

MEMORANDUM

VOLGNUMMER : WM 02-06

ONDERWERP : Springerigheid van modellen
als functie van de runfrequentie

SAMENSTELLER : Kees Kok

DATUM : oktober 2002

Dit is een persoonlijk memorandum. Slechts de auteur is verantwoordelijk voor de inhoud.
Indien niet meer nodig, graag retourneren aan de auteur.

Springerigheid van modellen als functie van de runfrequentie

Kees Kok

“Levendiger numerieke modellen vragen om levendiger forecasters”

Han Mellink

“Do not complain about inconsistency, use it!”

Anders Persson

Inhoudsopgave

1.	Inleiding	2
2.	Experimentele opzet	3
3.	Experimenten	6
	A. geen foutengroei	6
	B. lineaire foutengroei	9
4.	Discussie	17
	Referenties	19
	Appendix. Exponentiële foutengroei	20

1. Inleiding

Recentelijk is er enige discussie ontstaan over een vermeende toename van de "springerigheid" van het nieuwe operationele 22km Hirlam, in vergelijking tot het "oude" (55km) Hirlam en tot het ECMWF model. Vooral in situaties waarin er "extreme" weersituaties op zouden kunnen treden lijken de opeenvolgende modelverwachtingen niet erg consistent te zijn. Een recent voorbeeld is de situatie waarin winden tot windkracht 12 voorspeld werden, de run van 6 uur later "slechts" 10 Bf gaf (en op een iets andere positie), terwijl weer 6 uur later toch de eerste verwachting dichterbij de waarheid leek. Zo'n grote springerigheid wordt als een groot bezwaar gezien door de forecasters en tevens ook als een uiting dat er iets niet deugt.

Springerigheid of inconsistentie van modellen, bijvoorbeeld gedefinieerd als een groot verschil tussen de opeenvolgende runs van hetzelfde model, kan een veelheid van oorzaken hebben. Deze kunnen zowel gelegen zijn in het analyse- als in het voorspelsysteem. Bijvoorbeeld het al of niet verwerpen van waarnemingen, een verschil in kwaliteit van de randen van Hirlam, een verschil in kwaliteit van de opeenvolgende analyses (door verschil in beschikbaarheid van het aantal waarnemingen). Daarnaast kan het aan het model zelf liggen: misschien is het model "te actief", of is er wat mis met het convectieschema of numerieke formulering of parameterisaties. Op al deze mogelijke oorzaken die recent geopperd zijn zullen we niet ingaan. We willen slechts één aspect beschouwen dat in ieder geval ook een rol hoort te spelen in de discussie. Dat is de relatie die er bestaat tussen de runfrequentie van een model, i.e. het aantal maal per dag dat een model gedraaid wordt, en de springerigheid. We zullen laten zien dat wanneer een model frequenter gedraaid gaat worden noodzakelijkerwijs de springerigheid drastisch zal toenemen en ook dat de grootte van de "sprongen" toe zal nemen. We doen dit in een zeer simpele opzet waarin louter dit statistische effect gesimuleerd wordt. Alle andere aspecten zullen buiten beschouwing gelaten worden.

De kans dat er een groot verschil optreedt tussen 2 runs wordt uiteraard groter bij het vaker draaien. Je vergelijkt immers meer verwachtingen voor hetzelfde tijdstip. De grootte van zo'n verschil wordt enigszins "tegengewerkt" door het feit dat iedere volgende run op een later tijdstip een gemiddeld (!) betere skill heeft. De combinatie van deze 2 aspecten zullen we in een sterk vereenvoudigde setting bekijken als functie van de runfrequentie voor een paar kwantitatieve definities van – of maten voor – springerigheid. Deze definities voor springerigheid, de experimentele setup en de (statistische) modelaannamen komen aan de orde in sectie 2. De resultaten voor een model met geen foutengroei en voor lineaire foutengroei staan in hoofdstuk 3, terwijl die voor exponentiële groei in de Appendix gegeven worden. In de discussie (hoofdstuk 4) wordt nader ingegaan op de consequenties van enkele nuances op onze aannamen. Tevens wordt ingegaan op de vraag waarom juist bij extreme weersituaties een grotere springerigheid verwacht mag worden (ongeacht de runfrequentie). Ten slotte komt kort aan de orde waarom datzelfde geldt voor de overgang naar een hoge resolutie model dat meer voorspellende waarde heeft bij het voorspellen van extreem weer dan zijn lagere resolutie versie.

2. Experimentele opzet

a. definities "springerigheid"

Om qua gedachtegang zo dicht mogelijk bij de Hirlam opzet te blijven gaan we uit van series van 8 opeenvolgende forecasts voor steeds hetzelfde tijdstip. Deze worden genoteerd volgens hun initiële analyse als $t=-48, -42, -36, -30, -24, -18, -12, -6$. De verifiërende analyse wordt genoteerd als $t=0$. De uitleg van de experimentele opzet en van de definities van springerigheid wordt gedaan aan de hand van de laatste 24 uur; voor 48 uur is alles analoog. Alle resultaten worden getoond voor beide perioden.

Hoewel iedereen er een min of meer helder idee over heeft, is er geen eenduidige definitie van springerigheid. Hier zullen we dan ook geen poging doen om te komen tot een consensus-definitie. We zullen hier 3 definities beschouwen. Het is duidelijk dat springerigheid van modellen sterk samenhangt met het maximale verschil in de verwachting tussen opeenvolgende runs. Dit is dan ook de definitie of maat die de meeste aandacht krijgt. Precieser: bij deze definitie kijken we naar *het maximale verschil dat optreedt bij de verifiërende analyse tussen de opeenvolgende verwachtingen* startend op $-24, -18, -12$ en -6 uur en de verifiërende analyse zelf.

$$\begin{array}{rcccc} -24 & \leftarrow & \text{---} & \text{---} & > 0 \\ & & -18 & \leftarrow \text{---} & \text{---} > 0 \\ & & & -12 & \leftarrow \text{---} > 0 \\ & & & & -6 & \leftarrow \text{---} > 0 \\ & & & & & | & 0 \end{array}$$

De maximale waarde van deze 4 getallen (de 4 opeenvolgende verschillen) wordt vergeleken met de op dezelfde manier verkregen waarde als slechts de runs van -24 en -12 beschikbaar zouden zijn, en ook nog met als alleen de -24 run er zou zijn. Deze 3 worden in het vervolg kort genoteerd als de $4x/\text{dag}$, $2x/\text{dag}$ en $1x/\text{dag}$ voorspelcyclus, of ook wel als respectievelijk de 6 -uurlijkse, 12 -uurlijkse en 24 -uurlijkse voorspelcyclus. De experimenten zijn zo opgezet dat de laatste 2 forecastseries deelverzamelingen zijn van de eerste (en dus niet onafhankelijke andere forecasts). De kansverdelingen van deze maximale verschillen tussen opeenvolgende runs verschillen uiteraard. En dus ook de springerigheid. Op welke wijze deze verdelingen variëren als functie van een aantal min of meer realistische aannamen (zie b.) is onderwerp van dit pamflet.

In de uitleg hierboven zowel als in de rest van het verhaal is er voor het gemak dus van uitgegaan dat de runs om de $6, 12$ of 24 uur starten. Dat wil zeggen $4, 2$ of 1 maal per dag. Dit is voor de interpretatie van de resultaten niet cruciaal; het is alleen bedoeld om de gedachten te bepalen. Alle resultaten blijven exact hetzelfde als het een vergelijking betreft tussen runs die $4, 2$, en 1 maal per bijvoorbeeld 12 uur of per 6 uur gestart worden. Bij een uurlijkse of 3 -uurlijkse cyclus van het Hirlam spelen dus analoge fenomenen als hier beschreven zullen worden.

Naast bovenstaande definitie van springerigheid is ook gekeken naar een tweede maat. Hierbij is springerigheid gedefinieerd als het maximale verschil tussen 2 van de runs (dus niet noodzakelijkerwijs 2 opeenvolgende), ofwel als *de grootte van de range van alle forecasts* geldig voor hetzelfde tijdstip. Het moge duidelijk zijn dat de waarden van dit soort sprongen groter of gelijk is aan die volgens de eerste definitie. Ook hier zal weer gekeken worden naar verschillen tussen de 3 voorspelcycli.

Een derde maat die kort bekeken wordt is *de maximale som van 2 opeenvolgende "sprongen"* in de verwachting. Een forecaster kan soms danig in de war gebracht worden als niet alleen een bepaalde modelverwachting flink afwijkt van de voorgaande, maar als bovendien de daaropvolgende wederom een flinke sprong te zien geeft (en misschien wel weer lijkt op de eerste verwachting). Om ook een beetje gevoel te krijgen voor hoe vaak dit soort gedrag kan voorkomen en hoe dit afhankelijk is van de runfrequentie wordt dit ook meegenomen in de analyse. Omdat deze maat voor springerigheid bepaald wordt uit tenminste 3 opeenvolgende runs kunnen we alleen een voorbeeld laten zien bij een 48 uren periode.

b. modelaannamen

Om het foutengedrag van atmosfeermodellen na te kunnen bootsen in een zeer versimpelde opzet hebben we een aantal aannamen nodig. Deze aannamen vallen uiteen in 2 categorieën. De eerste hangt samen met de *initiële fout*. De initiële toestand van de atmosfeer is niet voor de volle 100% bekend. De analyse is ongetwijfeld een niet volledig "goede" beschrijving van de werkelijkheid. Er bestaat dus een kansverdeling van mogelijke initiële toestanden waarvan aangenomen mag worden dat die een min of meer bell-shape heeft (grote kans op bijna "goed", kleine kans op een relatief grote "fout"). De exacte keuze van deze verdeling beïnvloedt onze resultaten nauwelijks. We hebben om praktische redenen (begrensdheid, analytisch oplosbaar) niet voor de normale verdeling gekozen maar voor een cosinusvormige verdeling. Deze verdeling hebben we zodanig geschaald dat de halve periode van de cosinus valt tussen -.5 en +.5. Tenslotte is de hoogte zo genomen dat de integraal van de verdeling gelijk is aan 1. De functie is $(\pi/2)\cos\pi x$. Zie Fig.1:

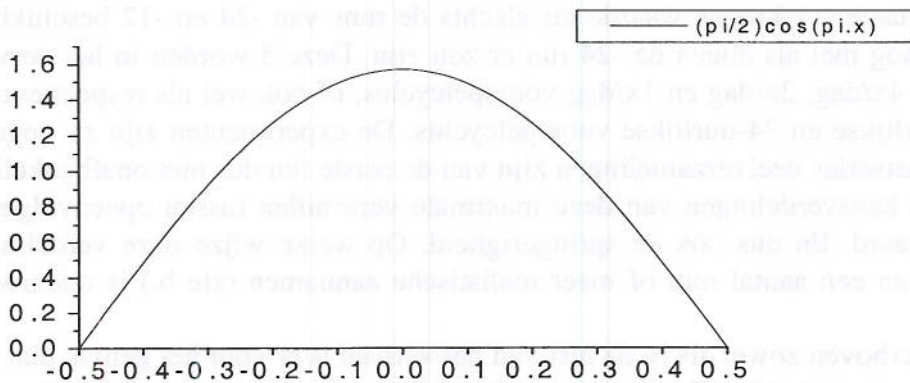


Fig.1 Kansverdeling van de analysefout.

Deze curve representeert dus de onzekerheid in de analyse (in principe inclusief de meegenomen invloed hierop van de first guess). Tenzij anders vermeld geldt deze curve voor alle analyses van -48 tot en met 0 (dus inclusief de verifiërende analyse) en voor alle experimenten. De grootte die uitgezet is op de x-as is niet zo relevant; in alle experimenten die in dit pamflet beschreven worden gaat het om de kwalitatieve aspecten. Gedacht kan worden aan de positie van een lagedrukgebied die niet 100 procent exact op de goede plaats geanalyseerd wordt. Maar ook kan gedacht worden aan een of andere scalargrootte die de toestand van de atmosfeer beschrijft, bijvoorbeeld de totale kinetische energie (TKE) over een

deel van het voorspeldomein. Het gaat niet om een kleinschalige grootheid zoals de 2-meter temperatuur op een bepaalde locatie.

De tweede aanname hangt samen met de onvoorspelbaarheid van de atmosfeer. De atmosfeer is een niet volkomen deterministisch bepaald systeem en ieder model zal dan ook *foutengroei* vertonen (gemiddeld genomen) naarmate de voorspeltermijn toeneemt. Het effect van lineaire en van exponentiële foutengroei zal nader worden bekeken. De voorgeschreven foutengroei wordt in principe voor ieder experiment constant genomen; dus niet de situatie waarin de ene run lineaire foutengroei ondergaat terwijl in de run van een tijdstip later de fouten exponentieel groeien. Afvlakking van de foutengroei wordt buiten beschouwing gelaten omdat we steeds korte voorspeltermijnen beschouwen. In werkelijkheid kan in individuele gevallen bij toenemende forecasttijd de fout soms juist kleiner worden. Dit wordt echter in geen van de voorbeelden aangenomen.

De fout in de forecast is dus de som van een analysefout, die random wordt getrokken uit de bovenstaande verdeling, en een foutengroei die lineair, exponentieel, of i.d. verondersteld wordt. Dit resulteert in een totale fout. Hierbij kan weer gedacht worden aan een positiefout van een depressie of aan een fout in de totale kinetische energie.

Door deze aannames wordt de verwachte fout bij kortere voorspeltermijn steeds kleiner. Doordat de verifiërende analyse dezelfde verdeling heeft als alle andere analyses en de foutengroei de initiële fout alleen maar vermenigvuldigt, is de gemiddelde fout bij onze modelformulering gelijk aan nul (geen bias). De (verwachte) absolute fout, maar ook de rms of de sd nemen af bij runs die dichter bij de verificatietijd starten.

Door deze aannames kunnen we de springerigheid onderzoeken als functie van alleen de runfrequentie zonder dat andere oorzaken de resultaten beïnvloeden.

3. Experimenten

Dit hoofdstuk valt uiteen in 2 delen. In deel A. kijken we onder de in het vorige hoofdstuk beschreven aannamen naar de situatie zonder foutengroei en in deel B. naar het effect van lineaire foutengroei. In de Appendix zijn de resultaten bij exponentiële foutengroei samengevat. Het effect van 1, 2 en 4 maal runnen per dag (of per 12 of 6 uur) wordt bekeken voor de in het vorige hoofdstuk beschreven maten van springerigheid. voor de forecasts vanaf 24 en vanaf 48 uur vóór de verifiërende analyse. De experimenten zijn uitgevoerd door (een groot aantal maal) random te trekken uit de voorgeschreven verdeling van de analysefout (zie Fig. 1). Onder de aanname van een bepaalde foutengroei wordt geïntegreerd tot het tijdstip van de verifiërende analyse. Dit wordt gedaan vanaf -24 , -18 , -12 en -6 met steeds nieuwe random trekkingen uit de beginverdeling. En ook analoog vanaf -48 . De verifiërende analyse wordt ook door random trekking bepaald. Om de verwachte kansverdelingen te genereren wordt zo'n cyclus van trekkingen plus integraties steeds 10000 maal herhaald.

A. geen foutengroei

Laten we eerst kijken naar puur het effect van het 1 of 2 of 4 maal runnen van het model per etmaal, zonder foutengroei. We beschouwen allereerst het grootste verschil tussen opeenvolgende runs voor hetzelfde verificatietijdstip. Voor de 4 maal per dag suite kijken we dus naar het maximum van: het verschil tussen -24 en -18 run, het verschil tussen -18 en -12 run, tussen -12 en -6 en tussen -6 en 0 . Voor de 2 andere suites analoog. De verdelingen van het maximale absolute verschil voor deze 3 voorspelcycli is te zien in Fig.2a bovenste panel. In deze figuur is, net als in de andere figuren die volgen, het percentage gevallen getoond per bin. Omdat er 3 verdelingen door elkaar getoond worden is er gekozen voor een presentatie in lijnen en niet in bars. De bins hebben in deze grafiek de breedte van $1/16$. De y-waarden (in procenten) zijn geplaatst in het midden van de bins. De dimensie van de x-waarden is de breedte van de initiële verdeling. Omdat random getrokken wordt uit de verdeling van Fig.1 die een breedte heeft van 1 is het maximale verschil tussen twee opeenvolgende trekkingen uiteraard ook gelijk aan 1.

Door het maximum te nemen bij meerdere keren trekken neemt de kans op grotere waarden uiteraard toe. Dit is in de figuur te zien door het naar rechts verschuiven van de verdelingen: bij de 24-uurs cyclus (i.e. 2 maal trekken) ligt het grootste aantal gevallen tussen 0 en $1/8$ (dashed line), bij de 12-uurs cyclus tussen $2/8$ en $3/8$ (dotted line), en bij de 6 uurs-cyclus tenslotte tussen $3/8$ en $1/2$ (getrokken lijn), terwijl bij deze laatste de laagste bin bijna volledig leeg is. Dit resultaat is slechts in geringe mate afhankelijk van de precieze vorm van de beginverdeling; iedere bell-shape curve resulteert in soortgelijk gedrag. (Zelfs bij een uniforme beginverdeling zien we dit gedrag)

Deze figuur is niks anders dan de onderbouwing van het statistische feit dat als je maar meer 2-tallen kiest uit een gegeven verdeling de kans toeneemt dat er een groot verschil optreedt (een grote "sprong") tussen de waarden van het tweetal. Dit is een van de fenomenen die meespelen bij het verhogen van de voorspelcyclus van 24 naar 6 uur. Bij geen foutengroei zal de kans op relatief grote (maar in dit voorbeeld weliswaar minieme) fouten sterk toenemen. Deze initiële fouten zullen in de atmosfeer sterk toenemen. In de atmosfeer is er uiteraard een tegenwerkende "effect" dat veroorzaakt wordt doordat er bij iedere volgende run (6, 12, 18 of 24 uur later) een gemiddeld(!) betere voorspelling gedaan kan worden. Dit doet

de mogelijkheid voor het optreden van grote fouten (en dus van grote sprongen) afnemen. Het gevolg van deze effecten tesamen zal in deel B. worden nagegaan.

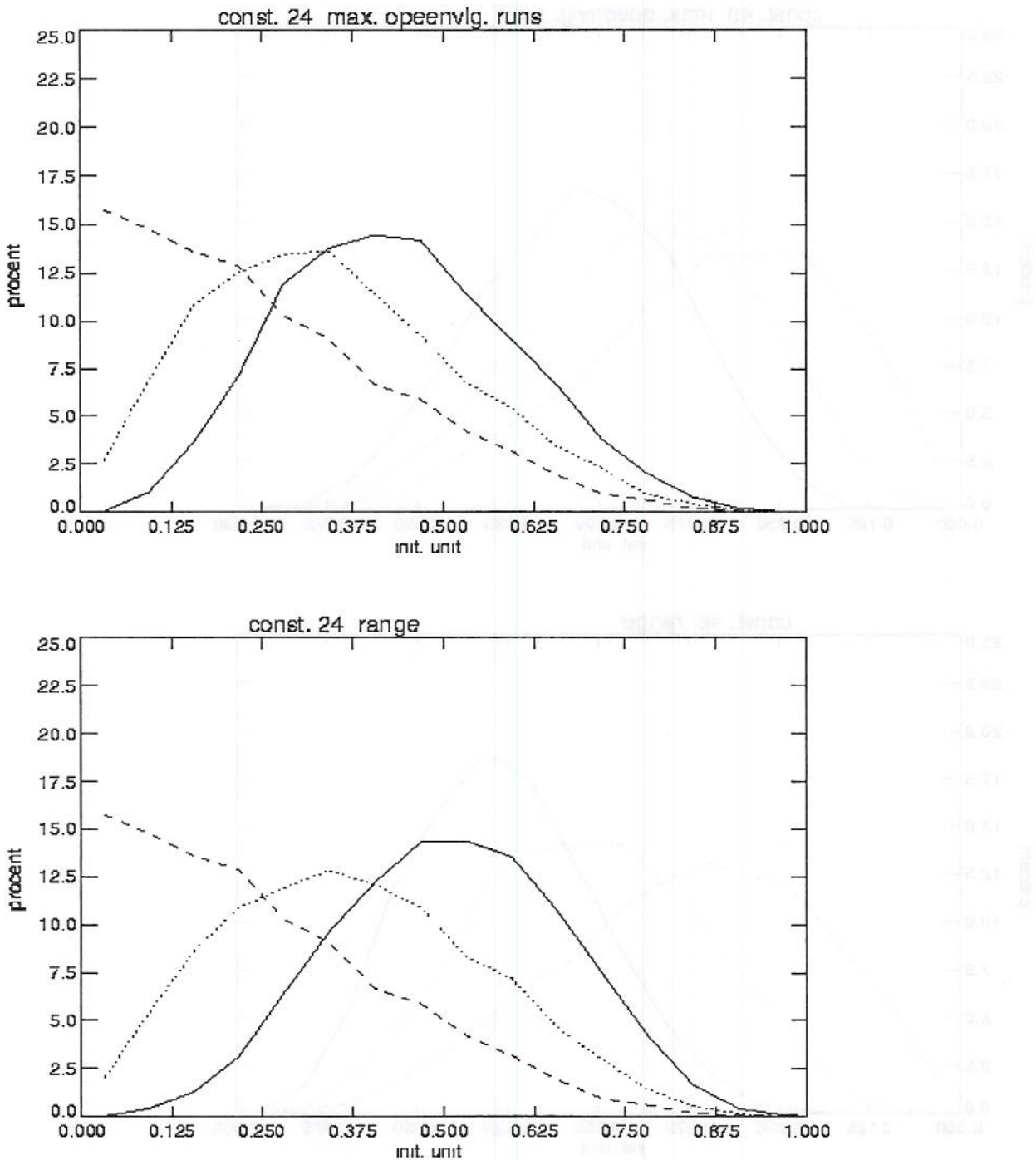


Fig. 2a. Frekwentieverdeling van het verwachte maximale verschil tussen 2 opeenvolgende runs (top panel) en van de verwachte maximale range (bottom panel) bij het 1x (dashed line), 2x (dotted line) en 4x (full line) runnen per 24 uur onder de aanname van geen foutengroei ('const').

De effecten van het vaker samplen op de tweede maat voor springerigheid – het grootste verschil tussen alle trekkingen – staat gegeven in het onderste panel. De lijn voor het 1x per dag runnen (dashed) verschilt uiteraard niet met die uit het bovenste panel. De andere

twee verdelingen zijn iets meer naar rechts verschoven als gevolg van het feit dat deze maat voor springerigheid altijd groter of gelijk is aan de eerste.

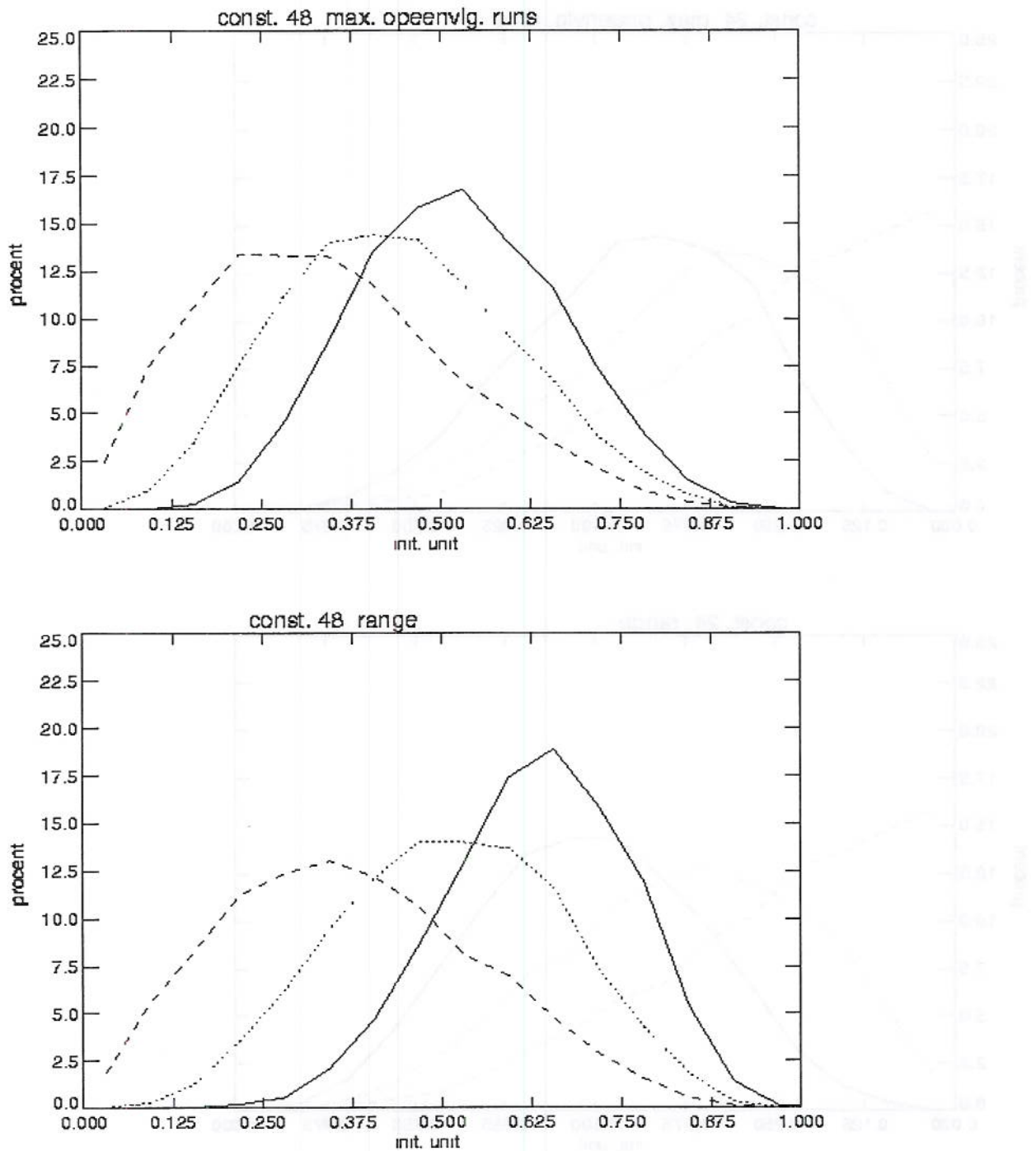


Fig. 2b. Als in Fig. 2a., maar de runs startend 48 uur voor de verifiërende analyse.

In Fig. 2b staan de analoge figuren als in Fig. 2a., maar nu voor het geval dat er gekeken wordt naar een periode van 48 uur. De kansen op relatief grote waarden zijn uiteraard veel groter dan bij de 24 uren periode.

B. Lineaire foutengroei

In het nu volgende voorbeeld wordt een lineaire foutengroei verondersteld die de analysefout lineair met de tijd laat groeien. Een schematisch voorbeeld staat gegeven in Fig.3. Hierin staat de range van de fouten gegeven voor de 5 runs vanaf $t=-24$ tot en met $t=0$ (met de 5^{de} "run" wordt bedoeld de random analyse op $t=0$). De range van de initiële fouten is weergegeven door middel van een dikke verticale streep. Deze representeert de verdeling van Fig. 1. De verdeling van de fouten op het tijdstip van de verifiërende analyse kent door de lineaire foutengroei ook eenzelfde (maar afgeplatte) vorm als die in Fig.1. In het schema staat de 4x/dag cyclus gegeven met runs om de 6 uur. Analoog kan een figuur gedacht worden van de 9 runs vanaf -48 uur. Bij het schema moet weer bedacht worden dat het tijdsinterval waarin er 4, 2 of 1 maal een run plaatsvindt er niet toe doet; alleen om de gedachten te bepalen wordt steeds een periode van een dag uur aangegeven. Het kan ook 4, 2 en 1 run per 12 of per 6 uur betreffen.

Het resultaat voor het maximale verschil tussen opeenvolgende runs is getoond in Fig. 4. De 3 curves uit Fig. 4a vertonen veel overeenkomst met die van Fig. 2. Bij de 24-uurs cyclus loopt de curve weer monotoon af, terwijl die van de 12- en vooral de 6-uurs cyclus een piek vertonen, de laatste wederom de grootste. Het statistische effect als gevolg van het meerdere malen "samplen" is nog steeds de dominante factor. De grootte van het maximale verschil is nu uiteraard een stuk groter. De x-as is gegeven in aantal malen de range van de initiële fouten, die op 1 gesteld was (Fig. 1). De bingrootte is 0.5 en de y-waarden staan weer afgebeeld boven het centrum van de bins. In het hier gegeven voorbeeld is de foutengroei zo gekozen dat de voorspelfout na 24 uur een factor 9 groter is dan de analysefout. Dit getal representeert dus de totale range van mogelijke fouten na 24 uur. De waarde van deze gekozen groei is niet echt van belang omdat we die hetzelfde nemen voor alle runs van de 3 voorspelcycli. Door een andere lineaire groei te veronderstellen veranderen de karakteristieken van de curves niet. Het enige wat gebeurt is dat de figuren naar rechts "opgerekt" worden, zonder dat de hoogtes veranderen (mits uiteraard de binbreedte mee "opgerekt" wordt). Of anders gezegd, alleen de schaal van de x-as verandert; de figuren blijven exact hetzelfde. Fig. 4 kan dus beschouwd worden als de verwachtingswaarde van de (eerste maat voor) springerigheid bij de 3 verschillende voorspelcycli, gegeven dat de (willekeurige) foutengroei hetzelfde blijft over alle runs die starten in de 24-uursperiode. Is er een totaal andere lineaire foutengroei dan blijft dus de onderlinge verhouding van de sprongen tussen de verschillende voorspelcycli hetzelfde, alleen de grootte verandert. Dit geldt analoog voor alle figuren uit de rest van dit rapport.

Een belangrijk verschil tussen Fig. 4 en Fig. 2 is de langere staart bij hogere waarden van x. (ofwel het maximum van de curve dat relatief meer naar links is verschoven). De meest extreme waarden van de getrokken lijn (de 6-uurs voorspelcyclus) zijn afkomstig van de verschillen tussen de -24 en de -18 run. In ons voorbeeld geldt dit noodzakelijkerwijs voor alle gevallen boven de 6. Gevallen onder de 6 kunnen ook het gevolg zijn van verschillen tussen 2 latere (opeenvolgende) runs. De maximaal mogelijke "sprong" is nu afhankelijk van hoe frequent - hoe korter achter elkaar - een model gerund wordt (in tegenstelling tot in Fig.2). Deze is in ons voorbeeld voor de 12-uurlijkse cyclus gelijk aan 7 en voor de 24-uurlijkse cyclus gelijk aan 5 (vergelijk Fig. 3). De kans op grotere waarden is veel groter naarmate het model vaker gerund wordt. Bijvoorbeeld, waarden boven 4 blijken ongeveer 5 maal zo vaak voor te komen in de 6-uurlijkse suite dan in die van eens per 24 uur. Dit is te zien in Fig. 4a onderste panel waarin de cumulatieve verdelingen uit het bovenste panel getoond worden. Hierbij zijn de waarden uit het bovenste panel van rechts naar links (!) opgeteld, om de

"kansen" op overschrijdingen makkelijker zichtbaar te maken. Ook voor andere waarden van de maximale "sprong" is het verschil in kans gemakkelijk af te lezen. Samengevat, de grote sprongen tussen opeenvolgende modelruns komen veel vaker voor bij toenemende runfrequentie en tevens zijn er grotere sprongen mogelijk. Let wel, we hebben het nog steeds over hetzelfde model; alleen de runfrequentie verschilt. Het toenemen van het aantal grote sprongen of het groter worden van de sprongen bij het overgaan naar een hogere runfrequentie, of bij het vergelijken van 2 modellen met verschillende runfrequentie, hoeft dus helemaal niks te maken te hebben met een verschil in skill. Bij de interpretatie van verschillen in springerigheid moet altijd rekening gehouden worden met dit effect.

Bovenstaand voorbeeld kan door de aannames van hoofdstuk 2 ook gemakkelijk analytisch berekend worden. Dit is hier achterwege gelaten.

In Fig. 4b. staan de overeenkomstige plaatjes voor de tweede definitie of maat van springerigheid: de maximale range van alle verwachtingen voor hetzelfde tijdstip. De dashed line komt uiteraard overeen met die van Fig. 4a. De maximaal mogelijke waarden zijn ook hetzelfde als die bij de eerste definitie. De verwachte waarde van de "sprong" is echter bij de tweede definitie voor de 12-uurlijkse en 6-uurlijkse cyclus aanmerkelijk hoger. Dit resulteert in een verschuiving naar rechts van de twee corresponderende curves in Fig. 4b ten opzichte van die in Fig. 4a. Hierdoor zijn bij deze definitie ook de verschillen in kansen op relatief grote sprongen tussen de 3 forecast-suites groter. Zo blijken de waarden boven 4 nu ongeveer 7 maal zo vaak voor te komen in de 6-uurlijkse suite dan in die van eens per 24 uur.

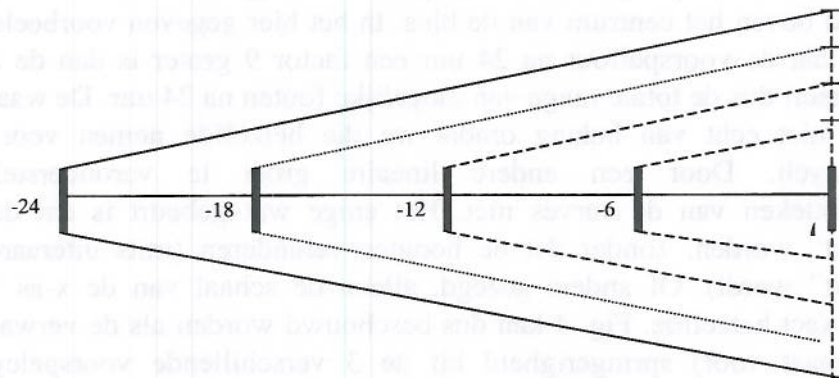


Fig.3. Schematisch overzicht van de range van de fouten als functie van de forecasttijd bij lineaire foutengroei.

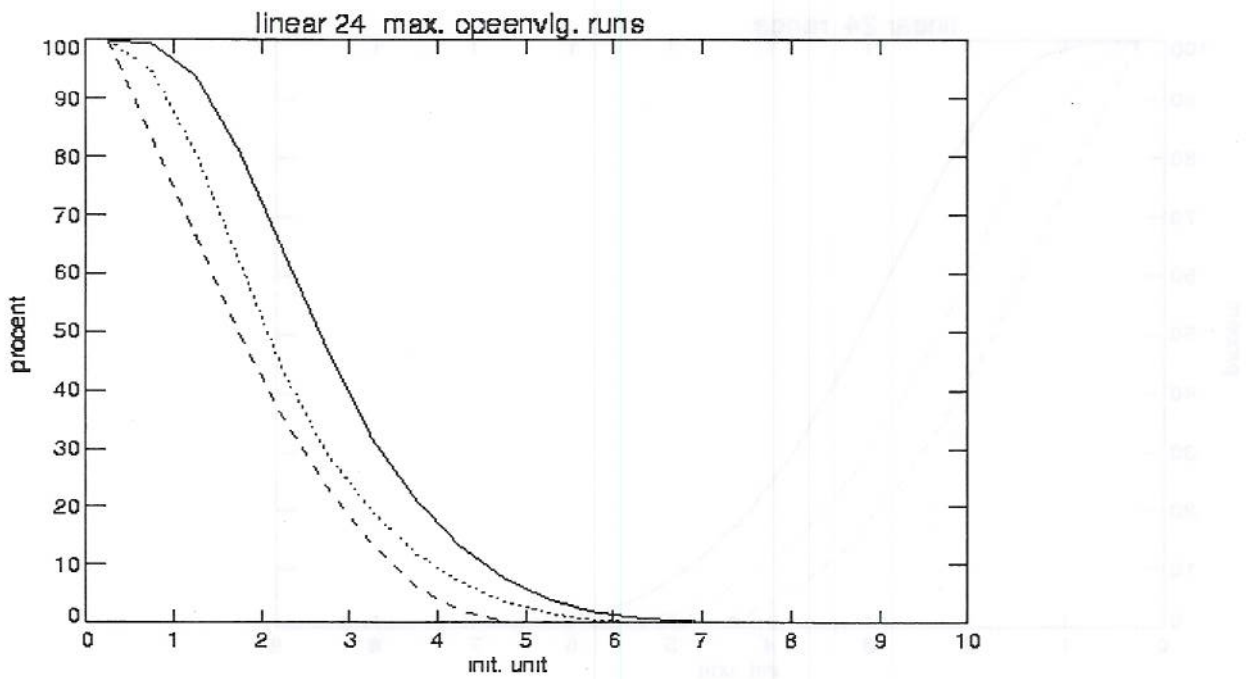
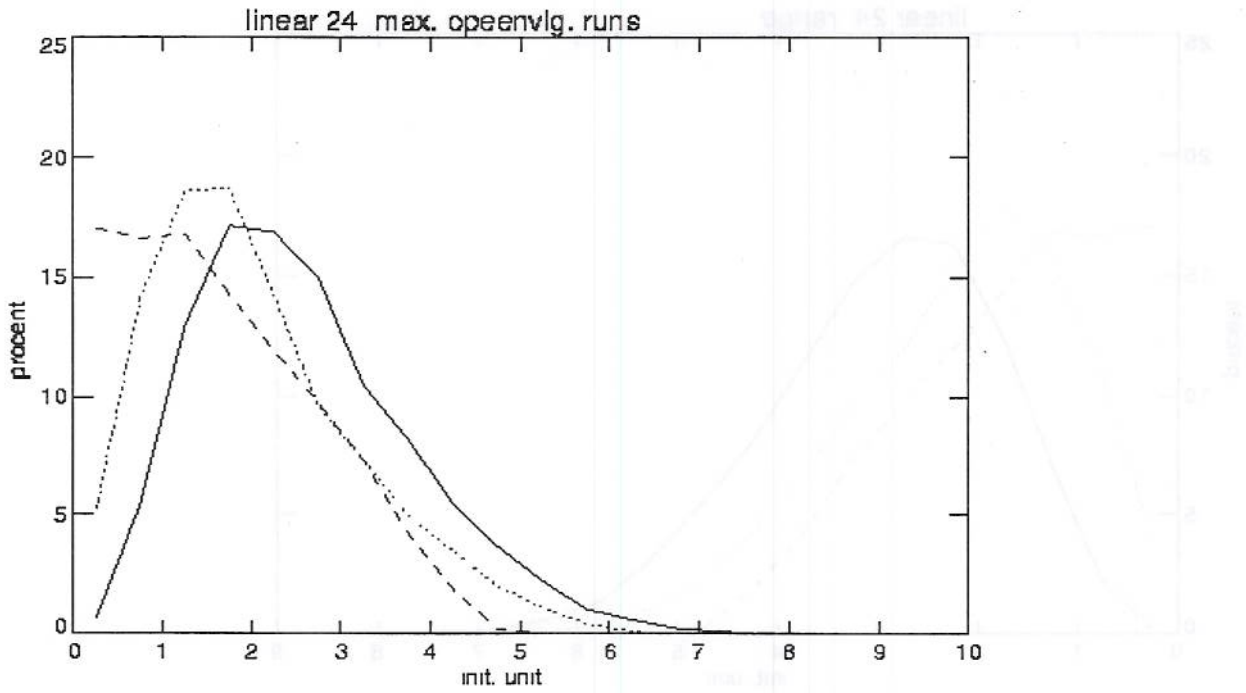


Fig. 4a. Frekwentieverdeling van het verwachte maximale verschil tussen 2 opeenvolgende runs (bovenste panel) en de cumulatieve verdeling (onderste panel) bij het 1x (dashed line), 2x (dotted line) en 4x (full line) runnen per 24 uur onder de aanname van lineaire foutengroei ('linear').

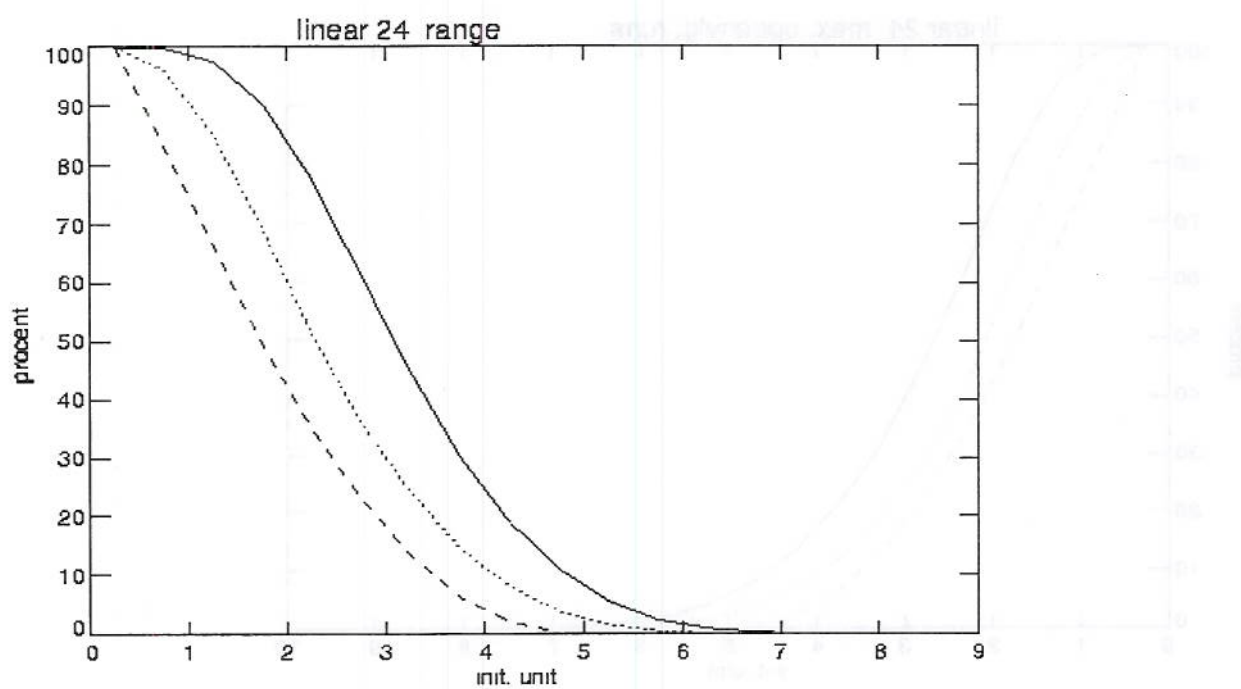
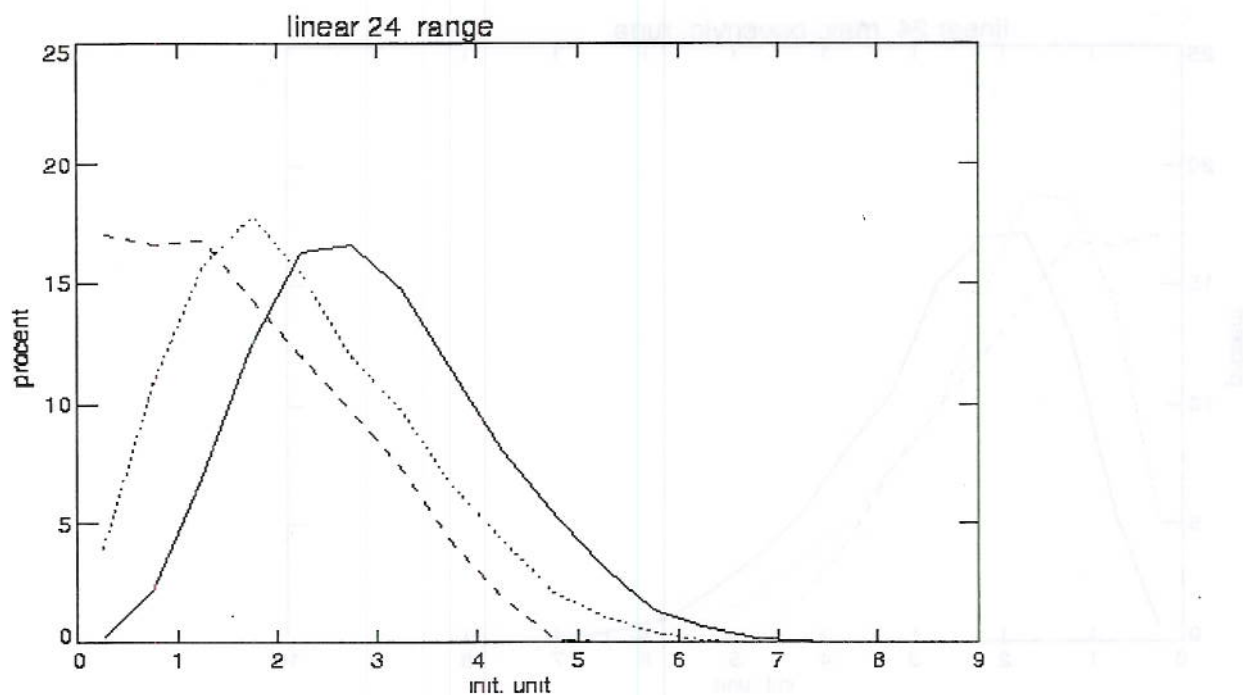


Fig. 4b. Als Fig. 4a, maar voor de maximale range van alle runs verifiërend op hetzelfde tijdstip.

Misschien ten overvloede laten we ook de resultaten zien als we de 4, 2 en 1 maal per dag suite beschouwen in een termijn van 48 uur (Fig. 5) in plaats van 24 uur. Dit is gedaan om een directe vergelijking te kunnen maken met de operationele Hirlam suite. We hebben nu de (lineaire) foutengroei de helft genomen van hierboven, i.e. wederom een factor 9, maar nu over een 48-uursperiode. Hierdoor is een goede vergelijking met de resultaten van de 24-uursperiode mogelijk. (Het gaat ons immers toch om de relatieve verschillen als gevolg van verschillen in runfrequentie). Hierdoor komen de curves voor het 1x per dag en 2x per dag runnen van het model, bekeken over een periode van 48 uur (Fig.5), nagenoeg overeen met die voor het 2x per dag en 4x per dag runnen bekeken over een periode van 24 uur (Fig. 4). (Kleine verschillen komen doordat voor de constructie van beide figuren gebruik is gemaakt van een andere set random trekkingen). De curve in Fig. 5 voor het 4x per dag runnen (getrokken lijn) is equivalent met het om de 3 uur runnen in een 24-uursperiode. Deze ligt in het onderste panel altijd (relatief ver) boven de andere curves. Het overgaan van het om de 6 uur runnen naar het om de 3 uur runnen geeft dus wederom aanleiding tot mogelijkere grotere sprongen alsmede tot een aanzienlijke toename van het aantal relatief grote sprongen. Bovenstaande geldt ook als we naar de tweede definitie van springerigheid kijken. De resultaten hiervoor staan in Fig. 5b.

Tot slot voor de volledigheid de resultaten voor de derde definitie (of maat) van springerigheid, i.e. de maximale som van 2 opeenvolgende sprongen (Fig. 5c). We onderzoeken dus de gevallen waarin de tweede run veel afwijkt van de eerste, en de derde weer veel van de tweede, bijvoorbeeld doordat het model “flip-flopt” tussen 2 mogelijke oplossingen. Dit kan uiteraard alleen onderzocht worden in de 48-uursperiode. De maximale waarden zijn hier uiteraard veel groter (bijna twee maal) dan bij de andere 2 onderzochte maten voor springerigheid. Een ander opvallend verschil zijn de langere staarten van de verdelingen die bovendien nog eens ver van elkaar liggen (zie onderste panel). Dit resulteert wederom in grote verschillen in sprongen voor de verschillende runcycli. Nemen we bijvoorbeeld een som van 2 opeenvolgende sprongen die slechts zelden, zeg in 1 à 2 procent van de gevallen, voorkomt in de 1x per dag suite, dan zien we dat zo'n sprong in circa 20 procent van de gevallen verwacht mag worden in een 4x per dag cyclus.

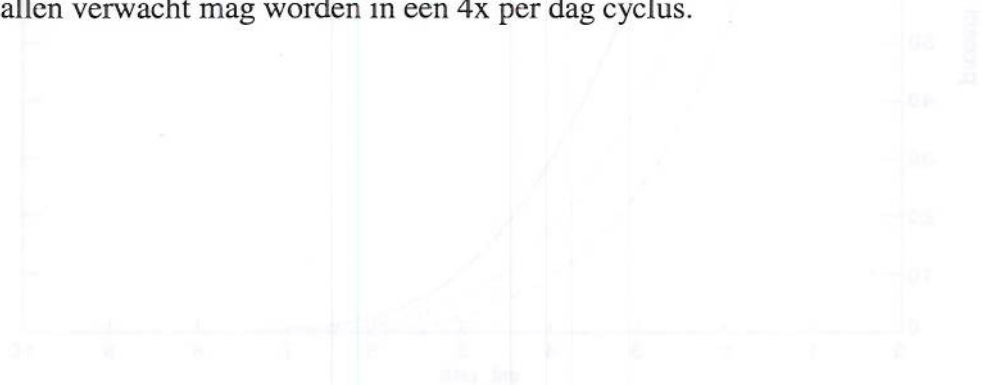


Fig. 5. Resultaten voor de tweede definitie van springerigheid (of maat) van springerigheid, i.e. de maximale som van 2 opeenvolgende sprongen. De x- en y-assen zijn respectievelijk de maximale som van 2 opeenvolgende sprongen en de frequentie van deze sprongen. De drie curves corresponderen met de runfrequenties 1x, 2x en 4x per dag. De curve voor 4x per dag is getrokken, de andere twee zijn gestippeld. De y-assen zijn in het onderste panel op een logaritmische schaal weergegeven.

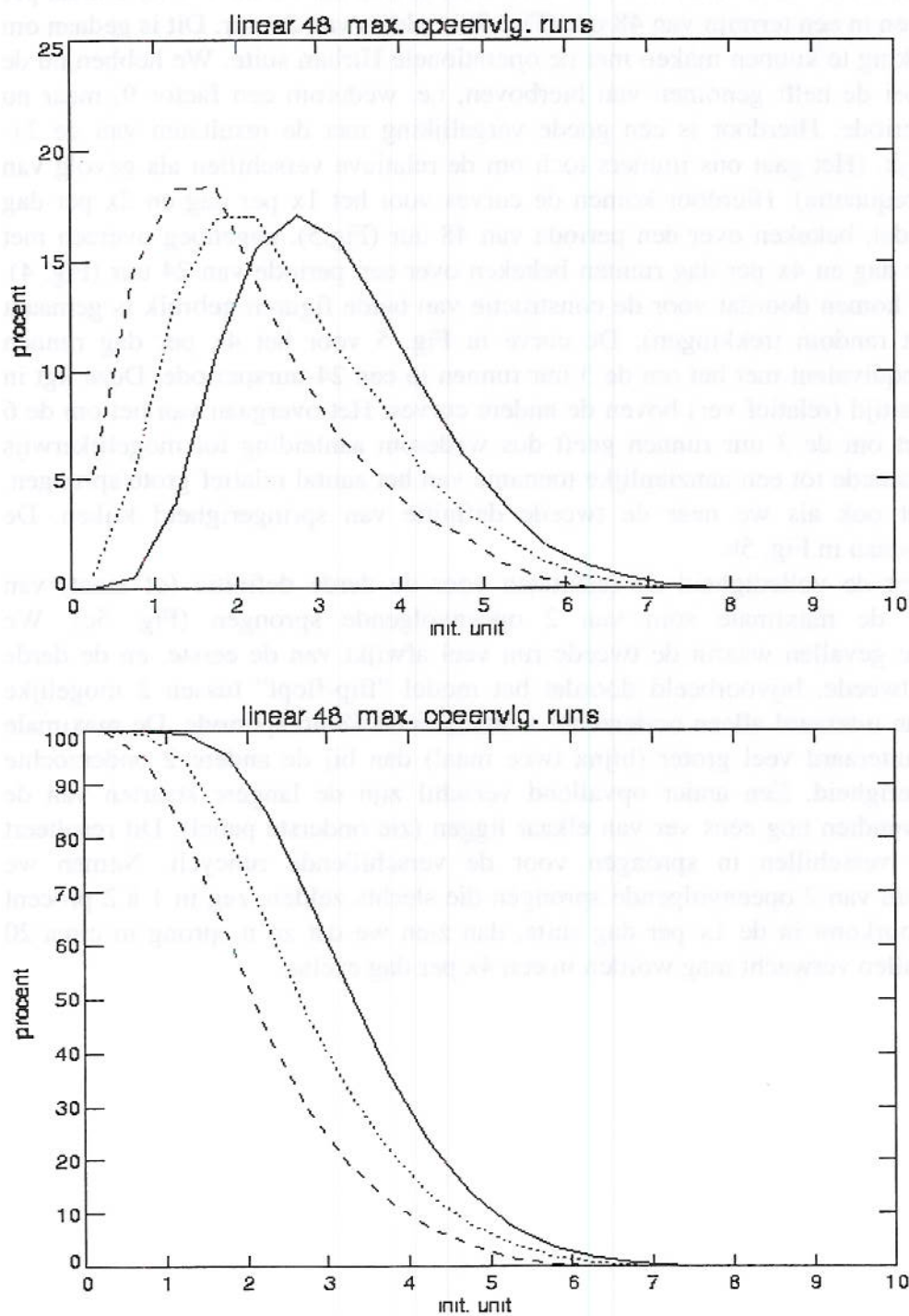


Fig. 5a. Frekwentieverdeling van het verwachte maximale verschil tussen 2 opeenvolgende runs (bovenste panel) en de cumulatieve verdeling (onderste panel) bij het 1x (dashed line), 2x (dotted line) en 4x (full line) runnen per dag bekeken over 48 uur onder de aanname van lineaire foutengroei ('linear').

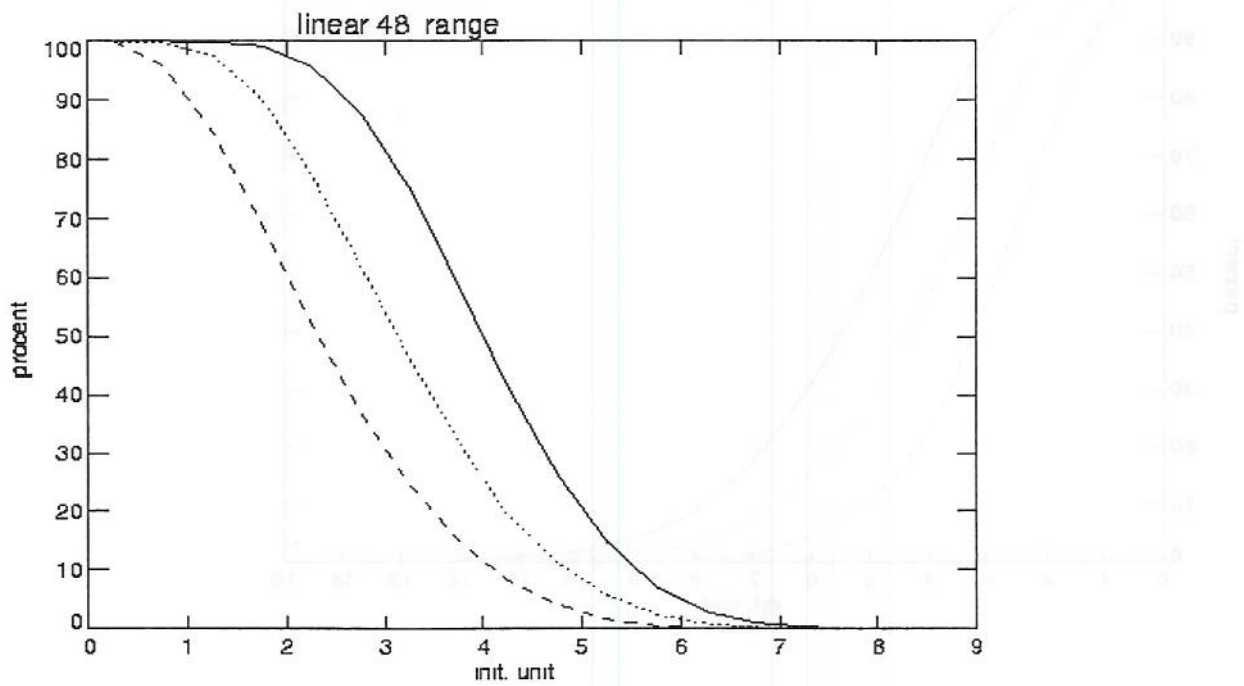
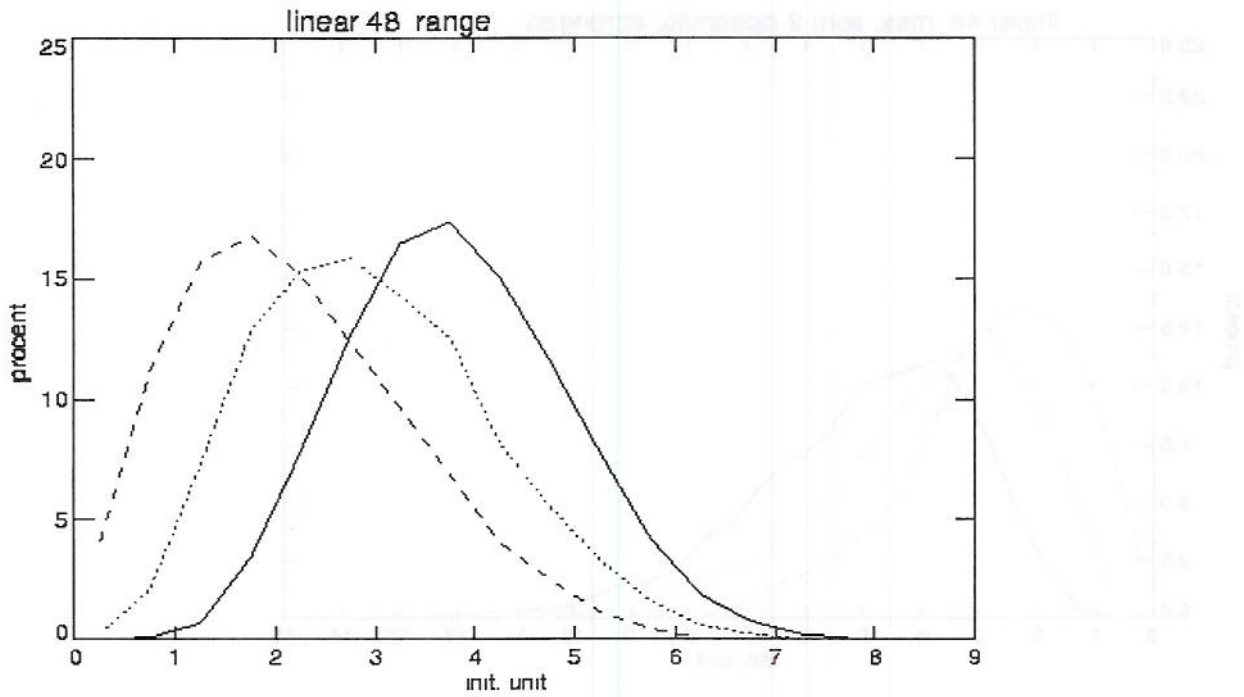


Fig. 5b. Als in Fig. 5a, maar voor de maximale range van de verwachtingen voor hetzelfde verifiërende tijdstip.

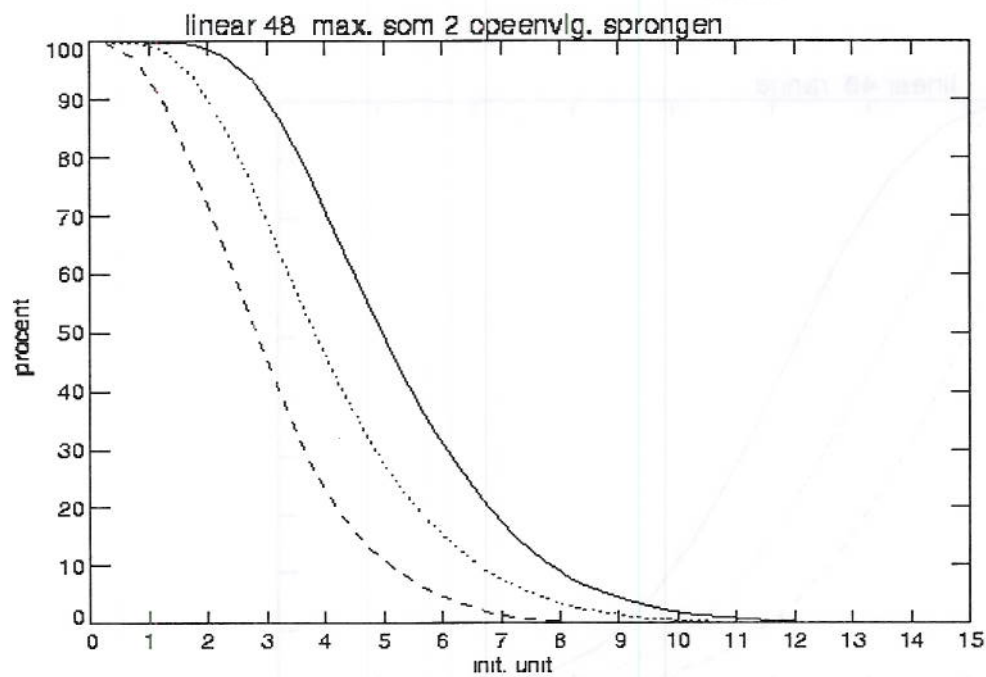
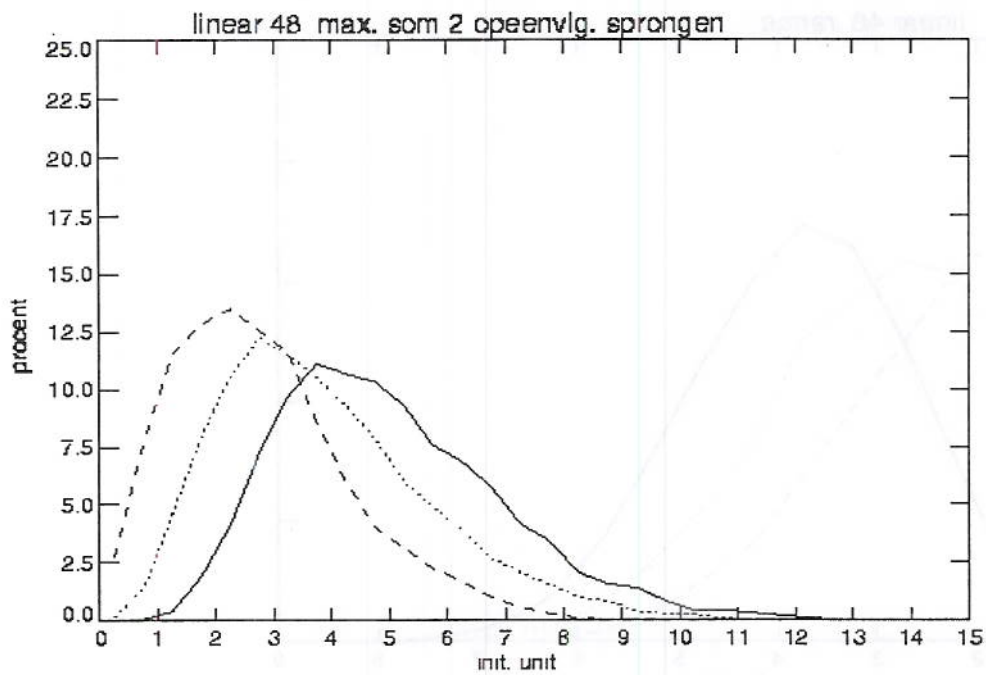


Fig. 5c. Als in Fig. 5a, maar voor de maximale som van 2 opeenvolgende sprongen voor hetzelfde verifiërende tijdstip.

4. Discussie

a. *summary*

"Springerigheid" ofwel inconsistentie van modellen wordt gezien als een hinderlijk fenomeen bij het maken van weersverwachtingen. In bovenstaand verhaal is gepoogd om één van de vele aspecten die een rol kunnen spelen bij de gesignaleerde springerigheid van modellen te isoleren van alle andere. We hebben voor een drietal definities (maten) van springerigheid de toename onderzocht louter als functie van een toenemende runfrequentie, hier alleen het 1, 2 en 4 maal runnen in een vaste periode van 24 en 48 uur. Dit is gedaan door middel van een paar simpele aannamen waaronder die van lineaire foutengroei. Dit is een realistische aanname bij korte voorspeltermijnen. Op onze andere aannamen komen we verder in deze paragraaf nog kort terug. Aangetoond is dat simpelweg door het vaker draaien van hetzelfde model springerigheid veel vaker te zien zal zijn en bovendien dat de grootte van de sprongen ook zal toenemen. Dit is een puur statistische fenomeen dat inherent is verbonden aan het vaker draaien van modellen en dat niks te maken heeft met hun skill. Het Hirlam model zal dan ook vaker een grotere sprong en een grotere inconsistentie vertonen dan het ECMWF model (afgezien van eventuele tekortkomingen van de beide modellen). Dit wordt nog "erger" als de runfrequentie verder opgevoerd wordt.

b. *wanneer de grootste springerigheid?*

Hoe groot de amplitude is van de sprongen (uitgedrukt in fysische grootheden) is niet in de discussie betrokken. En hoe deze zich verhoudt tot die van andere oorzaken is ook niet onderzocht. Andere aspecten zijn misschien belangrijker, zeker als niet alleen gekeken wordt naar de springerigheid van opeenvolgende runs van hetzelfde model maar ook nog eens naar die tussen verschillende modellen. Zie ook onder *d*. In onze analyse hebben we ons niet bekommerd om de grootte van de foutengroei. We hebben steeds aangenomen dat die hetzelfde was voor alle runs waarvan de forecasts vergeleken worden, i.e. 5 of 9 voor resp. de 24 of 48-uursperiode. Alle conclusies gelden voor kleine zowel als grote foutengroei, al zullen we in het eerste geval weinig tot niets van een grotere springerigheid merken.

In veel gevallen is de gevoeligheid van de atmosfeer voor kleine verstoringen relatief klein, i.e. de foutengroei is klein. (En hopelijk is in die gevallen ook de gevoeligheid van de modelverwachting gering voor kleine fouten in de analyses of in het model.) In dat geval is de voorspelbaarheid relatief groot en zal de forecaster niet veel last hebben van de springerigheid van het model. Soms echter zal de atmosfeer lokaal juist extreem "gevoelig" zijn (bijv. in sterk barokliene situaties). Dit zijn juist de gevallen waarin de foutengroei groot is en waarin de verschillen in springerigheid zichtbaar worden.

c. *verschil in kwaliteit van analyses*

In bovenstaand verhaal is aangenomen dat de kwaliteit van de analyses binnen één verwachtingencyclus hetzelfde is. Door het feit dat niet op ieder uur evenveel waarnemingen beschikbaar zijn zal dit echter niet altijd zo zijn. Runs startend op 06 en 18 uur GMT hebben een relatief slechtere uitgangspositie. Dit heeft tot gevolg dat de kans op grote sprongen toeneemt ten opzichte van bovenstaande verhaal. Omdat de 24-uurs en de 12-uurs cyclus dit nadeel niet hebben, zal het verschil in springerigheid tussen de 6-uurlijkse suite en de andere

alleen nog maar toenemen. Dit onverlet het feit dat de 06 en 18 runs duidelijk een toegevoegde waarde hebben boven de runs van 6 uur eerder.

d. systematische modelfout

In bovenstaande voorbeelden is uitgegaan van een model dat geen systematische fouten vertoont. De analysefout is random (en dus opeenvolgende analysefouten zijn onafhankelijk!) en de voorspelfout is louter het gevolg van de beperkte voorspelbaarheid van de atmosfeer. Vertoont het model echter wel een systematische fout dan kan dit gevolgen hebben voor de springerigheid van het model.

Veronderstellen we bijvoorbeeld een systematische fout in de analyses. Deze kan bijvoorbeeld zijn ontstaan door een modeldrift die door de invloed van de first guess niet volledig gecorrigeerd wordt door de waarnemingen. Hierdoor kan de kansverdeling van de fout in de analyses een asymmetrische vorm krijgen (in tegenstelling tot Fig. 1). Bijvoorbeeld te hoge of te lage waarden van de betreffende grootte (bijvoorbeeld de totale kinetische energie) komen te vaak voor. Dit resulteert bij gelijkblijvende andere aannames nu allereerst in een bias in de verwachtingen. Interessanter voor de context van dit pamflet is dat er ook een systematisch effect optreedt in de springerigheid. Deze zal afnemen. Het is voor te stellen dat als de staarten van de verdeling van de initiële fout wezenlijk veranderen door het introduceren van een systematische fout dat dan de verwachtingswaarde voor (bijv.) het maximale verschil tussen opeenvolgende runs ook afneemt. Dit betekent dus dat de springerigheid gemiddeld genomen afneemt terwijl de skill van het model ook afneemt. Weinig springerigheid of een hoge consistentie hoeven dus geenszins te duiden op een "goed" model (misschien zelfs omgekeerd).

Zo'n soort systematische modelfout kan resulteren in een toename van het aantal gevallen waarin het model in opeenvolgende runs de goede oplossing "geleidelijk te pakken krijgt" (ofwel min of meer monotoon naar de goede oplossing gaat), i.e. in de -24 run de depressie 200km ten W van de positie in de verifiërende analyse, en in de volgende runs steeds een beetje meer naar het oosten. Dit gedrag hoeft dus niet in alle gevallen wenselijk te zijn.

e. hogere versus lagere modelresolutie

Juist om de statistische springerigheid te isoleren van andere oorzaken is in ons verhaal steeds uitgegaan van hetzelfde model. De resolutie van dat model is niet relevant: dezelfde resultaten treden op bij iedere modelresolutie. Vergelijken we echter bijvoorbeeld een grootschalig model gerund in een 24-uurs voorspelcyclus met een hoog-resolutie model met een 6-uurs cyclus dan speelt bovenop de bovenstaande effecten nog iets anders een rol. Laag-resolutie modellen zijn in het algemeen minder goed in staat om zeer snelle ontwikkelingen en kleinschalige structuren te beschrijven. Hogere resolutie modellen zouden in principe in staat moeten zijn extremere (maar realistische) oplossingen te produceren, met andere woorden beter de werkelijke oplossingsruimte te vullen. Als we daarnaast accepteren dat juist deze kleine structuren veel minder goed voorspelbaar zijn (en in deterministische zin ook zullen blijven) dan zullen de fouten in individuele runs van het hoog resolutie model groter uit kunnen vallen. De springerigheid zou dan ook kunnen toenemen bij de overgang naar een hogere resolutie met eenzelfde runfrequentie. Net als in de eerdere voorbeelden heeft dit wederom niks te maken met verschillen in voorspellende waarde tussen de modellen. (Met

de *voorspellende waarde* wordt hier vooral bedoeld de probabilistische informatie - waarschuwingsfunctie, e.d. - die vervat is in de modelverwachtingen; zie Kok, 2001).

We kunnen echter nog een stap verder gaan: de toegevoegde waarde van hogere resolutie modellen over grootschalige zou wel eens onvermijdelijk vergezeld kunnen gaan van een af en toe optredende grotere springerigheid. Ofwel: een “beter” model vertoont een grotere inconsistentie. Of een gesignaleerde grote springerigheid inderdaad samengaat met een toegenomen skill (waarschuwingsfunctie) is niet uit een klein aantal gevallen af te leiden. Omgekeerd betekent het natuurlijk niet dat iedere toename van springerigheid een indicatie is voor een potentiële meerwaarde van het hoogresolutie model. In welke mate de grotere springerigheid die in een paar gevallen geconstateerd is, te verklaren is door een van bovengenoemde oorzaken is misschien de moeite waard om te onderzoeken.

Dankbetuiging

Veel dank is verschuldigd aan vooral Ben Wichers Schreur en Wim de Rooy voor de vele discussies over dit onderwerp. Daarnaast worden Gerard Cats, Robert Mureau, Wim de Rooy, Seijo Kruizinga en Daan Vogelesang bedankt voor het doorlezen van een eerdere versie van dit verhaal. Tenslotte Han Mellink voor zijn uitspraak (zie titelblad) die de boodschap van dit verhaal kernachtig samenvat.

Relevante Lectuur

ECMWF User Guide to ECMWF forecast products. Versie 3.2. 2002.

Kok, C. J., 2001. Probabilistic approach in comparative verification of mesoscale models.

In: Proceedings of the SRNWP workshop on the verification of mesoscale numerical weather forecasts. De Bilt, 23-24 April 2001.

Persson, A. (1995). On the skill and consistency in medium range weather forecasting.

ECMWF Newsletter Number 70 – Summer 1995.

Simmons, A.J., R. Mureau, and T. Petroligis (1995). Error growth and predictability

estimates for the ECMWF forecasting system. *Q. J. R. Met. Soc.* **121**, 1739-1771.

Appendix exponentiële foutengroei

Voor de volledigheid staan hier de resultaten onder de aanname van exponentiële foutengroei voor forecastperiodes van 24 en 48 uur (Figs. 6 en 7 resp.). Wederom voor twee definities van springerigheid. De schaal van de x-as is weer uitgedrukt in de eenheid van de breedte van de initiële verstoring. In de hier gegeven voorbeelden is de exponentiële groei zodanig gekozen dat de foutengroei na 24 uur een factor 9 groter is dan de initiële fout, dus analoog aan alle voorbeelden uit de eerdere hoofdstukken. Dit is gedaan om een makkelijke vergelijking met de voorbeelden eerdere hoofdstukken mogelijk te maken.

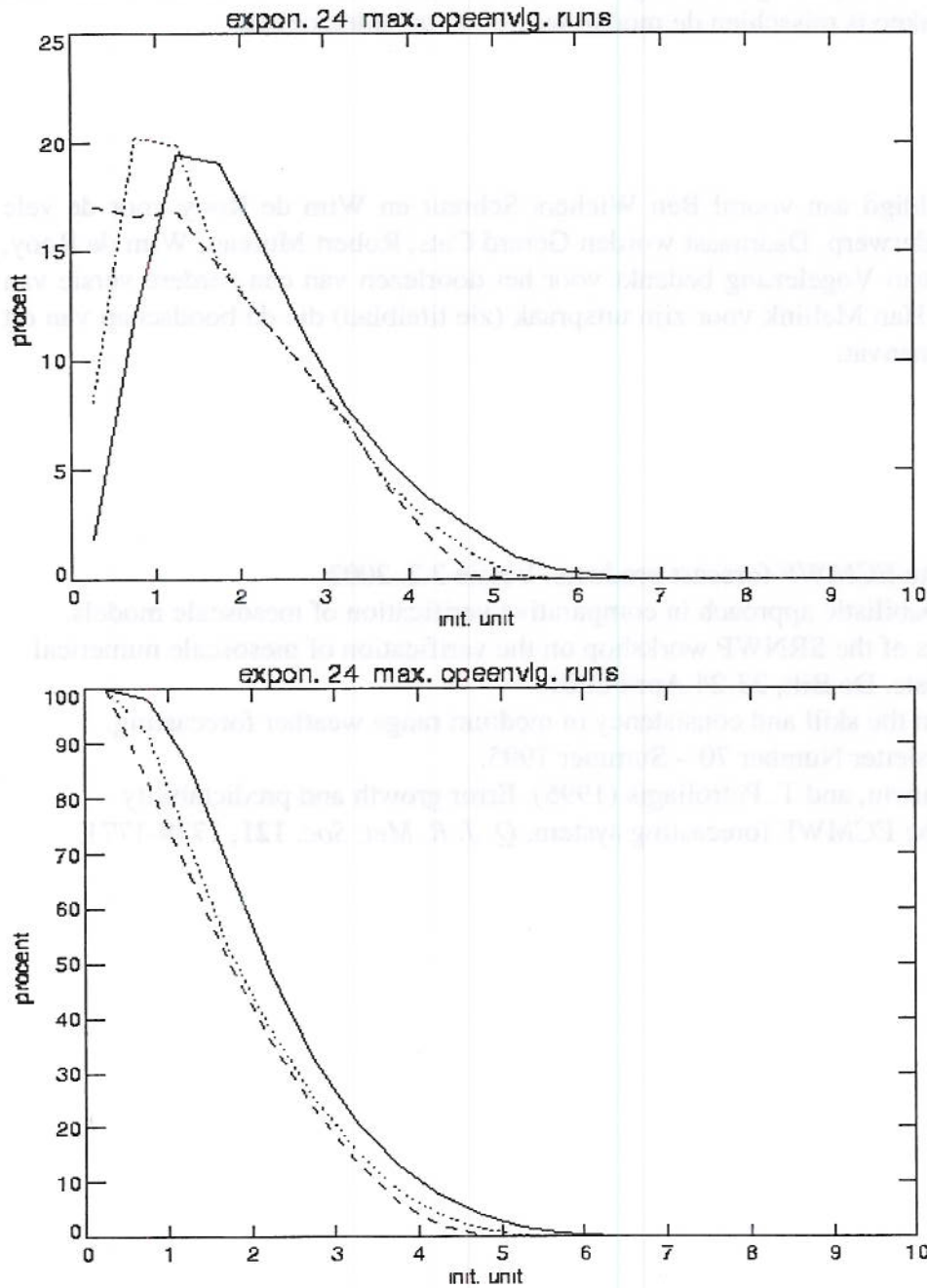


Fig. 6a. Frekwentieverdeling van het verwachte maximale verschil tussen 2 opeenvolgende runs (bovenste panel) en de cumulatieve verdeling (onderste panel) bij het 1x (dashed line), 2x (dotted line) en 4x (full line) runnen per 24 uur onder de aanname van exponentiële foutengroei ('expon.').

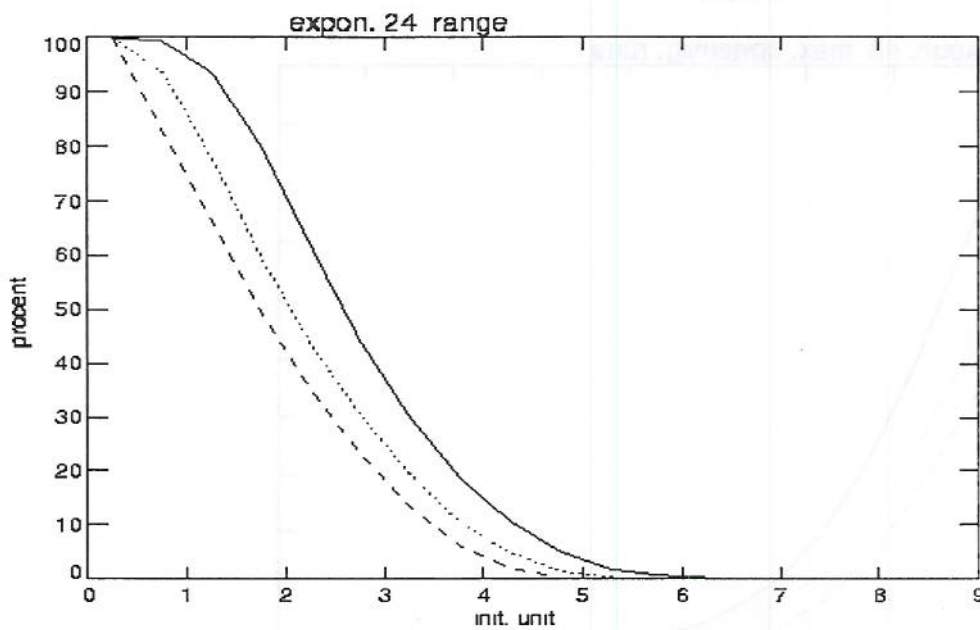
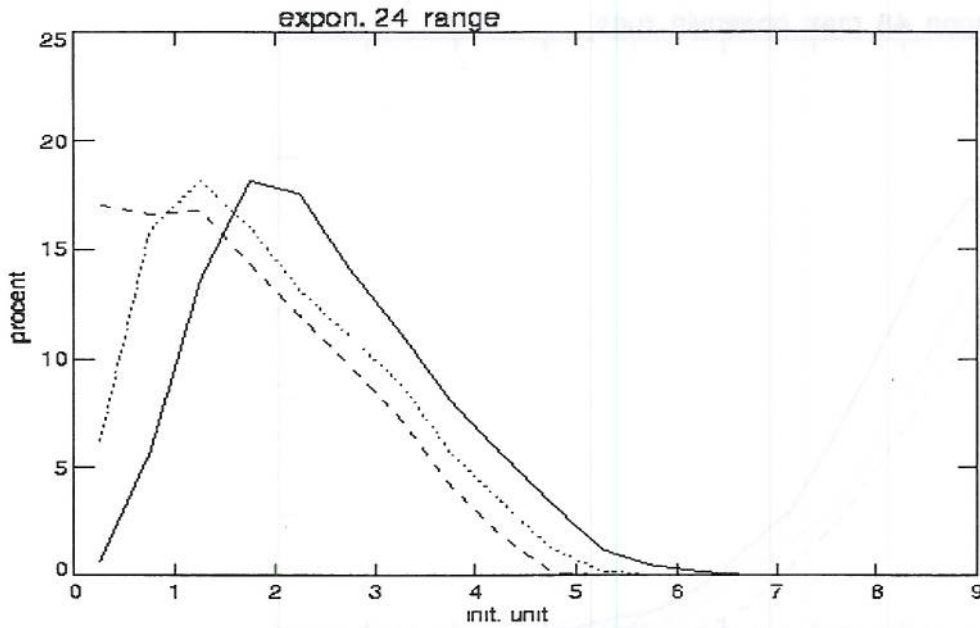


Fig. 6b. Als in Fig. 6a, maar dan voor de maximale range van de verwachtingen van de runs verifiërend op hetzelfde tijdstip.

De overeenkomsten met de resultaten bij lineaire groei zijn erg groot. Alle conclusies kunnen daarom overgenomen worden. Wel is nu uiteraard het verschil in springerigheid bij relatief grote waarden tussen de 3 suites relatief minder dan in het lineaire groei geval.

Door geen exponentiële groei te nemen maar juist de snelste foutengroei vlak na -24 of -48 te laten plaatsvinden en later af te laten nemen, kan juist een groter verschil tussen de 3 voorspelcycli gecreëerd worden (niet getoond).

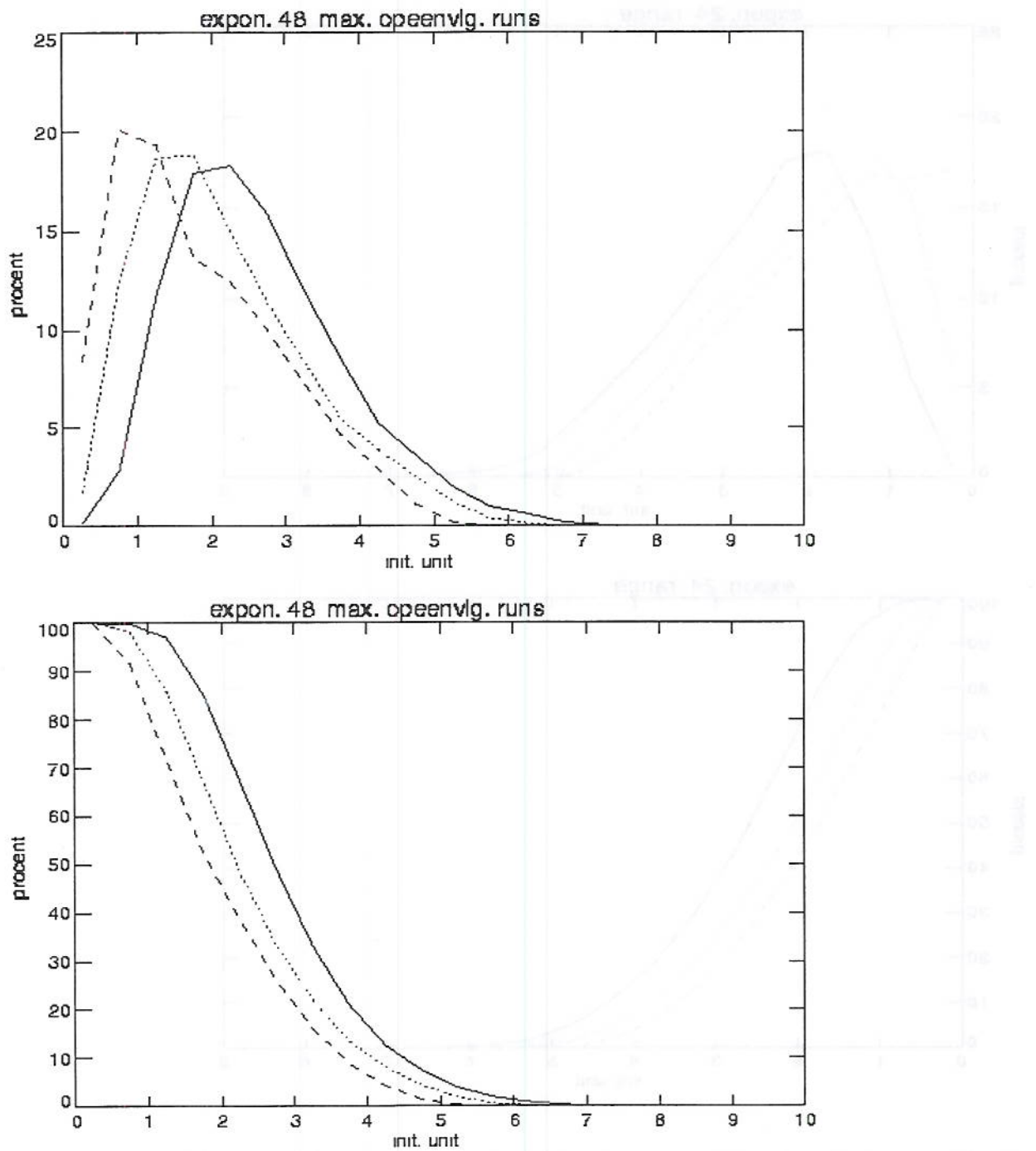


Fig. 7a. Frekwentieverdeling van het verwachte maximale verschil tussen 2 opeenvolgende runs (top panel) en van de cumulatieve verdeling (bottom panel) bij het 1x (dashed line), 2x (dotted line) en 4x (full line) runnen per dag bekeken over 48 uur onder de aanname van exponentiële foutengroei ('expon.').

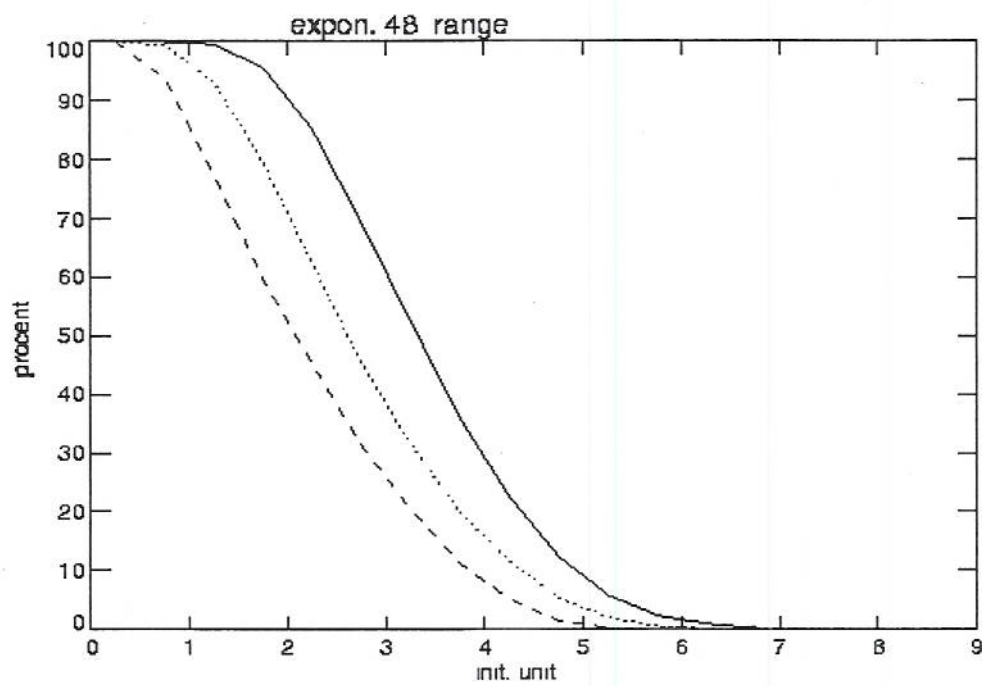
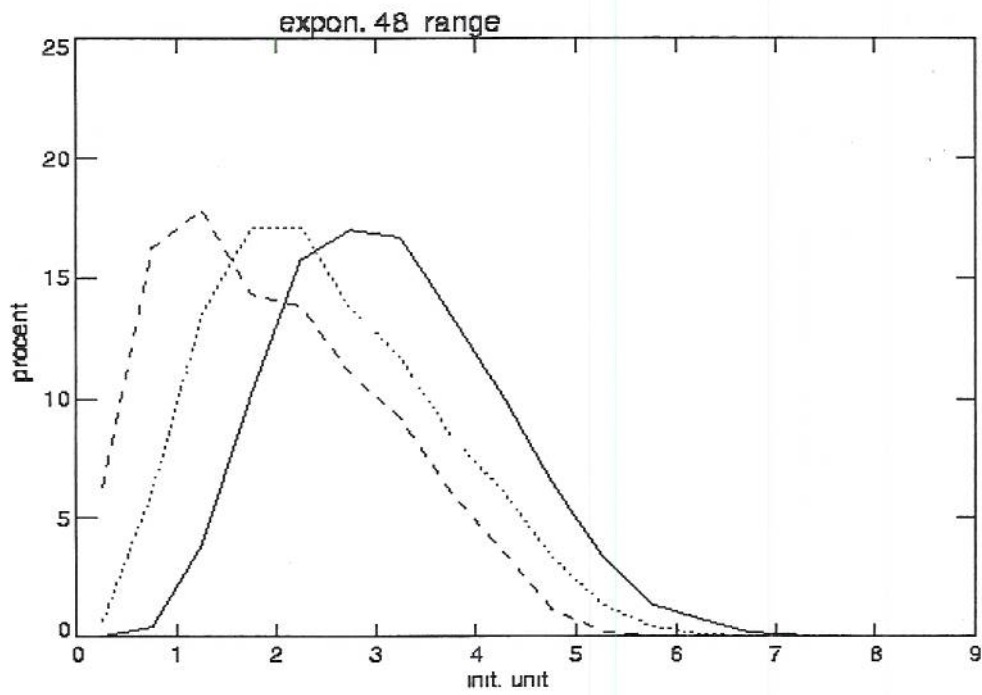


Fig. 7b. Als Fig. 7a, maar voor de maximale range van alle runs verifiërend op hetzelfde tijdstip.

