese verworpen.

4.7.2.3 Bootstrap variant

De bootstrap methode (zie 4.6) kan worden gebruikt om een betrouwbaarheidsinterval voor het verschil in de verwachtingen van twee verdelingen te berekenen. Hierbij wordt als schatter het verschil tussen de steekproefgemiddelden genomen.

Gegeven zijn 2 sets waarnemingen

$u_1,u_2,..,u_m$

$v_1,v_2,..,v_n$

Wanneer $u_1..u_m$ afkomstig zijn uit een verdeling met verwachting μ_1 en $v_1..v_n$ uit een verdeling met verwachting μ_2, dan is

$$\frac{1}{m}\sum_i u_i - \frac{1}{n}\sum_j v_j$$

een zuivere schatter voor μ_1-μ_2. Nu kan er met behulp van de bootstrap methode een benadering voor het 1-α betrouwbaarheidsinterval voor μ_1-μ_2 worden berekend. De bootstrap steekproeven die hierbij gebruikt worden, bevatten m trekkingen met teruglegging uit $u_1,..,u_m$ en n trekkingen met teruglegging uit $v_1,..,v_n$.

Wanneer 0 binnen het betrouwbaarheidsinterval ligt, is er geen reden om aan te nemen dat μ_1 en μ_2 verschillen. Ligt 0 niet binnen dit interval, dan zal de hypothese dat μ_1 en μ_2 gelijk zijn moeten worden verworpen. In dit geval zijn er twee mogelijkheden:

1. Het interval ligt geheel boven 0 hetgeen impliceert dat μ_1 groter is dan μ_2
2. Het interval ligt geheel beneden 0 hetgeen impliceert dat μ_1 kleiner is dan μ_2

5 DE ONTWIKKELDE PROGRAMMATUUR

In het kader van het berekenen van de MOE's en het uitvoeren van statistische toetsen zijn drie computerprogramma's geschreven. Deze zijn geprogrammeerd in Ada en draaien onder VAX-VMS. Deze drie programma's zijn:

- WRITE_UNITS
  Dit programma haalt de informatie uit de uitvoer van het Force Structure Model (FSM) die nodig is om de MOE's te berekenen. De uitvoer van dit programma is een tekstfile met gegevens over eenheden.
- CALCULATE_MOE
  Dit programma berekent waarden voor een aantal MOE's aan de hand van gegevens over eenheden. De uitvoer is een matrix van getallen waarin per simulatierun de waarden voor een aantal MOE's staan.
- DEC_MOE_VIS
  Dit is een interactief programma dat, aan de hand van één of meerdere matrices met gegevens, statistische analyses uitvoert.

Later in dit hoofdstuk zullen deze drie programma's uitgebreid worden beschreven. Daarbij worden voorbeelden van uitvoer en invoer van de programma's weergegeven door `kleine letters`. Vette letters geven een keuzemogelijkheid weer van het interactieve programma DEC_MOE_VIS. Vanwege afspraken die gemaakt zijn over de uitvoer van programmatuur die in het kader van FSM worden geschreven is deze uitvoer en de interactieve interface in het Engels.

De uitvoer van het ene programma dient als invoer voor een ander programma. Dit is gedaan om de afhankelijkheid van het FSM model kwijt te raken. Ook op andere manieren kan invoer gegenereerd worden voor de programma's CALCULATE_MOE en DEC_MOE_VIS. Figuur 7 geeft schematisch aan hoe de invoer van de verschillende programma's verkregen kan worden:

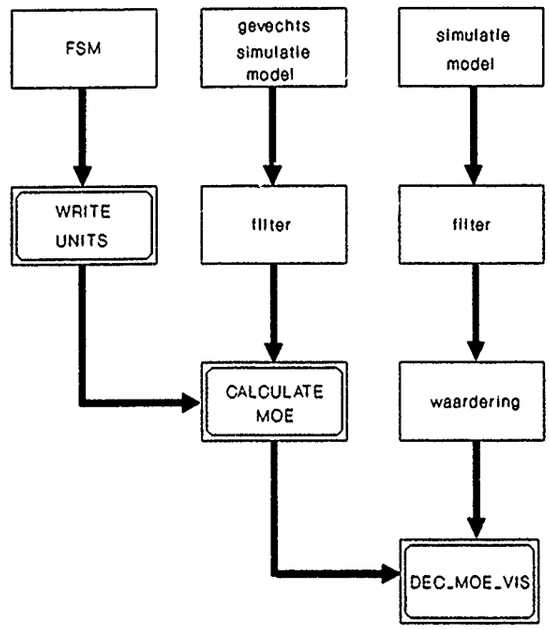

Fig. 7: Schematisch overzicht invoer verschillende programma's

Het programma WRITE_UNITS verkrijgt zijn invoer uit het gevechtssimulatiemodel FSM. Uit de uitvoer van WRITE_UNITS worden met het programma CALCULATE_MOE verschillende MOE's berekend. Dit programma kan ook MOE's voor uitkomsten van andere gevechtssimulatiemodellen berekenen, mits deze uitkomsten, eventueel met behulp van een filter, in het juiste formaat staan. Met behulp van het programma DEC_MOE_VIS kunnen de berekende MOE's vervolgens worden geanalyseerd. Dit programma kan ook gebruikt worden om de uitkomsten van een willekeurig stochastisch simulatiemodel te analyseren. Hiertoe moet aan de uitkomsten van het simulatiemodel een waardering worden toegekend.

5.1 Het computerprogramma WRITE_UNITS

Het programma WRITE_UNITS haalt uit de uitvoer van het FSM model de gegevens waarmee verschillende MOE's berekend kunnen worden. Voordat het programma gedraaid wordt, moet er in een aparte file staan op welk moment van de simulatie de gegevens weggeschreven moeten worden. Hiervoor zijn verschillende mogelijkheden:

- Als de toestand op een bepaalde tijdstip in de simulatie van belang is (bijvoorbeeld de toestand op 10:15:00 uur) dan komt in de file te staan:
  TIME = 10:15:00
- Als de toestand van belang is bij het bereiken van een bepaalde 'marker', dan komt in de file te staan:
  MARKER = 111 DOOR
  Deze markers worden door het FSM model in de uitvoerfile geschreven. Tijdens een simulatie kan door middel van condities in het scenario gezorgd worden voor het leggen van markers. Zo kan er bijvoorbeeld een marker gelegd worden op het moment dat de eenheid 111 door een mijnenveld is gebroken.
- Als de toestand van belang is waarop de sterkte van een bepaalde eenheid of een combinatie van eenheden beneden een op te geven drempel komt. Bijvoorbeeld de regel
  BLUE 30% : 4NT + 4NM + 4NAT < 4
  geeft aan dat de toestand wordt weggeschreven op het moment dat het totaal aantal voertuigen van 4NT, 4 NM en 4NAT kleiner dan of gelijk is aan 4. BLUE 30% geeft een naam aan zodat later kan worden teruggevonden dat de toestand om deze reden is weggeschreven.

In de file kunnen meerdere regels staan waarop de toestand moet worden weggeschreven. De toestand wordt echter maar één keer weggeschreven en wel zo vroeg mogelijk. Als bijvoorbeeld in de file staat:

RED 40% : 11 < 16
BLUE 30% : 4NT + 4NM + 4NAT < 4
TIME = 10:20:00

Dan wordt de toestand weggeschreven

- als het aantal voertuigen van de eenheid 11 kleiner dan of gelijk is aan 16, of
- als het gezamenlijk aantal voertuigen van 4NT, 4NM en 4NAT kleiner dan of gelijk is aan 4, of
- als het tijdstip 10:20:00 is bereikt, of
- als de simulatie gestopt is voordat één van de eerder genoemde voorwaarden bereikt is.

Van elke simulatierun wordt bepaald op welk tijdstip en om welke reden de toestand wordt weggeschreven. Met het programma DEC_MOE_VIS kan een overzicht gegeven worden hoe vaak de toestand om een bepaalde reden is weggeschreven (de dumpreden).
De set van dumpredenen staat in een gewone tekstfile. Er mogen ook commentaarregels en lege regels voorkomen in deze file. In de file FSM_USER_FILES moet de logical DUMP_SET aangeven in welke file de set van dumpredenen staat. De parameters FIRST_RUN_ID en LAST_RUN_ID uit de file FSM_USER_PARAMETERS geven aan van welke runs de toestand moet worden weggeschreven. Meer informatie over de files FSM_USER_FILES en FSM_USER_PARAMETERS staat in bijlage B.
De uitvoer van het programma WRITE_UNITS bestaat uit een aantal tekstfiles. Voor de beginsituatie en voor elke run wordt een aparte uitvoerfile gecreëerd. De naam van de files wordt gemaakt aan de hand van de logical UNIT_DATA, het versienummer van FSM en het runnummer. In de eerste regel staat informatie over de gehele run en in de andere regels staat informatie over de eenheden. De regels met "--" aan het begin van de regel zijn commentaarregels en zijn geen informatie. Ze zijn bedoeld om de uitvoer leesbaarder te maken.
De eerste regels uitvoer van dit programma zouden bijvoorbeeld kunnen zijn:

```
-- RUN_ID  DUMP REASON  DUMP_TIME
 101        RED  40%     10:17:00
--UNIT_NAME   COLOR       UNIT_TYPE    NR_OF_VEH   POSITION
  4NATD       BLUE        TOW2         0           0 0
  4NATE       BLUE        TOW2         0           0 0
  4NATR       BLUE        TOW2         0           0 0
              BLUE        TOW2         0           0 0
              BLUE        IEPG         0           0 0
              BLUE        IEPG         1           598880 5887690
  4NMC        BLUE        IEPG         1           598870 5887610
```

Uit de eerste regel blijkt dat dit de uitvoer is van run 101, dat de toestand weg is geschreven omdat aan de voorwaarde RED 40% werd voldaan. Dit gebeurde om 10:17:00 uur.

5.2 Het programma CALCULATE_MOE

Het programma CALCULATE_MOE berekent aan de hand van de gegevens van de eenheden de waarden van een aantal MOE's. De invoer is een tekstfile. De uitvoer van WRITE_UNITS kan als invoer gebruikt worden, maar eventueel kan ook op andere wijze een dergelijke tekstfile aangemaakt worden. De gegevens van de eenheden moeten dan echter wel op dezelfde manier in een file staan als de gegevens in de uitvoerfile van WRITE_UNITS.

De eisen die het programma CALCULATE_MOE oplegt aan de invoer zijn:

- Op de eerste regel staan de RUN_ID, de naam van de DUMP_REASON en het tijdstip waarop de toestand gedumpt is.
  De RUN_ID is een identificatie van de run waaruit deze gegevens komen. Dit is een geheel getal.
  De naam van de DUMP_REASON is een naam die aangeeft om welke reden deze toestand gedumpt is. Deze naam is 10 karakters lang.
  Het tijdstip van dumpen wordt aangegeven in het formaat UREN:MINUTEN:SECONDEN.
- Op elke andere regel staat informatie over één eenheid. Deze informatie bestaat uit:
  - De naam van de eenheid (maximaal 10 karakters, geen spaties)
  - De kleur van de eenheid (BLUE of RED)
  - Het type van de eenheid (maximaal 10 karakters, geen spaties)
  - Het aantal voertuigen van deze eenheid op het tijdstip waarop de toestand is gedumpt (geheel getal)
  - De positie van de eenheid. Deze wordt aangegeven door een x-coordinaat en een y-coordinaat. Beiden zijn gehele getallen en zijn gegeven in meters. Bijvoorbeeld UTM-coördinaten kunnen worden gebruikt. De positie (0, 0) wordt gegeven bij eenheden die geen voertuigen meer hebben.

Lege regels en commentaar (aangegeven door --) worden overgeslagen bij het inlezen.

In de file FSM_USER_FILES moet aangegeven worden welke files gebruikt moeten worden als invoer en als uitvoer. De logical UNIT_DATA geeft de naam van de invoerfiles aan. Deze naam wordt automatisch aangevuld met een versienummer en een run_id. De logical MOE_DATA

geeft aan hoe de uitvoerfile heet. In de file FSM_USER_PARAMETERS geven de parameters FIRST_RUN_ID en LAST_RUN_ID aan welke invoerfiles gebruikt worden.
De MOE's die momenteel berekend worden zijn (zie hoofdstuk 3 voor een beschrijving van de MOE's):

- FORCE_BLUE
- LOSS_RED
- FORCE_RATIO
- LOSS_RATIO
- WIN_FACTOR
- OBST_DEFENCE
- De tijdsduur tussen aanvang van het scenario en moment van dumpen. Op zich is dit niet echt een MOE (een tijdsduur van 20 is niet altijd beter of slechter dan een tijdsduur van 15), maar in combinatie met de dumpreden kan dit wel een MOE zijn.

De overige MOE's worden niet berekend omdat dit niet gemakkelijk en eenduidig kan. Voor specifieke gevallen kan het programma CALCULATE_MOE aangepast worden zodanig dat deze MOE's wel uitgerekend kunnen worden.

De uitvoer van CALCULATE_MOE wordt geschreven in een tekstfile. De file krijgt de naam zoals die in de FSM_USER_FILES door de logical MOE_DATA wordt gedefinieerd. De eerste en laatste regels zien er bijvoorbeeld als volgt uit:

```
    0  -- First moe_index
    7  -- Last moe_index
  101  -- First run_id
  150  -- Last run_id

 DUMPREASON  FORCE_BLUE  LOSS_RED    FORCERATIO  LOSS_RATIO  WIN_FACTOR  OBST_DEFEN  DURATION
 2.00000E+00 5.62500E-01 5.89744E-01 1.34799E+00 1.37109E+00 1.35954E+00 7.71406E-01 1.70000E+01
 2.00000E+00 6.87500E-01 5.89744E-01 1.88718E+00 1.67578E+00 1.78148E+00 8.13607E-01 1.65000E+01
 1.00000E+00 3.75000E-01 5.38462E-01 8.61539E-01 8.12500E-01 8.37019E-01 3.63521E-01 2.30000E+01
 2.00000E+00 6.25000E-01 5.89744E-01 1.57265E+00 1.52344E+00 1.54804E+00 7.58532E-01 7.50000E+00
 2.00000E+00 5.62500E-01 5.89744E-01 1.34799E+00 1.37109E+00 1.35954E+00 6.69680E-01 1.65000E+01
 2.00000E+00 7.50000E-01 5.89744E-01 2.35897E+00 1.82813E+00 2.09355E+00 8.92701E-01 7.50000E+00
:
:
 2.00000E+00 6.87500E-01 5.89744E-01 1.88718E+00 1.67578E+00 1.78148E+00 8.96836E-01 1.60000E+01
 2.00000E+00 5.00000E-01 5.89744E-01 1.17949E+00 1.21875E+00 1.19912E+00 6.83282E-01 1.70000E+01
BLUE 30%  -- Dump-reason 1
RED  40%  -- Dump-reason 2
```

De getallen op de eerste vier regels geven het formaat aan van de matrix met MOE-waarden. In elke rij staan de resultaten van één simulatierun. Elke kolom van de matrix geeft de waarden van

één MOE. De kolommen hebben titels. Dit zijn de namen van de betreffende MOE's. In de eerste kolom van de matrix staan getallen die aangeven om welke reden de toestand is weggeschreven. Onderaan staan de titels voor de dumpredenen.

5.3 Het programma DEC_MOE_VIS

Het programma DEC_MOE_VIS is een interactief programma waarmee een aantal statistische grootheden uitgerekend kan worden en/of grafische afbeeldingen gemaakt kunnen worden van de waarden van de MOE's. Eerst zal beschreven worden hoe met het programma gewerkt moet worden. Dit zal gedaan worden aan de hand van een aantal voorbeelden. Daarna zal een beschrijving van de benodigde invoer volgen. Het programma maakt gebruik van de IMSL bibliotheek voor een aantal statistische methoden en maakt gebruik van MONGO voor het tekenen van de figuren.

5.3.1 Handleiding programma DEC_MOE_VIS

Na het opstarten verschijnen er menu's met een aantal keuzemogelijkheden. Door middel van de pijltjestoetsen omhoog en omlaag wordt de gewenste mogelijkheid gekozen. Het pijltje naar links betekent terug naar het vorige menu en met return wordt een bepaalde keuze uitgevoerd.
Na het opstarten verschijnt het volgende menu in beeld:

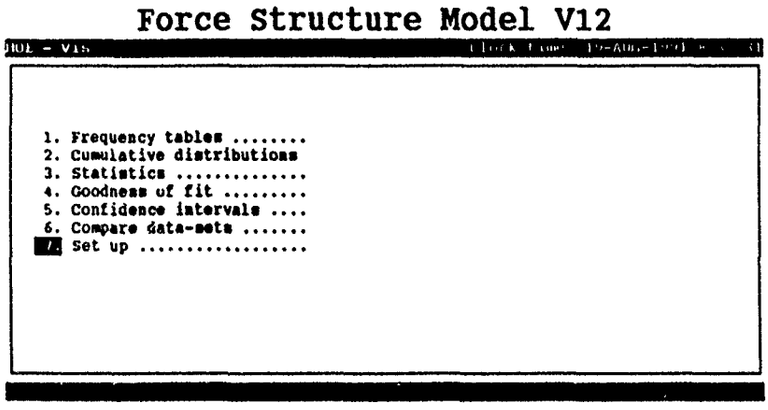

Fig. 8: Hoofdmenu

Allereerst zal voor keuzemogelijkheid 7 (Setup) gekozen moeten worden. Hiermee wordt ingesteld waar de uitvoer naar toegaat. De uitvoer is te verdelen in twee klassen:

- De grafische uitvoer. Hierbij kan gekozen worden uit een aantal mogelijke typen apparaten. Door middel van het menu kan het gewenste type uitvoerapparaat gekozen worden. Daarna zal het device aangegeven moeten worden. Dit kan een terminal zijn of een filenaam (afhankelijk van het gekozen type). Kiest men voor NONE dan wordt er geen grafische uitvoer gegenereerd.
- De tekst uitvoer. Kiest men in het setup menu voor text dan moet men daarna de filenaam invoeren van de tekstfile waar de uitvoer naar toe geschreven moet worden. Bij het vragen naar de filenaam verschijnt als standaard antwoord "NONE". Laat men dit staan als filenaam, dan zal er geen tekst uitvoer gegenereerd worden.

Na het instellen van de uitvoer parameters kunnen de statistische opties gekozen worden.

5.3.2 Voorbeelden van het werken met DEC_MOE_VIS

In deze paragraaf wordt met behulp van een aantal voorbeelden een overzicht gegeven van de mogelijkheden van DEC_MOE_VIS. Als iets vet gedrukt is dan geeft dit een keuzemogelijkheid weer van het programma DEC_MOE_VIS en als iets klein gedrukt is dan is dit een voorbeeld van de uitvoer zoals die in de tekstfile komt te staan die met **Setup** en **Text** is aangegeven. De voorbeelden die in deze paragraaf worden beschreven zijn:

Voorbeeld 1: Frequentiehistogram met gegevens uit één dataset.
Voorbeeld 2: Frequentiehistogram met twee datasets.
Voorbeeld 3: Empirische verdelingsfunctie van twee datasets
Voorbeeld 4: Statistieken van één dataset
Voorbeeld 5: Vergelijken dataset met normale verdeling
Voorbeeld 6: Betrouwbaarheidsinterval voor de verwachting
Voorbeeld 7: Het vergelijken van twee normaal verdeelde datasets.
Voorbeeld 8: Het vergelijken van twee datasets met behulp van de Kolmogorov-Smirnov toets
Voorbeeld 9: Het vergelijken van twee datasets met behulp van de toets van Wilcoxon
Voorbeeld 10: Het vergelijken van twee datasets met behulp van de Bootstrap methode

Voorbeeld 1: Frequentiehistogram met gegevens uit één dataset.
De gegevens zijn gesplitst naar hun dumpreden. Op deze manier kan zichtbaar gemaakt worden of er veel verschil is tussen de verdelingen van de uitkomsten bij de diverse dumpredenen.

Frequencytable

De gebruiker wil een frequentiehistogram.

1 set with dump_reasons

De gebruiker wil een frequentiehistogram waarbij de gegevens uit één set komen. De gegevens zijn opgesplitst naar een dumpreden.

dataset 1 : $dir1:moe_data

De gegevens van de betreffende MOE komen uit deze file.

dataset with dump_reasons : $dir1:moe_data

De gegevens betreffende de indeling van de waarnemingen naar dumpreden staan in deze file (dit hoeft niet noodzakelijkerwijs dezelfde file te zijn als de file met de MOE's).

Moe choice : forceratio

De gebruiker wil een frequentieverdeling van de MOE forceratio.

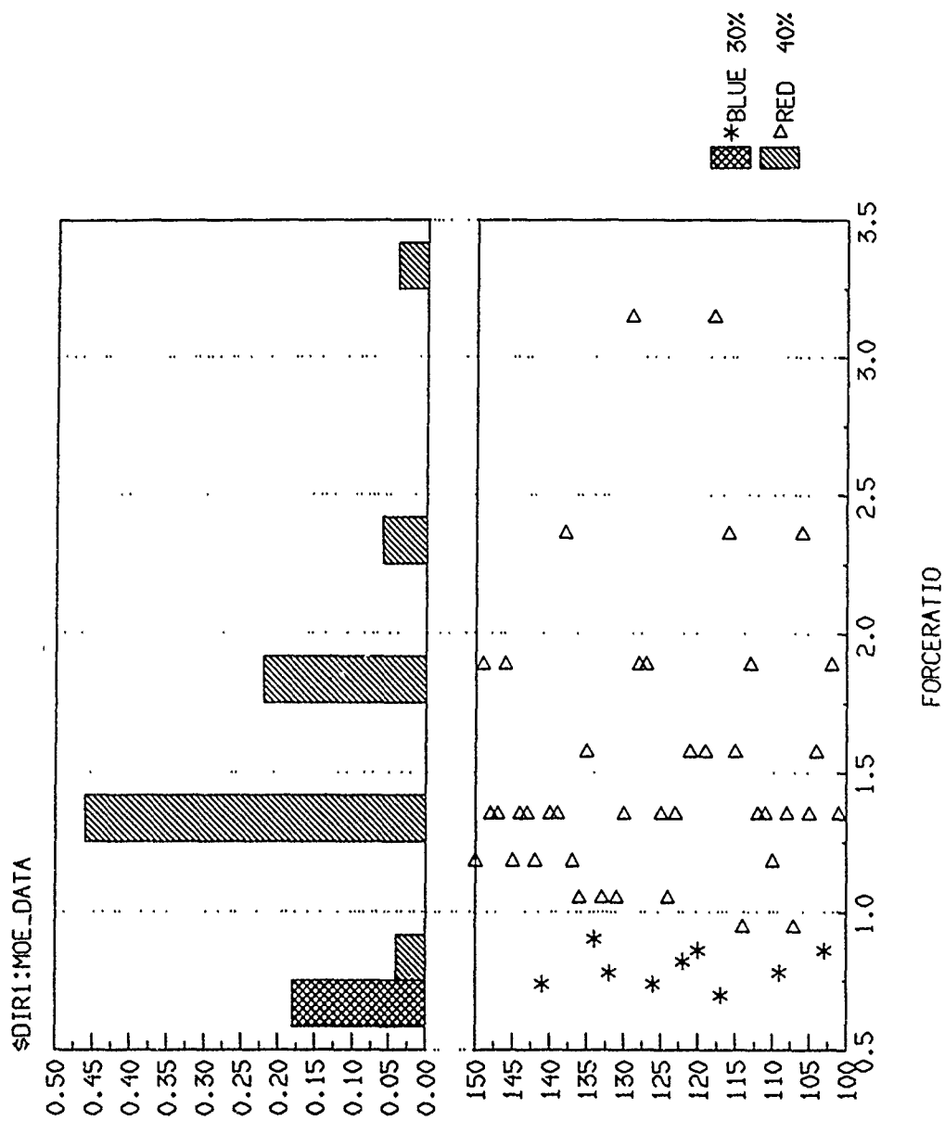

Fig. 9: Voorbeeld frequentiehistogram

De figuur bestaat uit twee delen:

1 In het onderste gedeelte staan de uitkomsten per run. Op de X-as staat de uitkomst voor de gekozen MOE en op de Y-as een index voor het nummer van de run. Een sterretje geeft aan dat de uitkomst uit een set waarnemingen komt met allen de eerste dumpreden en een driehoekje duidt op een uitkomst uit een set met de tweede dumpreden. Bijvoorbeeld de uitkomst van runnummer 101 is 1.35 en heeft als dumpreden RED 40%.

2 De bovenste helft geeft een relatieve frequentiehistogram. Deze geeft bijvoorbeeld aan dat 18% van de waarnemingen uit de set met de eerste dumpreden komt en tussen 0.5 en 1.0 ligt en dat 4% van de waarnemingen uit de set met de tweede dumpreden komt en tussen 0.5 en 1.0 ligt.

De bijbehorende uitvoer in de tekstfile ziet er als volgt uit:

```
$DIR1:MOE_DATA

FORCERATIO          Frequency (relative)
                    BLUE 30%        RED  40%
0.5 <= 1.0 :        9 ( 18.00%)     2 (  4.00%)
1.0 <= 1.5 :        0 (  0.00%)    23 ( 46.00%)
1.5 <= 2.0 :        0 (  0.00%)    11 ( 22.00%)
2.0 <= 2.5 :        0 (  0.00%)     3 (  6.00%)
2.5 <= 3.0 :        0 (  0.00%)     0 (  0.00%)
3.0 <= 3.5 :        0 (  0.00%)     2 (  4.00%)
Total               9 ( 18.00%)    41 ( 82.00%)
```

Uit de tabel valt af te lezen dat er 3 waarnemingen zijn in de set met de tweede dumpreden en met een waarde tussen 2.0 en 2.5, ofwel 6% van alle waarnemingen.
Het aantal klassen en de klassegrenzen worden door het programma bepaald aan de hand van de minimum- en maximumwaarde van de waarnemingen.

Voorbeeld 2: Frequentiehistogram met twee datasets.

Frequency table

De gebruiker wil een frequentiehistogram.

2 sets

De gebruiker wil de frequenties van twee datasets met elkaar vergelijken.

Dataset 1: $dir1:moe_data

Dit is de file met gegevens van de eerste data set.

Dataset 2: $dir2:moe_data

Dit is de file met gegevens van de tweede data set.

Moe choice : forceratio

De gebruiker wil een frequentieverdeling van de MOE forceratio.

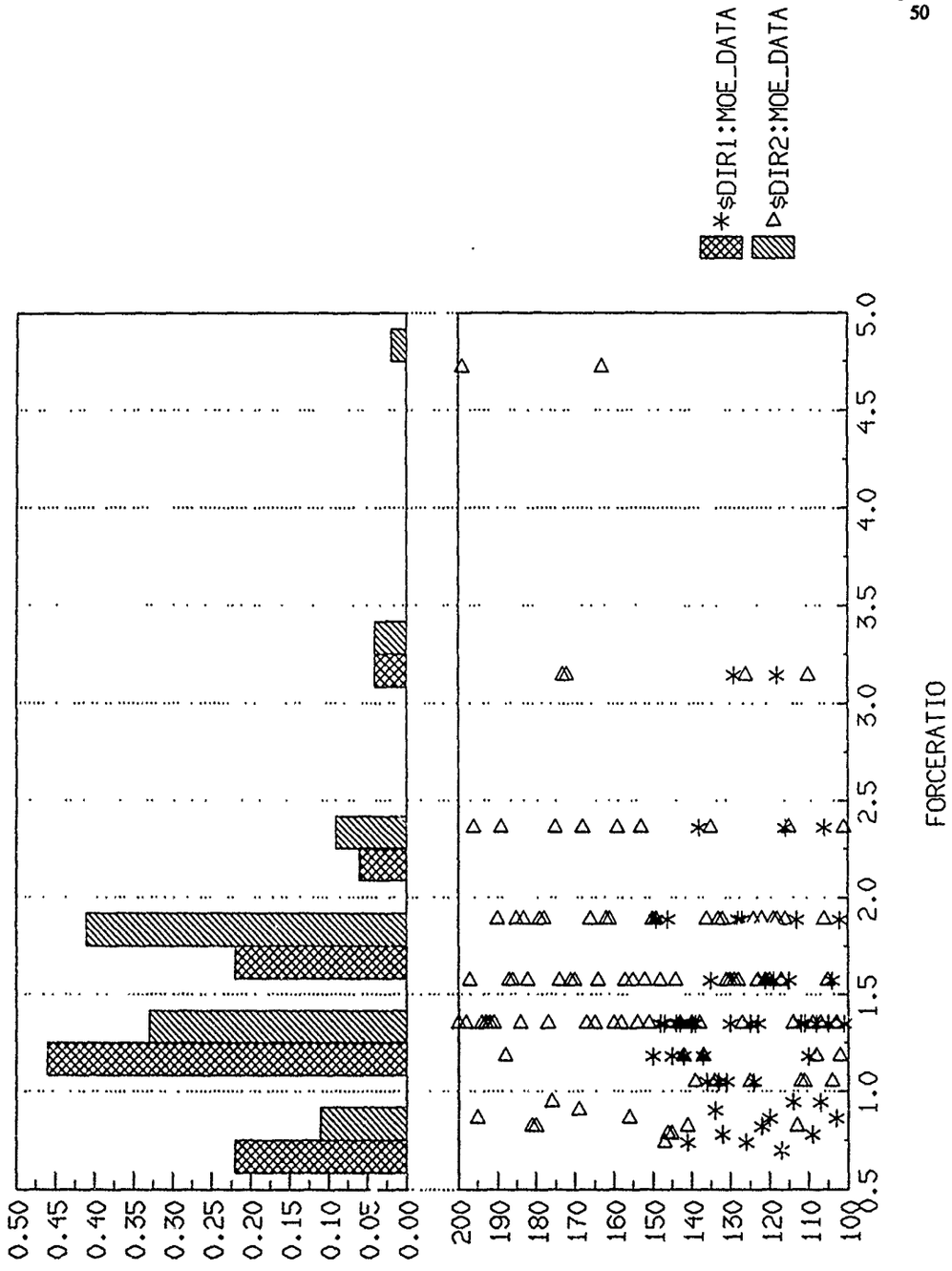

Fig. 10: Voorbeeld frequentiehistogram met twee datasets

De figuur bestaat uit twee delen:

1 In het onderste gedeelte staan de uitkomsten per run. Op de X-as staat de uitkomst voor de gekozen MOE en op de Y-as een index voor het nummer van de run. Een sterretje geeft aan dat de uitkomst uit de eerste datafile komt en een driehoekje duidt op een uitkomst uit de tweede datafile. Bijvoorbeeld de uitkomst van runnummer 101 uit de eerste dataset is 1.35 en uit de tweede dataset is 2.35.

2 De bovenste helft geeft een relatieve frequentiehistogram. Deze geeft bijvoorbeeld aan dat 46% van de waarnemingen uit de eerste dataset tussen 1.0 en 1.5 ligt en dat 33% van de waarnemingen uit de tweede dataset tussen 1.0 en 1.5 ligt.

De bijbehorende uitvoer in de tekstfile ziet er als volgt uit:

```
FORCERATIO          Frequency (relative)
                    Set 1          Set 2
0.5 <= 1.0 :    11 ( 22.00%)      11 ( 11.00%)
1.0 <= 1.5 :    23 ( 46.00%)      33 ( 33.00%)
1.5 <= 2.0 :    11 ( 22.00%)      41 ( 41.00%)
2.0 <= 2.5 :     3 (  6.00%)       9 (  9.00%)
2.5 <= 3.0 :     0 (  0.00%)       0 (  0.00%)
3.0 <= 3.5 :     2 (  4.00%)       4 (  4.00%)
3.5 <= 4.0 :     0 (  0.00%)       0 (  0.00%)
4.0 <= 4.5 :     0 (  0.00%)       0 (  0.00%)
4.5 <= 5.0 :     0 (  0.00%)       2 (  2.00%)

Set 1 = $DIR1:MOE_DATA
Set 2 = $DIR2:MOE_DATA
```

Uit bovenstaande tabel blijkt dat 23 waarnemingen uit de eerste set (46% van de waarnemingen uit de eerste set) een waarde hebben tussen 1.0 en 1.5.

Voorbeeld 3: Empirische verdelingsfunctie; van twee datasets

Cumulative distribution

2 sets

De gebruiker wil de verdelingsfuncties van twee datasets met elkaar vergelijken.

Dataset 1: $dir1:moe_data

Dit is de file met gegevens van de eerste data set.

Dataset 2: $dir2:moe_data

Dit is de file met gegevens van de tweede data set.

Moe choice : forceratio

De gebruiker wil de verdelingsfuncties van de MOE forceratio.

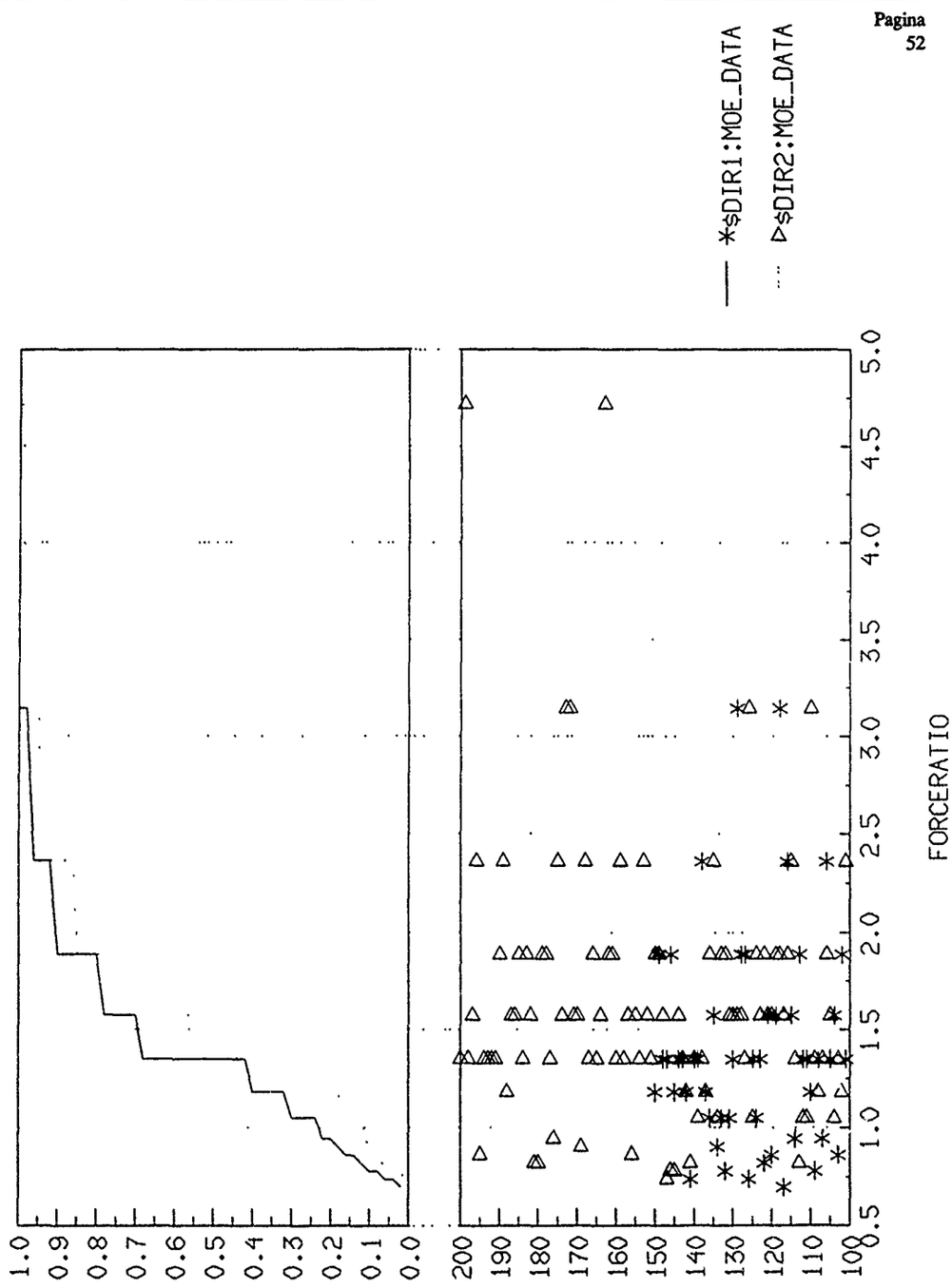

Fig. 11: Voorbeeld empirische verdelingsfunctie met twee datasets

De figuur bestaat uit twee delen:

1 In het onderste gedeelte staan de uitkomsten per run. Op de X-as staat de uitkomst voor de gekozen MOE en op de Y-as een index voor het nummer van de run.

2 De bovenste helft geeft de verdelingsfuncties. Deze geeft bijvoorbeeld aan dat 10% van de waarnemingen uit de eerste dataset kleiner of gelijk is aan 0.7 en dat 10% van de waarneming.n uit de tweede dataset kleiner of gelijk is aan 0.8.

Voorbeeld 4: Statistieken van één dataset

Statistics

De gebruiker wil statistieken van een dataset.

Dataset 1: $dir1:moe_data

Dit is de naam van de file met de gegevens

Dataset with dump_reasons : $dir1:moe_data

De gegevens betreffende de indeling van de waarnemingen naar dumpreden staan in deze file (dit hoeft niet noodzakelijkerwijs dezelfde file te zijn als de file met de MOE's).

Moe choice: forceratio

De gebruiker wil de statistieken van de verdeling van de MOE forceratio berekenen.

De uitvoer in de tekstfile ziet er als volgt uit:

```
FORCERATIO
All observations
---------------
Mean               :     1.4117
Variance           :     0.3043
Standard deviation :     0.5516
Skewness           :     1.3536
Kurtosis           :     2.0225
Minimum            :     0.6974
Maximum            :     3.1453
Nr. observations   :   50.0

BLUE 30%
---------------
Mean               :     0.7977
Variance           :     0.0047
Standard deviation :     0.0684
Skewness           :     0.1025
Kurtosis           :    -1.1967
Minimum            :     0.6974
Maximum            :     0.9026
Nr. observations   :    9.0
```

```
RED  40%
---------------
Mean               :     1.5464
Variance           :     0.2684
Standard deviation :     0.5181
Skewness           :     1.5893
Kurtosis           :     2.3636
Minimum            :     0.9436
Maximum            :     3.1453
Nr. observations   :    41.0
```

De eerste tabel geeft de statistieken van alle waarnemingen en de volgende twee tabellen geven de statistieken van de waarnemingen gesplitst naar de dumpreden. Skewness is de Engelse benaming voor scheefheid.

Voorbeeld 5: Vergelijken dataset met normale verdeling

Goodness of fit

De gebruiker wil verdeling vergelijken met normale verdeling.

Chi-square

De gebruiker wil een goodness of fit test doen met behulp van de chi-kwadraat toets.

Dataset 1: $dir1:moe_data

Dit is de naam van de file met de gegevens

Level of confidence: 95.0

Het betrouwbaarheidsniveau 1 - α is 95%

Moe choice: forceratio

De gebruiker wil de verdeling van de MOE forceratio vergelijken met een normale verdeling.

De uitvoer in de tekstfile ziet er als volgt uit:

```
Chi-square ..............
Dataset 1
$DIR1:MOE_DATA
Moe Choice
FORCERATIO
Level of confidence   :    95.0000
Chi-squared statistic :    17.1998
P-value               :     0.0006
DF                    :     3.0000
The hypothesis will be rejected
```

De P-value geeft aan hoe groot de kans is dat, als de nulhypothese waar is, er een toetsingsgrootheid gevonden wordt die groter of gelijk is aan de uitgerekende chi-kwadraat toetsingsgrootheid. Als deze P-value kleiner is dan α, dan wordt de nulhypothese verworpen. DF is het aantal vrijheidsgraden.

Voorbeeld 6: Betrouwbaarheidsinterval voor de verwachting

Confidence intervals

De gebruiker wil een betrouwbaarheidsinterval voor de verwachting berekenen.

Dataset 1: $dir1:moe_data

Dit is de naam van de gewenste dataset

Level of confidence: 95.0

Het betrouwbaarheidsniveau 1 - α is 95%

Moe choice: forceratio

De gebruiker wil het betrouwbaarheidsinterval van de MOE forceratio berekenen.

De uitvoer in de tekstfile ziet er als volgt uit:

```
Confidence intervals
Dataset 1
$DIR1:MOE_DATA
Moe Choice
FORCERATIO
Confidence intervals for the population mean value
Level of confidence             :     95.0000
Normal method                   :      1.2549 -      1.5684
Bootstrap standard method       :      1.2649 -      1.5585
Bootstrap percentile method     :      1.2727 -      1.5633
Bias-corrected percentile method:      1.2815 -      1.5701
```

Wanneer het betrouwbaarheidsinterval volgens de normale methode wordt berekend, krijgt men als antwoord dat de verwachting met een waarschijnlijkheid van 95% ligt tussen 1.2549 en 1.5684. Dit is echter met de aanname dat de MOE normaal verdeeld is. De andere betrouwbaarheidsintervallen zijn met behulp van de bootstrap methode verkregen.

Voorbeeld 7: Het vergelijken van twee normaal verdeelde datasets.

Compare datasets

De gebruiker wil datasets vergelijken.

Student's t

De gebruiker wil de Student's t-toets of daarmee samenhangende toetsen gebruiken.

Dataset 1: $dir1:moe_data

Dit is de naam van de eerste dataset

Dataset 2: $dir2:moe_data

Dit is de naam van de tweede dataset

Level of confidence: 95.0

Het betrouwbaarheidsniveau 1 - α is 95%

Moe choice: forceratio

De gebruiker wil de verwachtingen van de MOE forceratio vergelijken.

De uitvoer in de tekstfile ziet er als volgt uit:

```
Compare datasets
Student's t .............
Dataset 1
$DIR1:MOE_DATA
Dataset 2
$DIR2:MOE_DATA
Moe Choice
FORCERATIO
Level of confidence     :    95.0000
Confidence interval for the difference in means (first minus second) :
   ASSUMING EQUAL VARIANCES :
    -0.4552 -     -0.0154
   The hypothesis (means are equal) will be rejected !

   NOT ASSUMING EQUAL VARIANCES :
    -0.4406 -     -0.0300
   The hypothesis (means are equal) will be rejected !

Hypothese: variances are equal
Level of confidence     :    95.0000
F-value                 :     1.5325
Degrees of freedom      :    49,  99
Prob. of a larger F     :     0.0992
The hypothesis will be rejected !
```

Deze procedure mag alleen gebruikt worden als aangenomen mag worden dat beide sets waarnemingen uit een normale verdeling komen. Uit dit voorbeeld blijkt dat met een kans van 95% het verschil tussen de twee verwachtingen ligt tussen -0.4552 en -0.0154 als ook aangenomen mocht worden dat de varianties aan elkaar gelijk zijn. Mag de aanname van gelijke

varianties niet gedaan worden dan is het betrouwbaarheidsinterval in dit voorbeeld -0.4406 tot -0.0300 (berekend met de Satterthwaite procedure).
Ook wordt een toetsingsgrootheid (F-value) gegeven, waarmee de hypothese getoetst kan worden dat de twee varianties aan elkaar gelijk zijn (F-toets). De kans op een grotere F moet vergeleken worden met de onbetrouwbaarheid α. Is de kans kleiner dan α, dan dient de nulhypothese verworpen te worden. Het is af te raden om deze toets te gebruiken om te bepalen welk van de twee gegeven betrouwbaarheidsintervallen gebruikt wordt. Bij twijfel over gelijkheid van de varianties moet Satterthwaite's procedure worden gebruikt.

Voorbeeld 8: Het vergelijken van twee datasets met behulp van de Kolmogorov-Smirnov toets

Compare datasets

De gebruiker wil datasets vergelijken.

Kolmogorov-Smirnov

De gebruiker wil de Kolmogorov-Smirnov toets gebruiken.

Dataset 1: $dir1:moe_data

Dit is de naam van de eerste dataset

Dataset 2: $dir2:moe_data

Dit is de naam van de tweede dataset

Level of confidence: 95.0

Het betrouwbaarheidsniveau 1 - α is 95%

Tolerated difference: 0.0

De tolerantie is 0.0

Moe choice: forceratio

De gebruiker wil de verdelingen van de MOE forceratio vergelijken.

De uitvoer in de tekstfile ziet er als volgt uit:

```
Compare datasets
Kolmogorov-Smirnov ......
Dataset 1
$DIR1:MOE_DATA
Dataset 2
$DIR2:MOE_DATA
Moe Choice
FORCERATIO
Level of confidence  :    95.0000
Tolerated difference :     0.0000
D+                   :     0.2400
D-                   :     0.0000
P-value              :     0.0289
The hypothesis will be rejected !
```

De P-value geeft aan hoe groot de kans is dat, als de nulhypothese waar is, er een toetsingsgrootheid gevonden wordt die groter of gelijk is aan de gevonden toetsingsgrootheid MAX (D+, D-). Als deze P-value kleiner is dan α, dan wordt de nulhypothese verworpen.
In dit voorbeeld wordt de nulhypothese dat de verdelingen van de MOE's gelijk zijn verworpen.

Voorbeeld 9: Het vergelijken van twee datasets met behulp van de toets van Wilcoxon

Compare datasets

De gebruiker wil datasets vergelijken.

Wilcoxon

De gebruiker wil de toets van Wilcoxon gebruiken.

Dataset 1: $dir1:moe_data

Dit is de naam van de eerste dataset

Dataset 2: $dir2:moe_data

Dit is de naam van de tweede dataset

Level of confidence: 95.0

Het betrouwbaarheidsniveau 1 - α is 95%

Moe choice: forceratio

De gebruiker wil de verwachtingen van de MOE forceratio vergelijken.

De uitvoer in de tekstfile ziet er als volgt uit:

```
Compare datasets
Wilcoxon ................
Dataset 1
$DIR1:MOE_DATA
Dataset 2
$DIR2:MOE_DATA
Moe Choice
FORCERATIO
Level of confidence     :    95.0000
Wilcoxon test statistic :  3134.5000
P-value                 :     0.0097
The hypothesis will be rejected !
```

De P-value geeft aan hoe groot de kans is dat, wanneer de nulhypothese waar is, er een toetsingsgrootheid gevonden wordt die groter of gelijk is aan de uitgerekende Wilcoxon toetsingsgrootheid. Als deze P-value kleiner is dan α, dan wordt de nulhypothese verworpen.

Voorbeeld 10: Het vergelijken van twee datasets met behulp van de Bootstrap methode

Compare datasets

De gebruiker wil datasets vergelijken.

Bootstrap

De gebruiker wil de bootstrap methode gebruiken.

Dataset 1: $dir1:moe_data

Dit is de naam van de eerste dataset

Dataset 2: $dir2:moe_data

Dit is de naam van de tweede dataset

Level of confidence: 95.0

Het betrouwbaarheidsniveau 1 - α is 95%

Moe choice: forceratio

De gebruiker wil de verwachtingen van de MOE forceratio vergelijken.

De uitvoer in de tekstfile ziet er als volgt uit:

```
Compare datasets
Bootstrap ..............
Dataset 1
$DIR1:MOE_DATA
Dataset 2
$DIR2:MOE_DATA
Moe Choice
FORCERATIO
Level of confidence :     95.0000
Confidence interval for the difference in means (first minus second) :
    -0.4313 -     -0.0364
The hypothesis (means are equal) will be rejected !
```

Uit dit voorbeeld blijkt dat met een kans van 95% het verschil tussen de twee verwachtingen ligt tussen -0.4313 en -0.0364. Omdat 0.0 niet in dit interval ligt is wordt de nulhypothese (beide verwachtingen zijn aan elkaar gelijk) verworpen.

5.3.3 De benodigde invoer

De invoer van het programma DEC_MOE_VIS bestaat uit een matrix met daarin de waarden voor de MOE's. Op elke regel (rij) staan de uitkomsten van één run. In elke kolom staan de uitkomsten van één MOE. In de file met deze matrix worden eerst enkele getallen verwacht die de grootte van de matrix weergeven. Verder moeten de namen van de MOE's in de file staan voor de matrix. Een voorbeeld van de eerste en laatste regels van een invoerfile ziet er als volgt uit:

```
    0  -- First moe_index
    7  -- Last moe_index
  101  -- First run_id
  150  -- Last run_id

 DUMPREASON  FORCE_BLUE  LOSS_RED    FORCERATIO  LOSS_RATIO  WIN_FACTOR  OBST_DEFEN  DURATION
 2.00000E+00 5.62500E-01 5.89744E-01 1.34799E+00 1.37109E+00 1.35954E+00 7.71406E-01 1.70000E+01
 2.00000E+00 6.87500E-01 5.89744E-01 1.88718E+00 1.67578E+00 1.78148E+00 8.13607E-01 1.65000E+01
 1.00000E+00 3.75000E-01 5.38462E-01 8.61539E-01 8.12500E-01 8.37019E-01 3.63521E-01 2.30000E+01
 2.00000E+00 6.25000E-01 5.89744E-01 1.57265E+00 1.52344E+00 1.54804E+00 7.58532E-01 7.50000E+00
 2.00000E+00 5.62500E-01 5.89744E-01 1.34799E+00 1.37109E+00 1.35954E+00 6.69680E-01 1.65000E+01
 2.00000E+00 7.50000E-01 5.89744E-01 2.35897E+00 1.82813E+00 2.09355E+00 8.92701E-01 7.50000E+00
:
:
 2.00000E+00 6.87500E-01 5.89744E-01 1.88718E+00 1.67578E+00 1.78148E+00 8.96836E-01 1.60000E+01
 2.00000E+00 5.00000E-01 5.89744E-01 1.17949E+00 1.21875E+00 1.19912E+00 6.83282E-01 1.70000E+01
BLUE 30%  -- Dump-reason 1
RED  40%  -- Dump-reason 2
```

De eerste vier regels geven de grootte van deze matrix weer. Er zijn 8 kolommen in de matrix (0 tot en met 7) en 50 rijen (101 tot en met 150). In kolom 0 staat een getal dat aangeeft om welke dumpreden de toestand is weggeschreven. Als er verschillende dumpredenen gebruikt zijn moet dit altijd via kolom nul aangegeven worden. De nummers van de rijen (101 tot en met 150) horen bij de run-id's van de verschillende runs.
De vijfde regel geeft de namen van de kolommen, dus van de MOE's. Een naam bestaat uit maximaal 10 karakters en mag geen spaties bevatten.
Na de matrix met reële getallen kunnen enkele regels volgen met op elke regel een naam van een dumpreden.
De uitvoer van het programma CALCULATE_MOE kan als invoer dienen voor dit programma.

In FSM_USER_FILES moet aangegeven worden welke files als invoer gebruikt moeten worden. Dit wordt gedaan door de logicals MOE_DATA_1, MOE_DATA_2 enz. In bijlage B staat meer over het gebruik van FSM_USER_FILES.

6 MOGELIJKHEDEN VOOR VERDER ONDERZOEK EN AANBEVELINGEN

Nu er de mogelijkheden zijn om op eenvoudige wijze maten van effectiviteit (MOE's) uit te rekenen en te analyseren, zijn er ook uitgebreidere statistische onderzoeken mogelijk.
De mogelijkheden tot verder onderzoek die in dit hoofdstuk behandeld worden zijn:

- Gevoeligheidsanalyse
- Metamodellen

Gevoeligheidsanalyse

Gevoeligheidsanalyses worden uitgevoerd om te bepalen of het model gevoelig is voor kleine veranderingen in de invoer. Stel voor een invoerparameter een interval vast waartussen deze parameter kan schommelen. Neem nu de twee uitersten en ga daarmee scenario's doorrekenen. De andere invoerparameters worden constant gehouden. De gevoeligheid kan worden bepaald door de uitkomsten met elkaar te vergelijken. Zie paragraaf 4.7 (Datasets vergelijken).
Het kan nodig zijn om enkele sets van invoerparameters door te rekenen, omdat in één set de uitkomst van het model ongevoelig kan zijn voor deze invoerparameter terwijl in een andere set wel gevoeligheid optreedt. Als bij één van de sets de uitkomsten gevoelig zijn voor deze invoerparameter dan is het model gevoelig voor deze parameter.

Metamodellen

Bij metamodellering worden relaties (F) geschat tussen een set invoerparameters ($x_1, .. ,x_n$) en de verwachtingswaarde van één of meerdere maten van effectiviteit (MOE).

$$MOE = F(x_1, .. ,x_n)$$

Zo kan bijvoorbeeld een relatie geschat worden tussen de initiële gevechtskrachtverhouding en verwachtingswaarde van de terreinwinst van de vijand. Deze relaties kunnen globale inzichten verschaffen in de gevoeligheid voor veranderingen in de invoerparameters:

$$MOE_{\delta} = F(x_1, .. , x_i + \delta, .., x_n)$$

Ook kunnen de relaties gebruikt worden om globale schattingen te maken van verwachtingswaarden van MOE's met waarden van invoerparameters waarvoor geen simulaties zijn gedraaid:

$$MOE' = F(x_1', .., x_n')$$

Om de schatting van de relatie F te kunnen maken worden de waarden van de invoerparameters gevarieerd. Uit theoretisch oogpunt wordt verondersteld dat de simulatie-uitkomst afhankelijk is van deze invoerparameters.

Voor elke set invoerwaarden worden een aantal simulaties gedraaid en aan de hand van de simulatie-uitkomsten kan een schatting gemaakt worden van de verwachting van de MOE. Doordat er meerdere sets zijn met waarden van de invoerparameters en bijbehorende schattingen van de verwachting kan de relatie geschat worden tussen de invoerparameters en de verwachting.

Om dit soort analyses te kunnen doen is het nodig dat er goede inzichten zijn in de theoretische samenhang tussen invoerparameters en de gekozen MOE.

Verder is een opzet nodig voor het uitgebreid beschrijven van de invoer. Als er een goede opzet is en bij elk scenario worden tevens gegevens genoteerd over de gevonden MOE('s) dan kan na verloop van tijd onderzoek gestart worden naar verbanden tussen waarden van invoerparameters en de uitkomst van de MOE.

LITERATUUR OVERZICHT

[Alink] Drs. G.A. Alink, Mw. Drs. E.A.M. Boots, Ir. M.J. le Mahieu, Mw. H.A. Mol, Ir. H. Philippens
Verslag van de werkgroep "Validatie van simulatiemodellen"
TNO-Fysisch en Elektronisch Laboratorium, FEL-90-I261, 1990
ONGERUBRICEERD

[Efron 1] B. Efron
Bootstrap methods: another look at the jackknife
The annals of statistics, Vol.7, No.1, 1979 (1-26)

[Efron 2] B. Efron, R. Tibshirani
Bootstrap Methods for Standa :d Errors, Confidence Intervals, and Other Measures of Statistical Accuracy
Statistical Science, Vol.1, No.1, 1986 (54-77)

[Efron 3] B. Efron
Better Bootstrap Confidence Intervals
Journal of the American Statistical Association, Vol.82, No.397, 1987 (171-185)

[Huber] R.K. Huber
System Analysis and Modeling in Defense, Development, trends, and Issues
Plenum Press, New York and London, 1984

[IMSL] IMSL
User Guide STAT/Library, Fortran subroutines for statistical analysis
IMSL, 1987

[Joppe] Drs. W. Joppe, Drs. R. Kiel
Onderzoek naar de gevoeligheid van de parameter wapensys:eembetrouwbaarheid op de resultaten van FSM simulaties.
TNO-Fysisch en Elektronisch Laboratorium, FEL-91-A177, 1991,
ONGERUBRICEERD

[Kendall] M.G. Kendall, A. Stuart
The advanced theory of statistics, Vol. 2
Charles Griffen & company Ltd. London, 1961

[Kiel] Drs. R. Kiel
MONGO - An Interactive Plotting Program
TNO-Fysisch en Elektronisch Laboratorium, FEL-90-I344, 1991,
ONGERUBRICEERD

[Merr] Ir. S.A. van Merriënboer
Vliegbasismodel SUSTAINED: Onderzoek naar betrouwbaarheidsintervallen
TNO-Fysisch en Elektronisch Laboratorium, FEL-91-I290, 1991,
ONGERUBRICEERD

[Mood] A.M. Mood, F.A. Graybill, D.C. Boes
Introduction to the theory of statistics
McGraw-Hill, 1974

[Mural] K. Muralidhar, G.A. Ames, R. Sarathy
Bootstrap confidence intervals for estimating audit value from skewed populations and small samples
Simulation Vol.56, No.2, 1991 (119-127)

[Neave] H.R. Neave
Statistics Tables
George Allen & Unwin, 1978

[Soest] Ir. J. van Soest
Elementaire Statistiek
Delftse uitgevers maatschappij, 1983

[VS 2-7200] VS 2-7200, 3e druk (dienstgeheim)
Militair woordenboek Koninklijke Landmacht

Drs. P.A.B. van Schagen
(groepsleider)

Ir. M. van der Kaaij
(auteur)

Drs. J.K. Vink
(auteur)

INDEX

Betrouwbaarheidsinterval	9, 28, 29, 30, 32, 33, 37, 38, 46, 55, 57
Betrouwbaarheidsniveau	8, 54, 55, 56, 57, 58, 59
Bootstrap	9, 29, 30, 33, 37, 38, 46, 55, 59
Chi-kwadraat	9, 25, 27, 54, 55
F-toets	31, 33, 57
Frequentiehistogram	8, 19, 20, 23, 30, 46, 47, 48, 49, 50, 51
Frequentietabel	24
Gevoeligheidsanalyse	61
Goodness of fit	9, 19, 25, 54
Kolmogorov-Smirnov	8, 9, 25, 33, 34, 46, 57
Nulhypothese	8, 25, 26, 27, 28, 31, 32, 33, 34, 35, 36, 37, 55, 57, 58, 59
Onbetrouwbaarheid	26, 27, 34, 37, 57
Satterthwaite	9, 31, 32, 57
Statistieken	9, 22, 24, 46, 53, 54
Student's t-toets	31, 56
Toets van Wilcoxon	9, 33, 36, 46, 58
Toetsingsgrootheid	26, 27, 31, 32, 33, 34, 35, 36, 37, 55, 57, 58
Tolerantie	34, 35, 57
Verdelingsfunctie	8, 19, 21, 25, 26, 27, 30, 34, 35, 46, 51, 52, 53
Vrijheidsgraden	27, 29, 32, 33, 55

HET GEBRUIK VAN FSM_USER_FILES EN FSM_USER_PARAMETERS

In deze bijlage wordt beschreven hoe de files FSM_USER_FILES en FSM_USER_PARAMETERS eruit zien. Deze files zijn om parameters en filenamen in te stellen die afhankelijk zijn, of kunnen zijn van de omgeving. Bijvoorbeeld de namen van invoer- en uitvoerfiles of de indices van de runs die gedraaid worden. Het gebruik van deze files impliceert dat ook een aantal andere files aanwezig moeten zijn. Hieronder staan alle benodigde files weergegeven en is aangegeven wat er minimaal in moet staan.

De file FSM_SYSTEM_FILES.

In deze file staan verwijzingen naar de andere files. De file kan er als volgt uitzien:

FSM_SYSTEM_PARS	FILE	[]FSM_SYSTEM_PARAMETERS
FSM_USER_PARS	FILE	[]FSM_USER_PARAMETERS
ERROR_FILE	FILE	[]MOE_ERRORS
EML_FILE	FILE	[]ERR$LOG
LOG_TERM	TERMINAL	nl:

EML_FILE geeft de logical voor de logfile en LOG_TERM voor de log_terminal. In deze file en op deze terminal kunnen door de programma's meldingen en waarschuwingen worden weggeschreven die het mogelijk maken in de gaten te houden wat het programma doet. De programma's die in dit rapport beschreven worden maken nauwelijks gebruik van deze mogelijkheid. Het is dan ook aan te raden om de logical LOG_TERM te definiëren als nl:, dit betekent dat er geen log_terminal wordt geopend.

De file FSM_SYSTEM_PARAMETERS.

In deze files staan parameters die in alle programma's noodzakelijk zijn. De file kan er als volgt uit zien:

RUN_NAME	STRING_TYPE	MOE
START_MODE	START_MODES	SCRATCH
FSP_LOG_LOGFILE	BOOLEAN_TYPE	FALSE
FSP_LOG_TERMINAL	BOOLEAN_TYPE	FALSE
FSP_AC_LOG_LOGFILE	BOOLEAN_TYPE	FALSE
FSP_AC_LOG_TERMINAL	BOOLEAN_TYPE	FALSE

De laatste vier parameters geven aan of er meldingen over het gebruik van de parameters weggeschreven worden in de logfile of op de log_terminal.

De file PARAMETER_DATABASE.

In deze file staat een overzicht van de aangeroepen parameters en files door de verschillende programma's.

De file MOE_ERRORS.DIRECT

In deze file staan de foutmeldingen die in het programma DEC_MOE_VIS gegenereerd kunnen worden.

De file FSM_USER_FILES.

Deze files geven aan welke files als invoer en uitvoer voor de verschillende programma's worden gebruikt. Onderstaand voorbeeld geeft aan hoe deze file eruit zou kunnen zien. Achter elke logical staat aangegeven welk programma deze logical nodig heeft. Als een ander programma gedraaid wordt kan deze logical weggelaten worden.

```
DBWI_FILE       FILE    $DATDIR:XXXX_                  -- WRITE_UNITS
UDT_FILE        FILE    []USER_DEFINED_TYPES           -- WRITE_UNITS
PDT_FILE        FILE    $DATDIR:PRE_DEFINED_TYPES      -- WRITE_UNITS
ANALYSIS_FILE   FILE    []ANALYSIS                     -- WRITE_UNITS
MUF_FILE        FILE    []UNITS                        -- WRITE_UNITS
RIF_FILE        FILE    []RIF                          -- WRITE_UNITS
DUMP_CRITERIA   FILE    []DUMP.TXT                     -- WRITE_UNITS

UNIT_DATA       FILE    []UNIT_DATA                    -- WRITE_UNITS
                                                         & CALCULATE_MOE

GRID_TERRAIN    FILE    $DATDIR:GRID_TERRAIN           -- CALCULATE_MOE
MOE_DATA        FILE    []MOE_DATA                     -- CALCULATE_MOE

MOE_DATA_1      FILE    $DIR1:MOE_DATA                 -- DEC_MOE_VIS
MOE_DATA_2      FILE    $DIR2:MOE_DATA                 -- DEC_MOE_VIS
```

$DATDIR, $DIR1 en $DIR2 zijn in dit voorbeeld logicals die in de VMS omgeving zijn gedefinieerd. Het grid-terrein wordt alleen gebruikt om de MOE OBST_DEFEN (hindernis verdedigingskracht) uit te rekenen.

De file FSM_USER_PARAMETERS

In deze file staan enkele parameters die in enkele van de programma's gebruikt worden. In onderstaand voorbeeld wordt aangegeven in welk programma de parameters gebruikt worden.

```
RUN_ID          INTEGER_TYPE 100                -- ALLE
FIRST_ID        INTEGER_TYPE 101                -- WRITE_UNITS
                                                   & CALCULATE_MOE
LAST_ID         INTEGER_TYPE 150                -- WRITE_UNITS
                                                   & CALCULATE_MOE

START_TIME      TIME_TYPE    08 08 1991 10 00 00 -- CALCULATE_MOE
START_OBSTACLE_1 COORD_TYPE  32U-PD004877        -- CALCULATE_MOE
END_OBSTACLE_1  COORD_TYPE   32U-PD001870        -- CALCULATE_MOE
START_OBSTACLE_2 COORD_TYPE  32U-PD004880        -- CALCULATE_MOE
END_OBSTACLE_2  COORD_TYPE   32U_PD000885        -- CALCULATE_MOE
```

RUN_ID wordt gebruikt om de naam van de logfile te maken. FIRST_ID en LAST_ID bepalen van welke simulaties de units worden weggeschreven en van welke simulaties de MOE's worden berekend. START_TIME is van belang bij het berekenen van de DURATION. De start- en eindcoördinaten zijn coördinaten van de hindernissen die verdedigd moeten worden.

UNCLASSIFIED

REPORT DOCUMENTATION PAGE (MOD-NL)

1. DEFENSE REPORT NUMBER (MOD-NL)
TD91-3696

2. RECIPIENT'S ACCESSION NUMBER

3. PERFORMING ORGANIZATION REPORT NUMBER
FEL-91-A323

4. PROJECT/TASK/WORK UNIT NO.
20565

5. CONTRACT NUMBER
A89KL619

6. REPORT DATE
DECEMBER 1991

7. NUMBER OF PAGES
68 (INCL. 2 APPENDICES, EXCL. RDP & DISTRIBUTION LIST)

8. NUMBER OF REFERENCES
15

9. TYPE OF REPORT AND DATES COVERED
INTERIM REPORT

10. TITLE AND SUBTITLE
WAARDERING EN ANALYSE VAN SIMULATIERESULTATEN
(VALUATION AND ANALYSIS OF SIMULATION RESULTS)

11. AUTHOR(S)
M. VAN DER KAAIJ, J.K. VINK

12. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
TNO PHYSICS AND ELECTRONICS LABORATORY, P.O. BOX 96864, 2509 JG THE HAGUE
OUDE WAALSDORPERWEG 63, THE HAGUE, THE NETHERLANDS

13. SPONSORING/MONITORING AGENCY NAME(S)
MOD-NL

14. SUPPLEMENTARY NOTES

15. ABSTRACT (MAXIMUM 200 WORDS, 1044 POSITIONS)
FOR ANSWERING QUESTIONS ABOUT THE DEPLOYMENT OF WEAPONSYSTEMS, THE STOCHASTIC COMBAT SIMULATION MODEL FSM (FORCE STRUCTURE MODEL) IS ONE OF THE MODELS DEVELOPED. THIS REPORT DESCRIBES THE WAY TO ANSWER A QUESTION BY MEANS OF THE STOCHASTIC SIMULATION MODEL AND STATISTICAL TECHNIQUES.
THE OUTCOMES OF COMBAT SIMULATIONS HAVE TO BE VALUED BEFORE THEY CAN BE COMPARED TO EACH OTHER. THIS CAN BE DONE WITH THE HELP OF ONE OR MORE MEASURES OF EFFECTIVENESS (MOE'S). WHEN USING A STOCHASTIC MODEL, THE MODEL HAS TO BE RUNNED SEVERAL TIMES WITH THE SAME INPUT. IN THIS WAY A SET OF OBSERVATIONS IS GENERATED AND SOMETHING CAN BE SAID ABOUT THE DISTRIBUTION OF POSSIBLE OUTCOMES. IN THIS REPORT A NUMBER OF STATISTICAL TESTS ARE DESCRIBED WITH WHICH TWO SETS OF OBSERVATIONS CAN BE COMPARED TO EACH OTHER.
SOME COMPUTER PROGRAMS ARE WRITTEN FOR MAKING THE ANALYSIS OF SIMULATION RESULTS EASIER. THESE PROGRAMS ARE ALSO DESCRIBED.
AN IMPULSE IS GIVEN TO FURTHER RESEARCH.

16. DESCRIPTORS
COMPUTERIZED SIMULATION
STATISTICAL ANALYSIS

IDENTIFIERS
MEASURES OF EFFECTIVENESS
COMBAT SIMULATION

17a. SECURITY CLASSIFICATION (OF REPORT)
UNCLASSIFIED

17b. SECURITY CLASSIFICATION (OF PAGE)
UNCLASSIFIED

17c. SECURITY CLASSIFICATION (OF ABSTRACT)
UNCLASSIFIED

18. DISTRIBUTION/AVAILABILITY STATEMENT
UNLIMITED AVAILABILITY

17d. SECURITY CLASSIFICATION (OF TITLES)
UNCLASSIFIED

UNCLASSIFIED