

Hydrologisch modelleren, voorspellingsfouten en data assimilatie

Om over te kunnen gaan tot een doeltreffend beheer van water in stroomgebieden, dient men goed te weten hoe en waar water verdeeld is en wat de voornaamste processen binnen deze gebieden zijn. De toestand van een systeem en de evolutie ervan kan men kennen via 2 mogelijkheden: metingen of modelsimulaties. De beste schatting wordt bekomen door modelresultaten en metingen te combineren tijdens verschillende fasen in het modelleren. Deze combinatie wordt data-assimilatie genoemd. Het is belangrijk te beseffen dat iedere data-assimilatie techniek veronderstelt dat het model noch de metingen systematische fouten bevatten. Aan deze voorwaarde is echter zelden voldaan, zeker als men in detail naar modelvoorspellingen kijkt. Daarom wordt in dit artikel besproken hoe men via data-assimilatie tot optimale resultaten kan komen en hoe men rekening kan houden met systematische fouten.

Iedereen kent de kringloop van water, maar het is niet vanzelfsprekend om hydrologische processen precies en nauwkeurig te simuleren met een model. Het is ook onmogelijk om overal en altijd goede metingen van verschillende variabelen te bekomen. In dit artikel wordt getoond hoe men door een combinatie van metingen en fysische wetten in een model tot verbeterde schattingen kan komen. Dit wordt geïllustreerd voor één van de belangrijkste variabelen in de hydrologische cyclus, nl. het bodemvochtgehalte.

Het bodemvochtgehalte heeft een belangrijke invloed op de verdeling van water en energie over het land en op verscheidene fysische en scheikundige processen in de bodem, planten en atmosfeer. Op wereldschaal is het bodemvochtgehalte een bepalende factor voor het weer en het klimaat. Binnen Vlaanderen zien we gemakkelijk in dat het bodemvochtgehalte een cruciale rol speelt in de landbouw voor de plantengroei, de verspreiding van meststoffen en pesticiden, de erosie en de afvoer van water naar beken en rivieren. Verder kunnen we niet naast de problematiek van overstromingen en droogtes kijken.

Een eerste bron van informatie over bodemvochtgehalten zijn de klassieke metingen. Bodemonsters laten toe het bodemvochtgehalte vrij nauwkeurig te bepalen, maar het nadeel is dat deze metingen slechts op enkele tijdstippen en locaties kunnen genomen worden. Afstandswaarnemingen vanuit vliegtuigen of satellieten laten een meer gebiedsdekkende bodemvochtschatting toe, maar deze kunnen de variaties binnen kleine oppervlakten moeilijker in kaart brengen. Deze schattingen bevatten ook een relatief grote graad van onzekerheid, vooral onder gewasbedekking. Verder kan geen enkele meting volledig continu zijn in ruimte en tijd.

Hydrologische modellen vormen een tweede informatiebron. Ze simuleren de interactie tussen atmosferische variabelen (vooral neerslag), de vegetatie en het land. Het bodemvochtgehalte wordt berekend uit een reeks fysische vergelijkingen en kan bepaald worden op eender welke plaats en continu in de tijd.

Door een combinatie van metingen en fysische wetten in een model kan men tot verbeterde schattingen komen. Deze combinatie kan tijdens verschillende stadia in het modelleren gebeuren en wordt in aardwetenschappen dikwijls data-assimilatie genoemd.

Data-assimilatie

Om enigszins betrouwbare modelresultaten voor eender welke variabele te verkrijgen, zorgt men ervoor dat (i) de modelstructuur, i.e. de numerieke oplossing van de formules die de achterliggende fysische wetten beschrijven, correct zijn, (ii) de parameters geoptimaliseerd zijn voor de bestudeerde omstandigheden in het gemodelleerde gebied, (iii) de initiële toestand juist gedefinieerd is en (iv) de invoergegevens correct zijn.

De optimalisatie van deze factoren is typisch afhankelijk van observaties of metingen. De invoergegevens bestaan uit meteorologische (gemeten) data. Voor Vlaanderen beroept men zich vaak op gegevens uit Ukkel of op andere meteorologische stations dicht bij het bestudeerde gebied gelegen, maar niets garandeert dat dergelijke meteorologische gegevens ook echt weergeven wat het weer in een volledig stroomgebied is. De algemene fysische wetten die modelstructuur bepalen, zijn dikwijls (niet altijd) afgeleid uit (historische) geobserveerde verbanden tussen variabelen. Parameters zorgen er voor dat het model kan afgestemd worden op de bestudeerde plaatsen of omstandigheden. Optimale parameterschattingen worden bekomen door ervoor te zorgen dat het modelresultaat zo goed mogelijk overeenstemt met metingen gedurende een bepaalde kalibratieperiode. Het kalibreren van een model betekent dus dat gemeten informatie gebruikt wordt om optimale schattingen van parameters te verkrijgen. Analooeg wordt de beste initiële toestand gevonden door gemeten observaties op te nemen in de initiële toestandsschatting.

Zelfs indien men dan modelleert met de best mogelijke modelstructuur, meteorologische invoerdata, geoptimaliseerde parameters en initiële toestand, dan kan men nog onmogelijk verwachten dat de modelresultaten een perfecte weergave van de natuur zijn. Daarom worden modelresultaten best geregeld bijgestuurd door de toestand of parameters te optimaliseren met nieuwe (gemeten) informatie tijdens het simuleren. Dikwijls reserveert men de term 'data-assimilatie' enkel voor het bijsturen van de toestand, of m.a.w. het herinitialiseren van een modelsimulatie. Het best gekende voorbeeld hiervan is de dagelijkse weersvoorspelling: indien men vandaag 10 °C had verwacht, maar men meet 15 °C, dan zal men bij de voorspellingen voor morgen uiteraard niet volledig vertrouwen op het gesimuleerde vervolg van de verwachtingen van vandaag, maar veeleer rekening moeten houden met de observaties van vandaag.

Metingen zijn echter ook nooit foutloos. Daarom is het aangewezen om bij assimilatie een gewogen gemiddelde te nemen van de voorspellingen en de metingen op een bepaald ogenblik voor een bepaalde variabele. Zo kan men best voorspellingen voor morgen opnieuw berekenen uitgaande van bij voorbeeld 13 °C en niet 15 °C voor de temperatuur (deel van de weerstoestand) van vandaag. De beste schatting van de toestand kan berekend worden via een statistische filter waarbij de gewichten bepaald worden door de onzekerheden op de voorspellingen en de eigenlijke metingen. De Kalman filter (Kalman, 1960) zal de optimale combinatie geven, indien de onzekerheden gekend zijn en zowel het model als de observaties vrij zijn van systematische fouten. Het resultaat van de filtering wordt terug in het model gebruikt om verder te voorspellen. Dit wordt schematisch weergegeven in figuur 1.

Onzekerheden en ensembles

De onzekerheid of foutenmarge op metingen is dikwijls door de fabrikant meegegeven met het meetinstrument. Indien een observatie in een model moet gebruikt worden, moet de gebruiker bij de meetfout nog de onzekerheid optellen die

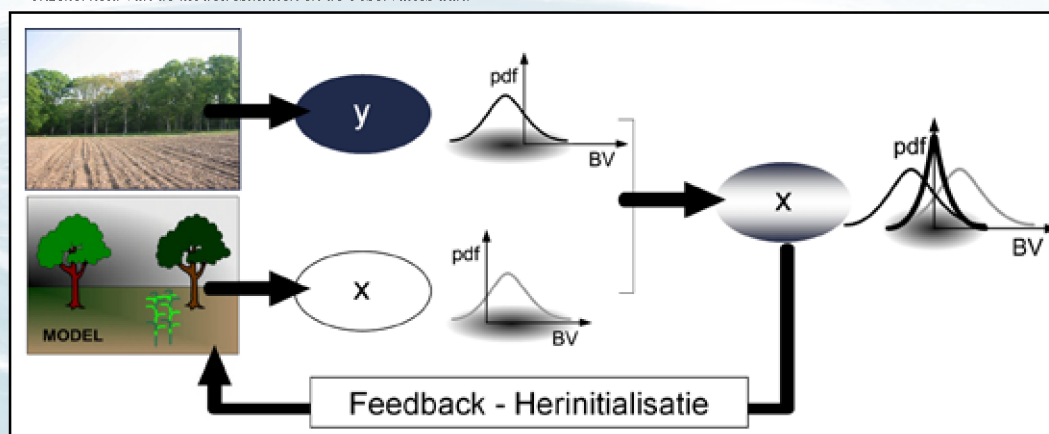
veroorzaakt wordt door het verschil in schaal waarop de meting werd genomen en de schaal waarop hij wordt gebruikt (De Lannoy et al., 2007a). In dit artikel wordt aangenomen dat er geen schalingsprobleem is en dat de onzekerheid enkel te wijten is aan instrumentele beperkingen in de meting van bodemvocht.

Voor de onzekerheid op de modelresultaten is geen voorkennis voorhanden. We weten wel dat zowel de invoergegevens als de parameters en initiële toestand onzeker zijn. Ook de modelstructuur is onzeker, maar dit wordt voor dit artikel buiten beschouwing gelaten. Men kan de totale voorspellingsonzekerheid schatten met ensembles. Dit gebeurt door verschillende parallelle modelsimulaties over eenzelfde periode te genereren met licht verschillende invoergegevens, parameters en initiële toestanden. De spreiding die men dan verkrijgt in de modelresultaten is een maat voor de voorspellingsonzekerheid. Indien men met ensembles werkt om onzekerheden te schatten, gebruikt men voor het optimaal combineren van modelresultaten en observaties de ensemble Kalman filter. Ensembles bieden de meest praktische uitweg, wanneer men wil filteren met niet-lineaire modellen, zoals meestal het geval is in de hydrologie.

Toepassing

Als illustratie wordt hier de bodemvochttoestand geschat in een maïsveld waar het *Optimizing Production Inputs for Economic and Environmental Enhancement* (OPE³) project wordt uitgevoerd. Dit veld ligt in Beltsville, dicht bij Washington, D.C. (VS) en wordt beheerd door de *Beltsville Agricultural Research Center* van het Amerikaanse departement voor landbouw (USDA). De dataset bestaat uit tijdreeksen van bodemvocht gemeten met capaciteitsprobes gedurende 1 jaar (1 mei 2001 t.e.m. 30 april 2002) op 36 punten in de ruimte en op 3 tot 7 dieptes (De Lannoy et al., 2006a). De 36 profielen worden gesimuleerd met het *Community Land Model* (CLM2.0, Dai et al., 2003). De kalibratie en initialisatie van dit model wordt besproken door De Lannoy et al. (2006b). Figuur 2 toont een ensemble aan bodemvocht-

Figuur 1: Assimilatie van observaties (y) in de modeltoestand (x) van bodemvocht (BV). De probabilitydensiteitsfuncties (pdf) geven de onzekerheid van de modelresultaten en de observaties aan.



voorspellingen voor 1 van de 36 profielen op 10 cm diepte. Het ensemble gemiddelde is uiteindelijk de modelschatting van het bodemvocht en de spreiding bepaalt de onzekerheid.

In de eerste kolom van tabel 1 wordt het gesimuleerde (ensemble gemiddelde, Ens gemid) bodemvocht over het volledige modelprofiel getoond, als een gemiddelde over alle 36 gemodelleerde profielen, samen met de ruimtelijke spreiding op deze waarde. De Root Mean Square Error (RMSE) geeft de gemiddelde afwijking t.o.v. de observaties weer in een validatieperiode van 2 Oktober 2001 t.e.m. 30 April 2002. Bodemvocht is een toestandsvariabele en fluxen zoals bvb. evapotranspiratie en afvoer zijn ervan afhankelijk. Daarom worden ook de gemiddelde waarden voor de evapotranspiratie en drainage hier weergegeven.

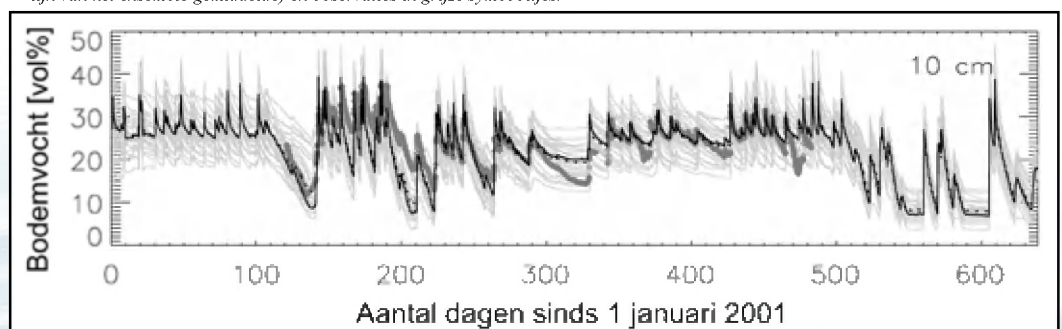
Vervolgens worden in de periode van 2 oktober 2001 t.e.m. 19 maart 2002 2-wekelijks alle beschikbare observaties in de profielen geassimileerd met een ensemble Kalman filter en worden de resultaten opnieuw gevalideerd over de validatieperiode van 2 Oktober 2001 t.e.m. 30 April 2002. Het is duidelijk dat de assimilatie de RMSE doet dalen. Het gemiddelde bodemvochtprofiel was onderschat. Het valt ook direct op dat de drainage een onrealistisch grote waarde aanneemt. Dit wordt veroorzaakt door het feit dat we via assimilatie het model dwingen een hoger bodemvocht te simuleren. Het model wil deze verhoogde waarde echter niet aanvaarden en verwijdt het toegevoegde water opnieuw door de drainage te verhogen. Het probleem zit dus in systematische fouten in de modelresultaten voor bodemvocht: het model simuleert bij te droge bodemvochtwaarden toch realistische afvoer en evapotranspiratie.

Systematische fouten

In de hydrologie en in de meeste aardwetenschappen is data-assimilatie tegenwoordig een standaardmethode geworden om simulaties bij te sturen. Helaas worden de standaardmethoden ook heel dikwijls onterecht of voor het verkeerde doel toegepast. Men vergeet heel vaak dat traditionele assimilatietechnieken niet ontworpen zijn om systematische fouten te corrigeren, maar slechts dienen om willekeurige (random) fouten te filteren en zo te voorkomen dat de voorspellingsfout zou cumuleren. Het schatten van systematische fouten in observaties en modelvoorspellingen krijgt tegenwoordig toenemende aandacht (De Lannoy et al., 2007b).

Friedland (1969) ontwikkelde een methode die een optimale schatting van de toestand met de Kalman filter mogelijk maakt in aanwezigheid van systematische voorspellingsfouten. Deze methode (EnBKF1) bestaat uit 2 parallelle filters: één voor de toestandschatter en één voor de systematische fout. De schatting van de fout kan dan gebruikt worden om de toestandschatter te corrigeren. In deze originele methode stuurt men de toestandschatter zonder foutcorrectie terug in het model, i.e. de filter voor de toestandschatter is perfect identiek aan de Kalman filter zonder systematische foutschatting. Het verschil zit enkel in de output (dus voor verwerking buiten het model) van de toestand: deze wordt gecorrigeerd voor de systematische fout. Dit wordt geïllustreerd in figuur 3. Het is daarom logisch dat in tabel 1 de resultaten voor bodemvocht (toestand) met deze methode EnBKF1 verbeterd worden t.o.v. de gewone EnKF (de RMSE daalt van 5 vol% naar 1.9 vol%), terwijl het probleem van overschatten drainage aanwezig blijft, omdat de toestandfilter dezelfde blijft als in de gewone EnKF.

Figuur 2: Ensemble van modelvoorspellingen (lichtgrijze 'spaghetti-plot') voor bodemvocht op 10 cm diepte op 1 locatie in het OPE3 veld, met aanduiding van het ensemble gemiddelde (volle zwarte lijn), gecalibreerde deterministische simulatie (streepjeslijn, dikwijls achter de lijn van het ensemble gemiddelde) en observaties in grijze symbooltjes.



Tabel 1: Profiel bodemvocht, evapotranspiratie (ET) en drainage, gemiddeld over 36 profielen in het OPE3 veld en met aanduiding van de ruimtelijke standaardafwijking. De waarden zijn gemiddeld over de validatieperiode van 2 oktober 2001 t.e.m. 30 april 2002. Ens gemid verwijst naar de ensemble gemiddelde modelrun zonder assimilatie, terwijl alle andere scenario's verwijzen naar simulaties met 2-wekelijkse assimilatie van alle beschikbare bodemvochtobservaties over het profiel in de periode van 2 oktober 2001 t.e.m. 19 maart 2002.

	Ens gemid	EnKF	EnBKF0	EnBKF1	EnBKF3
Profiel gem. BV [vol%]	16.5 ± 2.8	17.9 ± 3.6	19.2 ± 5.1	19.7 ± 5.3	18.2 ± 3.8
ET [mm]	67 ± 11	67 ± 11	67 ± 11	67 ± 11	67 ± 11
Drainage [mm]	109 ± 32	427 ± 393	109 ± 32	427 ± 393	571 ± 526
RMSE voor BV [vol%]	6.7 ± 3.5	5.0 ± 2.7	2.0 ± 0.7	1.9 ± 0.7	4.8 ± 2.7

Figuur 3 toont ook nog 2 alternatieve methoden om de systematische fout te schatten en de toestand te corrigeren. Men zou kunnen verwachten dat men met een gecorrigeerde modeltoestand verbeterde voorspellingen kan doen. Met methode EnBKF3 neemt men de correctie voor de systematische fout op in het model en het model wordt daarmee geherinitialiseerd. Uit tabel 1 is duidelijk dat op deze manier de bodemvochtschatting wel wordt verbeterd, maar dat de afhankelijke drainage enorm overschat wordt. Uiteraard ligt dit opnieuw aan de aard van de systematische modelfout: het model weigert verhoogde bodemvochtwaarden te simuleren en probeert alle bijgevoegde water te verwijderen via drainage. De evapotranspiratie wordt nagenoeg niet beïnvloed, omdat de validatieperiode in de winter ligt.

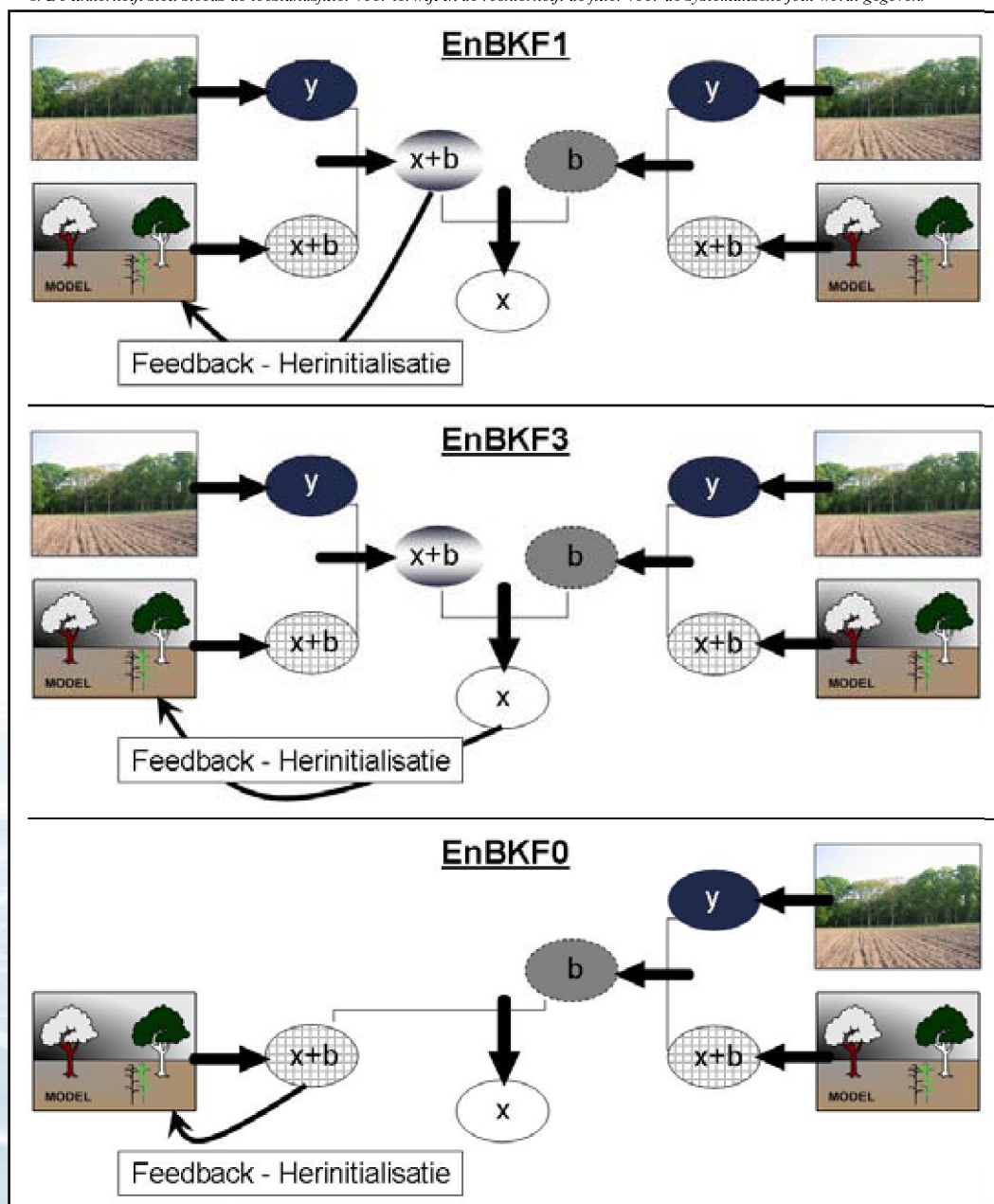
De beste methode is voor deze studie de methode

EnBKF0 (dit zal voor iedere modelleerstudie anders zijn!). Deze methode stuurt de toestand helemaal niet bij en schat enkel de systematische voorspellingsfout. De modelresultaten worden dan gecorrigeerd voor deze foutschatting buiten het model. Aangezien in deze studie realistische evapotranspiratie en drainage berekend werden bij een foutief bodemvocht, is het hier best de toestand onaangetaast te laten binnen de modelrun en de bodemvochtresultaten enkel achteraf te corrigeren. Tabel 1 toont dat de RMSE voor EnBKF0 zeer laag is en de bodemvochtafhankelijke fluxen realistisch blijven.

Besluit

Het is niet vanzelfsprekend om hydrologische processen precies en nauwkeurig te simuleren met een model. Het is ook onmogelijk om overal

Figuur 3: Verschillende methoden voor gecombineerde toestand- en systematische foutschatting. De systematische fout is aangegeven door b . De linkerhelft stelt steeds de toestandsfilter voor terwijl in de rechterhelft de filter voor de systematische fout wordt gegeven.



en altijd goede metingen van verschillende variabelen te bekomen. Nochtans moet een degelijk waterbeheer gesteund zijn op een goede kennis van de toestand en de processen in een hydrologisch systeem. Het optimaal combineren van observaties en modelresultaten met een statistische filter is een manier om schattingen van variabelen betrouwbaarder te maken. Een optimale combinatie gebeurt op basis van een afweging van de onzekerheden op verschillende informatiebronnen (observaties, model). De onzekerheid kan benaderend bepaald worden met ensembles.

Helaas steunen nagenoeg alle klassieke assimilatiemethoden op de veronderstelling dat voorspellingen en observaties vrij zijn van systematische fouten, terwijl aan deze voorwaarde in praktijk zelden voldaan is. In dit artikel wordt getoond hoe via dubbele filters de systematische fout kan geschat worden en hoe de toestand daarvoor kan gecorrigeerd worden. Dit wordt geïllustreerd voor bodemvocht in een klein maïsveld in Greenbelt, Maryland, VS. Daarbij wordt het effect van bodemvochtassimilatie op afhankelijke fluxen, zoals evapotranspiratie en drainage besproken. Het is duidelijk dat assimilatiesystemen veel baat hebben bij het schatten en corrigeren voor systematische fouten en dat de keuze van de exacte methode voor correctie afhangt van het type voorspellingsfout. Het wordt daarom aangeraden dat toekomstige modelleerstudies voor waterbeheer hiermee rekening houden.

Dankwoord

De auteurs danken de BARC-ARS (USDA) om de data voor het OPE³ veld ter beschikking te stellen. Dit onderzoek werd sterk gesteund door interessante discussies en bijdragen van Paul Houser (CREW/GMU, USA) en Rolf Reichle (NASA/GSFC, USA).

Referenties

Dai, Y., X. Zeng, R. Dickinson, I. Baker, G. B. Bonan, M. G. Bosilovich, A. Scott Denning, P. A. Dimeyer, P. R. Houser, G. Niu, K. W. Oleson, C. A. Schlosser, and Z.-L. Yang (2003). The Common Land Model, Bulletin of the American Meteorological Society, 84, 1013–1023.

De Lannoy, G.J.M., Gish T., Verhoest, N.E.C., Houser, P.R., Van Meirvenne, M., De Troch, F.P. (2006a). Spatial and temporal characteristics of soil moisture in an intensively monitored agricultural field (OPE³). Journal of Hydrology, 331(3-4), 719-730, doi:10.1016/j.jhydrol.2006.06.016.

De Lannoy, G.J.M., Houser, P.R., Pauwels, V.R.N., Verhoest, N.E.C. (2006b). Assessment of model uncertainty for soil moisture through ensemble verification. Journal of Geophysical Research, 111, D10101, doi:10.1029/2005JD006367.

De Lannoy, G.J.M., Houser, P.R., Verhoest, N.E.C., Pauwels, V.R.N., Gish, T. (2007a). Upscaling of point soil moisture measurements to field averages at the OPE³ test site. Journal of Hydrology, 343(1-2), 1-11, doi: 10.1018/j.jhydrol.2007.06.004.

De Lannoy, G.J.M., Reichle, R.H., Houser, P.R., Pauwels, V.R.N., Verhoest, N.E.C. (2007b). Correcting for forecast bias in soil moisture assimilation with the ensemble Kalman filter. Water Resources Research, 43, W09410, doi: 10.1029/2006WR005444.

Friedland, B. (1969). Treatment of bias in recursive filtering, IEEE Transactions on Automatic Control, AC-14, 359–367.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems, Transactions of the ASME - Journal of Basic Engineering, 82, 35–45.

*G. J.M. De Lannoy,
V. R.N. Pauwels en
N. E.C. Verhoest*

*Corresponderende auteur:
dr. ir. G. De Lannoy
Laboratorium voor Hydrologie en Waterbeheer
Universiteit Gent
Coupure links 653
9000 Gent
e-mail: Gabrielle.DeLannoy@UGent.be
Tel. +32 – (0)9 – 264.61.40
Fax. +32 – (0)9 – 264.62.36*