Dagelijkse variaties in de foutengroei en Ensemble verwachtingen

Leerzame Experimenten met een Eenvoudig Niet-lineair Systeem (2)

Seijo Kruizinga en Kees Kok

In een vorig artikel (Kruizinga, 2003) hebben we een eenvoudig niet-lineair systeem beschreven dat we willen gaan gebruiken om bepaalde aspecten van Ensemble verwachtingen te bestuderen. Met behulp van een lange integratierun hebben we een referentieatmosfeer gecreëerd die we gaan gebruiken om deze studies uit te voeren. We hebben in het vorige artikel gezien dat kleine verstoringen in de begintoestand leiden tot exponentiële foutengroei met als gevolg een ernstige beperking van de voorspelbaarheid in dergelijke (niet-lineaire) systemen.

In dit artikel willen we

(1) nog wat nader ingaan op die exponentiële foutengroei en op basis daarvan duidelijk maken waarom een verwachting voor de foutengroei nuttig kan zijn.
(2) de Ensemble verwachtingen introduceren en aannemelijk maken dat dat type verwachtingen in staat is om verwachtingen voor de foutengroei te produceren.

Bij de komende experimenten zullen we nog steeds gebruik maken van een perfect model en ons dus concentreren op de foutengroei die uitsluitend het gevolg is van fouten in de beschrijving van de begintoestand. Veel van de gevonden resultaten zullen overeenkomst vertonen met resultaten die ook bij atmosfeermodellen worden gevonden. We moeten echter voorzichtig zijn met de directe interpretatie in atmosferische termen.

Wat bedoelen we met “de fout”

Voor we verder gaan volgt eerst een aantal definities en afspraken. We zullen veel gebruik maken van de term "de fout in de verwachting" of de "fout in de begintoestand". Soms bedoelen we hiermee een op een bepaald moment opgetreden verschil tussen de verwachte (of waargenomen) waarde en de werkelijke waarde. Meestal echter bedoelen we daarmee de standaarddeviatie van de verschillen tussen verwachte (of waargenomen) waarden en werkelijke waarden. Om de leesbaarheid van het artikel te vergroten en niet steeds voluit "de standaarddeviatie van… " te moeten schrijven zullen we dit afkorten tot bijvoorbeeld de verwachtingsfout(sd) of beginfout(sd). Voluit betekent dan het laatste "de standaarddeviatie van de fout in de waargenomen (of geanalyseerde) begintoestand".

Foutengroei in afhankelijkheid van de beginfout(sd)

Mediaan

In figuur 1 is in de middelste curve nogmaals de foutengroei (beginfout(sd)=0,01) gepresenteerd zoals berekend in het eerste artikel. Nu hebben we echter niet de wortel uit het gemiddelde van de kwadratische fouten uitgezet maar de mediaan van 10.000 dagelijkse RMSE’s (Root Mean Square Errors). Voor de mediane waarde geldt dat 50% van alle verwachtingen een grotere RMSE (fout) heeft (slechter is) dan die gegeven waarde en de overige 50% van de verwachtingen een kleinere RMSE heeft en dus beter is. Het beeld van de foutengroei verandert daardoor niet wezenlijk maar de mediaan is iets gemakkelijker te interpreteren.

Foutengroei

Bij vergelijking met figuur 4 uit het vorige artikel valt op dat bij lage en hoge voorspeltijden de medianen ongeveer gelijk zijn aan het gemiddelde. Echter in het groeigedeelte van de curve is de mediaan aanzienlijk kleiner dan het overeenkomende gemiddelde. Dat verschil wordt veroorzaakt doordat in het groeigedeelte zeldzaam voorkomende grote fouten relatief veel bijdragen aan het gemiddelde maar de mediaan slechts in geringe mate beïnvloeden. De curve loopt dan ook aanzienlijk minder steil, de verdubbelingstijd van de fout is in deze curve dan ook hoger, namelijk 2,1 dag in plaats van ongeveer 1,5 dag. In die figuur zijn ook de foutengroei curven uitgezet voor andere waarden van de standaarddeviatie van de beginverstoring en wel voor een relatief grote beginfout(sd)= 0,3 (de bovenste curve) en voor een relatief kleine beginfout(sd)=0,001 (de onderste curve). Duidelijk is te zien dat deze curven in het groeigedeelte parallel lopen maar uiteindelijk op dezelfde verzadigingswaarde uitkomen. In het groeigedeelte van de curven is de groeisnelheid dus voor alle curven hetzelfde, een verdubbeling van de fout in omstreeks 2,1 dag. Van deze drie curven stemt de bovenste het beste overeen met de groeicurve van de huidige numerieke modellen. Bij de huidige kwaliteit van het analysesysteem en de numerieke modellen is namelijk de RMSE van de dag 1 verwachtingen eveneens ongeveer 5 à 10 % van de RMSE van de klimatologie (Simmons, 1995).

Voorspelbaarheidshorizon

Figuur1 illustreert heel mooi het probleem van de (gelimiteerde) voorspelbaarheid van de atmosfeer. Als we bijvoorbeeld een mediane waarde voor de fout(sd) van 1,0 acceptabel vinden dan halen we in de bovenste curve dag vier, in de middelste curve dag 15 en in de onderste curve dag 22. Om echter van dag vier naar dag 15 en vervolgens naar dag 22 te komen moeten we de begintoestand wel respectievelijk 30 en 300 keer zo nauwkeurig bepalen. In de huidige atmosfeermodellen zou dat overeenkomen met bijvoorbeeld een fout(sd) in de analyse van de 500 hPa hoogte, van 30 cm respectievelijk 3 cm als we uitgaan van de huidige fout(sd) in de analyse van omstreeks 10 meter.

Dagelijkse variaties in de foutengroei

De medianen zoals hiervoor gepresenteerd zijn vastgesteld met behulp van de frequentieverdelingen van de dagelijkse RMSE-waarden. Naast de mediaan zijn we ook in staat andere percentielen te bepalen. Bijvoorbeeld het 80% percentiel waarvoor geldt dat 80% van alle verwachtingen beter is (een lagere RMSE heeft) en een 20% percentiel waarvoor geldt dat slechts 20% van de verwachtingen beter is. In figuur 2 zijn voor de middelste curve van figuur 1 naast de mediaan ook de 80% percentielen en de 20% percentielen als functie van de voorspeltijd uitgezet. Wat opvalt in deze grafiek is dat zowel bij lage voorspeltijden als bij hoge voorspeltijden deze drie curven dicht bij elkaar liggen maar in de groeifase sterk uiteen lopen. Als de groeisnelheid, in dit geval een verdubbeling van de fout in 2,1 dagen, altijd hetzelfde is onafhankelijk van de bijvoorbeeld de begintoestand en de beginfout(sd) dan verwachten we dat deze curven, op de hier gebruikte logaritmische schaal, parallel lopen. Kennelijk varieert de groeisnelheid van dag tot dag, mogelijk afhankelijk van de begintoestand. Dit kan aanleiding geven tot aanzienlijke verschillen in voorspelbaarheid.

Bijvoorbeeld bij dag 15 is de mediane RMSE omstreeks 1,0. Echter in 20% van de gevallen is de RMSE kleiner dan 0,2 en eveneens in 20% van de gevallen is de RMSE groter dan 3,0.

De interpretatie kan ook in de andere richting. Indien we een RMSE van 1,0 acceptabel vinden dan reikt de voorspelbaarheid in 20% van de gevallen minstens tot dag 20 maar anderzijds in 20% van de gevallen nog niet tot dag 10. Ook bij andere beginfouten(sd) treden deze variaties op, maar niet altijd in even grote mate. De variatie in groeisnelheid kunnen we in beeld brengen door de verhouding tussen de waarde van het 80% percentiel en de waarde van het 20% percentiel te berekenen. Bij een kleine verhouding liggen de curven van de percentielen dicht bij elkaar. Is de verhouding groot dan liggen de curven van de percentielen juist ver uit elkaar en als de verhouding onafhankelijk is van de voorspeltijd dan lopen de curven parallel. In figuur 3 zijn deze verhoudingen geplot als functie van de voorspeltijd voor de drie groeicurven van figuur 1. We zien dat deze verhouding vrij klein is bij voorspeltijd 0 (ook bij constante standaarddeviatie van de beginverstoringen zullen de dagelijkse verstoringen in geringe mate variëren in sterkte) maar snel toeneemt met stijgende voorspeltijd tot omstreeks 4,5 als de beginfout(sd) groot is (standaardeviatie 0,3) en zelfs tot omstreeks 25 bij kleine beginfout(sd). Bij lange voorspeltijden daalt het quotiënt weer tot waarden die de statistische variatie in de grootte van de verstoring van de begintoestand weergeven. In het Ensemble systeem van het ECMWF is de maximale verhouding omstreeks 2,0 (Kruizinga, 2001) en dus veel kleiner dan bij het Lorenz-systeem.

Ensemble verwachtingen

Voorspelling van de betrouwbaarheid: skill prediction

De groeisnelheid van de fout kan sterk van dag tot dag variëren en het zou heel nuttig zijn indien we per verwachting aan zouden kunnen geven of we te maken hebben met een verwachting met hoge of lage groeisnelheid in de fout. Oftewel of we te maken hebben met een verwachting met mogelijk een lage dan wel hoge kwaliteit. Bij grote groeisnelheid kunnen we lokaal of incidenteel nog steeds een verwachting van hoge kwaliteit hebben maar de kans op lage kwaliteit is veel hoger. Bij lage groeisnelheid is in alle gevallen de kwaliteit van de verwachting hoog. Meestal duiden we een dergelijke uitspraak over de groeisnelheid aan als Skill-prediction. Het Ensemble systeem is een potentiële kandidaat voor een dergelijk Skill-prediction systeem.

Het ensemble systeem

Het systeem van Ensemble verwachtingen berust op de aanname dat we weliswaar niet exact de beginfout kennen, maar dat we wel de statistische karakteristieken van de beginfout kennen. In ons geval is dat natuurlijk waar omdat we zelf de beginfout genereren. Bij de numerieke atmosfeermodellen ligt dat veel complexer. Op basis van deze kennis kunnen we een conditionele kansverdeling (zie kader) berekenen voor de werkelijke begintoestand van het systeem uitgaande van de waarnemingen. Met behulp van Ensemble verwachtingen wordt deze conditionele kansverdeling van de werkelijke begintoestand omgezet in een conditionele kans voor de werkelijke toestand na de verwachtingsperiode. De techniek van de Ensemble verwachtingen zullen we toepassen op het Lorenz-systeem met onafhankelijke toevallige waarneemfouten bij ieder der X-en.

Het genereren van de beginverstoringen

Deze toevallige waarneemfouten van X_1 t/m X_K zijn onderling onafhankelijk maar volgen allemaal een normale verdeling met gemiddelde 0,0 en standaarddeviatie 0,01. Omdat de waarneemfout(sd) relatief klein is, is de conditionele kansverdeling van de werkelijke waarde van X_1 een simpele Gaussklok rond de waargenomen waarde met gemiddelde 0,0 en standaarddeviatie 0,01. Bij grotere waarneemfouten moeten we de klimatologische verdeling van X_1 ook nog betrekken bij het vaststellen van de conditionele kansverdeling maar in dit geval is dat een verwaarloosbaar effect.

Het genereren van de begintoestanden

Om een Ensemble verwachting te maken gaan we nu als volgt te werk. We verstoren de waargenomen toestand, analyse plus toegevoegde toevallige waarneemfout, nogmaals door bij iedere X_1 t/m X_k weer (voor alle X-en onafhankelijk) een toevalsgetal uit de kansverdeling van de waarneemfouten op te tellen, oftewel we trekken een mogelijke begintoestand uit de conditionele kansverdeling van de werkelijke begintoestand. Uitgaande van deze nieuwe verstoorde beginsituatie berekenen we een nieuwe verwachting. Deze verwachting is dan het eerste lid van ons Ensemble. Deze procedure herhalen we vervolgens 49 keer, steeds met nieuwe verstoringen, om een Ensemble met 50 leden te genereren.

Toets in één variabele

Omdat de beginwaarden getrokken zijn uit een conditionele kansverdeling van de werkelijke toestand van het systeem veronderstellen we nu dat deze Ensemble verwachtingen op zich ook weer trekkingen zijn uit een (conditionele) kansverdeling van de werkelijke toestand van het systeem na de verwachtingsperiode. Als we deze veronderstelling willen toetsen dan moeten we, in ons geval, toetsingen doen aan 8-dimensionale vectoren en dat is over het algemeen nogal ondoorzichtig. Daarom vereenvoudigen we de toetsing door bijvoorbeeld alleen naar de verwachtingen op één locatie te kijken. In ons geval vereenvoudigen we de toetsing door alleen de Ensemble verwachtingen voor X_1 te analyseren. In het experiment dat we hebben uitgevoerd met het Lorenz systeem beschikken we nu dus over een verwachting voor X_1 uitgaande van de onverstoorde waarneming (dus de werkelijke toestand plus de fout in de waarneming), overeenkomend met de Control run in de Ensemble verwachtingen van het ECMWF, en 50 verwachtingen voor X_1 startend van situaties die verstoord zijn ten opzichte van de waarneming. Verder beschikken we uiteraard over de waarneming van X_1 voor het moment waarvoor de verwachting geldig is. Deze gegevens zullen we gebruiken om de Ensemble techniek te verifiëren. We gebruiken hiervoor uiteraard weer de 10000 dagen waarvoor we de referentie-atmosfeer hebben berekend.

Verificatie van Ensemble verwachtingen

Zoals gezegd veronderstellen we dat de leden van het Ensemble een steekproef zijn uit de conditionele kansverdeling voor de werkelijke waarde van X_1. Dat betekent dat het gemiddelde van het Ensemble (EM), zie kader, een goede verwachting moet zijn voor X_1 en dat de standaarddeviatie van het Ensemble (SD), zie kader, een maat is voor de mogelijke fout in de verwachting van X_1. We zullen eerst de standaarddeviatie SD onder de loep nemen. Als SD een maat is voor de fout in de verwachting dan moet SD in ieder geval overeenkomstige groeikarakteristieken hebben. We beschikken in ons geval voor een gegeven waarde van de voorspeltermijn k over 10000 SD waarden van de Ensembles. Van deze 10000 waarden hebben we weer een frequentieverdeling gemaakt en in figuur 4 zijn de 20%, 50% en 80%

percentielen geplot als functie van k. De groeicurven van de SD in figuur 4 vertonen opvallend veel overeenkomst met de groeicurven van de RMSE in figuur 2. Vooral het uiteenwaaieren van de percentielcurven tussen dag 1 en dag 15 is opvallend gelijk. Wel zijn de overeenkomstige SD-waarden steeds iets hoger. Echter de verzadigingswaarde van SD is duidelijk lager dan de verzadigingswaarde van de RMSE. De eerste is gelijk aan de klimatologische standaarddeviatie van X_1, immers als we de Ensemble verwachtingen maar lang genoeg door laten lopen dan zijn de 50 leden op den duur willekeurige trekkingen uit de klimatologische verdeling van X_1, met een overeenkomstige standaarddeviatie ten opzichte van het gemiddelde van het Ensemble. De verzadigingswaarde van de RMSE is, omdat het de standaarddeviatie van het verschil tussen twee willekeurige niet gecorreleerde tijdseries van X_1 betreft, gelijk aan $\sqrt{2}$*de standaarddeviatie van X_1. Deze overeenkomst in groeicurves in figuur 2 en fig. 4 toont aan dat Ensemble verwachtingen de potentie hebben om skillverwachtingen te maken.

Slot

We hebben laten zien dat de foutengroei in het door ons gebruikte Lorenz-systeem sterk varieert van dag tot dag. De overeenkomst met het gedrag van de atmosferische modellen is niet perfect maar voldoende om met behulp van experimenten met het Lorenz-systeem ons begrip van foutengroei en skillverwachtingen te vergroten. We hebben vervolgens Ensemble verwachtingen geïntroduceerd en laten zien dat dit soort verwachtingen mogelijk in staat is om skillverwachtingen te produceren.

Kader: conditionele kansverdeling ---
We moeten ons goed realiseren dat we bij het ensemble systeem te maken hebben met een conditionele kansverdeling. Dat wil zeggen dat we op grond van de gegeven waarneming en onze kennis van de statistische karakteristieken van de waarneemfout aan ieder interval (voor de mogelijke waarden van X_1) een kans toe kunnen kennen dat de werkelijke waarde in dat interval ligt. Een andere waarnemer met andere waarneemfouten zal tot een andere conditionele verdeling komen en dus tot andere kansen voor een gegeven interval. Beide verdelingen zijn even goed of slecht. Er bestaat dus niet zoiets als de ware verdeling van de werkelijke waarde. De werkelijke waarde is slechts één getal. Een andere manier om deze verdeling te omschrijven is om te spreken van de verdeling van de geloofwaardigheid (credibility). We zullen meer geloofwaardigheid toekennen aan een interval in de buurt van de waarneming en veel minder aan een interval ver van de waarneming.

Kader: ensemble gemiddelde en spreiding ---

We zullen bij de verificatie gebruik maken van de waargenomen en verwachte waarden voor X_1. We zullen daarom vaak het subscript 1 weglaten en de subscriptpositie gebruiken om een aantal andere grootheden aan te duiden.

Met $X_e(i,k,j)$ duiden we aan: het j^e lid van het ensemble van de k-daagse verwachting voor dag i. (i variërend van 1 tot 10000; j variërend van 1 tot 50; k variërend van 1 tot 40).

Verder introduceren we het gemiddelde van het ensemble:

$$\mathrm{EM}(i,k) = \frac{1}{50}\sum_{j=1}^{50} X_e(i,k,j)$$

en de standaarddeviatie van het ensemble

$$\mathrm{SD}(i,k) = \sqrt{\frac{1}{49}\sum_{j=1}^{50} \left(X_e(i,k,j) - \mathrm{EM}(i,k)\right)^2}$$

Bij de berekening van EM(i,k) en SD(i,k) wordt dus niet de controlrun meegenomen.

Literatuur

Kruizinga, S., 2001: Statistical Analysis of Local 500 hPa ECMWF Ensemble Forecasts. Proceedings of the Eigth Workshop on Meteorological Operational Systems, ECMWF, 12-16 november 2001, 44-49

Kruizinga, S., 2003: Leerzame experimenten met een eenvoudig niet-lineair systeem. Meteorologica, Juni 2003.

Simmons, A.J., 1995: The skill of 500 hPa height forecasts. In *Predictability,* Seminar Proceedings, Volume 1, 4-8 September 1995, ECMWF, Reading, UK

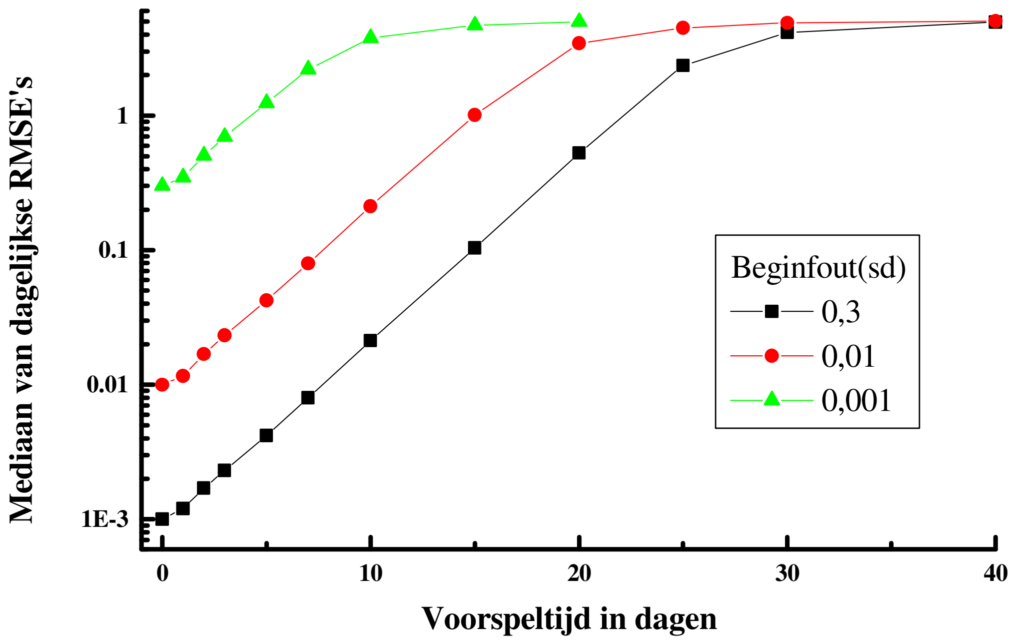

Fig. 1 Foutengroei in het Lorenz systeem als gevolg fouten in de begintoestand bij verschillende standaarddeviaties van de beginverstoringen.

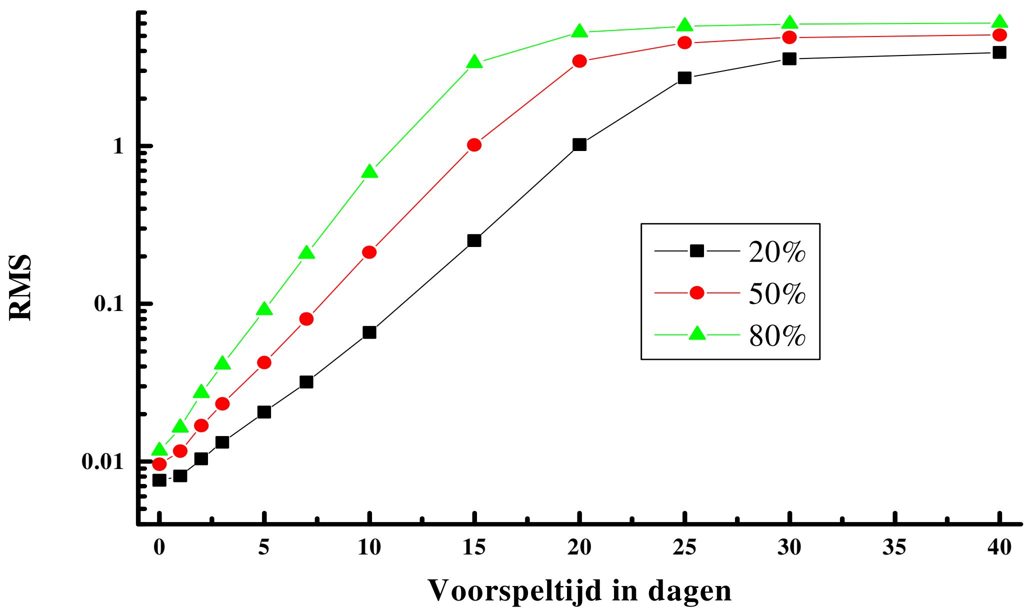

Fig.2 Mediaan en 20% en 80% percentiel van de foutengroei bij een beginsterkte van 0,01

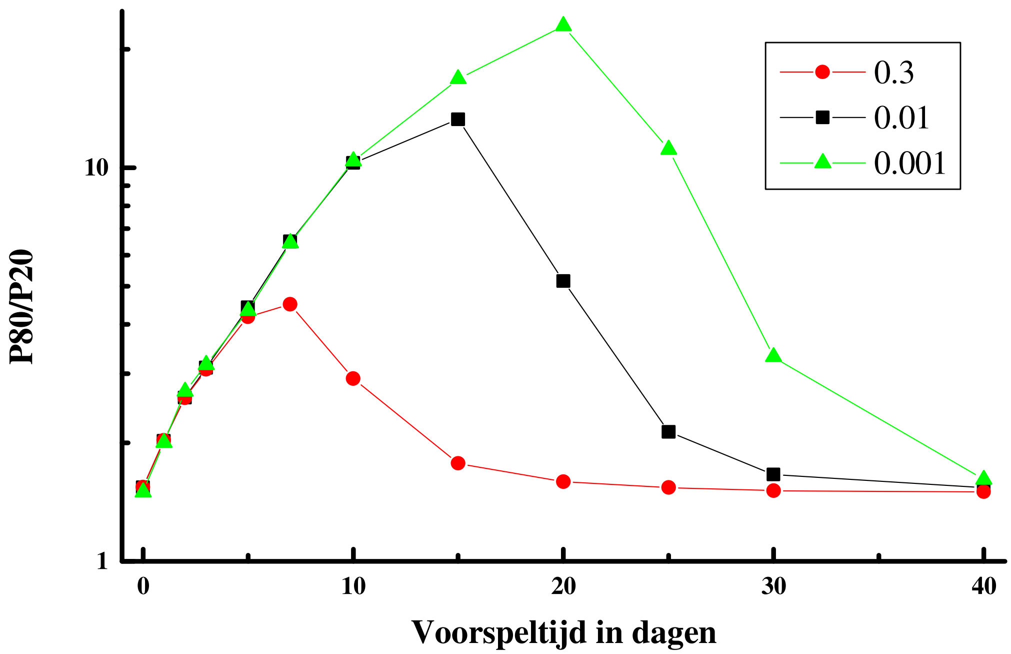

Fig. 3 Verhouding van 80% percentiel (P80) met het 20% percentiel (P20) als functie van de voorspeltijd voor verschillende sterktes van de beginverstoringen.

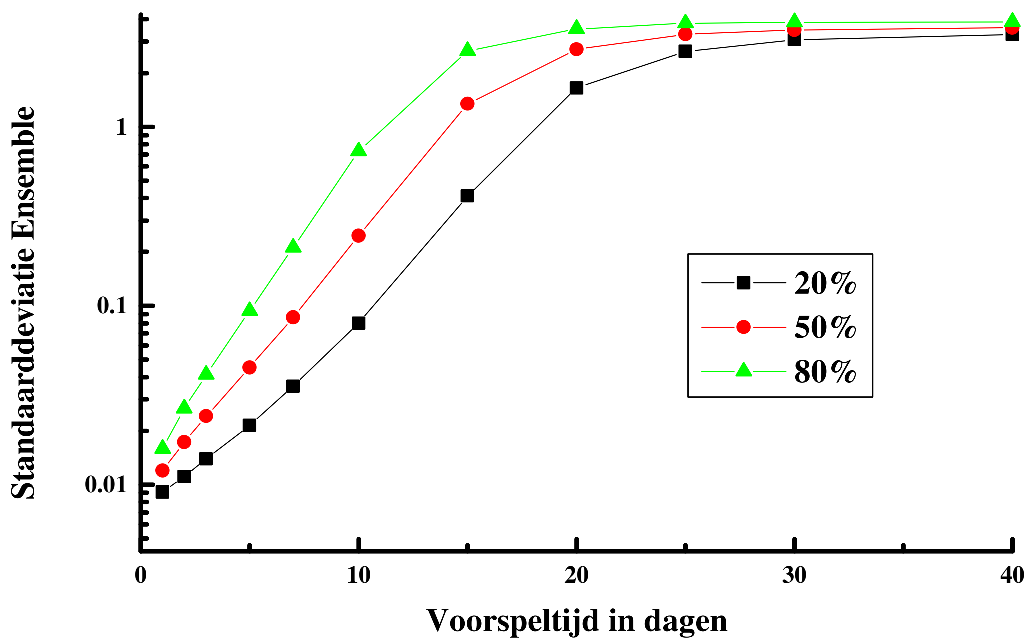

Fig. 4 Mediaan en het 20% en 80% percentiel van de standaarddeviaties van de Ensemble verwachtingen voor 10000 dagen versus de voorspeltijd in dagen.