

Inhoud | Contents

Wiskunde en planeet Aarde

Onder redactie van Daan Crommelin en Henk Schuttelaars

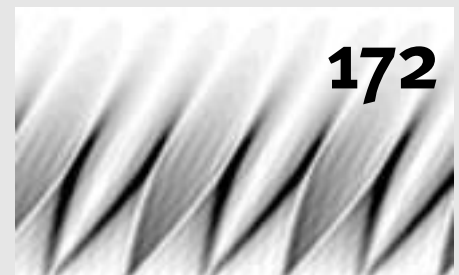
Artikelen | Features

- 164** Het onvoorspelbare venijn van de staart
Alef Sterk, Renato Vitolo, Henk Broer
- 168** Voorspellen met behulp van data-assimilatie
Arnold Heemink, Peter Jan van Leeuwen
- 172** Wiskundige aspecten van interne golven
Theo Gerkema, Leo Maas
- 177** Dynamica van bodempatronen in ondiepe zeeën
Henk Schuttelaars, Huib de Swart
- 184** Watergolven
Brenny van Groesen, Frits van Beckum
- 190** De complexe dynamica van fytoplankton
Arjen Doelman e.a.
- 195** El Niño: processes, dynamics and predictability
Anna von der Heydt, Henk Dijkstra
- 200** Klimaatsverandering, verwoestijning en biljart
Arjen Doelman e.a.
- 207** Shock waves and two phase porous media flows
Cornelis Johannes van Duijn, Iuliu Sorin Pop
- 212** Onweer: een bron van broeikasgassen en anti-materie
Christoph Köhn, Margreet Nool, Ute Ebert
- 217** De wiskunde van reflectieseismiek
Chris Stolk

Rubrieken | Departments

- 155** Redactioneel
Daan Crommelin, Henk Schuttelaars
- 156** Servicepagina
- 157** Agenda
Patrick Hafkenscheid
- 159** Nieuws
Patrick Hafkenscheid
- 182** Het keerpunt van Jan Barkmeijer
Ionica Smeets
- 205** In de verdediging
Geertje Hek
- 221** Problemen

Uitgelicht



Theo Gerkema en Leo Maas laten zien dat golven onder het zeeoppervlak zich heel anders gedragen dan golven aan het vrije oppervlak.



Arjen Doelman en anderen beschrijven conceptuele modellen voor het gedrag van vegetatiepatronen.

Colofon

Het Nieuw Archief voor Wiskunde is een uitgave van het Koninklijk Wiskundig Genootschap en verschijnt vier maal per jaar. Het tijdschrift richt zich op een ieder die zich beroepsmatig met wiskunde bezighoudt, als academisch of industrieel onderzoeker, student, leraar, journalist of beleidsmaker. Het stelt zich ten doel te berichten over ontwikkelingen in de wiskunde in het algemeen en in de Nederlandse wiskunde in het bijzonder.

ISSN 0028-9825

Adres

Nieuw Archief voor Wiskunde
Centrum Wiskunde & Informatica
Postbus 94079
1090 GB Amsterdam
tel. 020-5924288
redactie@nieuwarchief.nl
www.nieuwarchief.nl

Hoofredactie

Barry Koren (b.koren@tue.nl)
Jan van Neerven (j.m.a.m.vanneerven@tudelft.nl)

Eindredactie/Bladmanager

Fred Drissen (fred@nieuwarchief.nl)

Redactie

Patrick Hafkenscheid (p.hafkenscheid@vu.nl)
Nellie Verhoef (n.c.verhoef@utwente.nl)
Steven Wepster (s.a.wepster@uu.nl)

Bureau redactie

Gianni van den Braak
Lætitia Diemer

Medewerkers

Johan Bosman (IEM Essen), Hans Cuyper (TU/e), Gabriele Dalla Torre (UL), Christophe Debry (KU Leuven), Geertje Hek (UvA), Jinbi Jin (UL), Teun Koetsier (VU), Derk Pik (JSG Maimonides), Sjoerd Rienstra (TU/e), Hans Sterk (TU/e), Marco Streng (VU), Wouter Zomervrucht (UL)

Illustraties

Ryu Tajiri, Amsterdam

Advertenties

advertenties@nieuwarchief.nl

Abonnementen/Subscriptions

Abonnementenservice NAW
Postbus 2238, 5600 CE Eindhoven
admin@wiskgenoot.nl, tel. 088-2265256

Abonnementsperiode/Subscription period

Een abonnement gaat in op de dag dat uw aanmelding ontvangen wordt. De opzegtermijn is twee maanden vóór het verstrijken van de abonnementsperiode.

The subscription starts on the day that your application is received. The term of notice is two months before the end of the current subscription period.

Exchange subscriptions

Koninklijk Wiskundig Genootschap Library
Centrum voor Wiskunde en Informatica
Postbus 94079, NL-1090 GB Amsterdam
KWG-exchange@cw.nl, fax +31 20 5924026

Vormgeving

Kitty Molenaar (ontwerp)
Susanne Laws (omslag, begeleiding binnenwerk)
Programmatuur (CONTEX)
Hans Hagen, PRAGMA-ADE, Hasselt (Overijssel)
Druk
Drukkerij Ten Brink, Meppel

Koninklijk Wiskundig Genootschap

Het Koninklijk Wiskundig Genootschap (KWG) is een landelijke vereniging van beoefenaren van de wiskunde. In 1778 opgericht onder de zinspreuk: 'Een onvermoeide arbeid komt alles te boven' is het 's werelds oudste nationale wiskundegenootschap. Leden ontvangen het Nieuw Archief voor Wiskunde als onderdeel van hun lidmaatschap.

De contributie voor gewone leden bedraagt €80 per jaar. Voor gepensioneerden en werklozen €40. Voor studenten aio's en oio's van een Nederlandse universiteit of hbo-opleiding €30. Studenten en pas afgestudeerden kunnen een jaar lang gratis lid worden. Voor leden van de wiskundige vakverenigingen VVS, NVvW en NVORWO geldt het gereduceerd tarief van €60. Leden die het Nieuw Archief voor Wiskunde niet willen ontvangen, betalen slechts €40 contributie. Losse nummers zijn te bestellen voor €20.

Members of SIAM, the *Société Mathématique de France*, the *Gesellschaft für Angewandte Mathematik und Mechanik*, the *Deutsche Mathematiker-Vereinigung*, and of the *American, Australian, Belgian, Indian, London, Spanish, and South-African Mathematical Societies* living outside of the Netherlands pay €60 instead of the full membership fee of €80 a year. Conversely, these foreign societies, as well as the VVS and the NVORWO, have reduced membership fees for KWG-members. A postage charge of €10 is added for members living abroad but in Europe, and a charge of €12,50 for members living outside of Europe. Instituten in Nederland en buitenland kunnen zich abonneren op het Nieuw Archief voor Wiskunde voor €100 (Nederland), €110 (Europa) of €112,50 (buiten Europa).

Website: www.wiskgenoot.nl

Ledenadministratie

Abonnementenservice NAW
Postbus 2238, 5600 CE Eindhoven
admin@wiskgenoot.nl, tel. 088-2265256

Bestuur (wiskgenoot@wiskgenoot.nl)

André Ran (VU, voorzitter), Joke Blom (CWI, secretaris), Onno van Gaans (UL, penningmeester), Geurt Jongbloed (TUD, ondervoorzitter), Jenneke Krüger (SLO), Herman te Riele (CWI), Arjen Sevenster, Gert Vegter (RUG), Mark Veraar (TUD)

Informatie voor auteurs

Kopij

Het Nieuw Archief voor Wiskunde is een geredigeerd tijdschrift. Kopij dient elektronisch te worden aangeboden aan de hoofdredactie. Uitgebreide informatie en auteursinstructies kunt u vinden op www.nieuwarchief.nl.

Reprints

Auteurs ontvangen een elektronische reprint in pdf-formaat.

Volgende nummers

nummer	verschijningsdatum	deadline kopij
4	02-12-2013	verstreken
1	03-03-2014	01-11-2013
2	02-06-2014	01-02-2014
3	01-09-2014	01-05-2014

Instituutsleden

De publicaties van het Koninklijk Wiskundig Genootschap worden mede mogelijk gemaakt door de bijdragen van de volgende instituten.



Centrum Wiskunde
& Informatica



Rijksuniversiteit Groningen



Universiteit van Amsterdam



Universiteit Leiden



Vrije Universiteit



Universiteit Utrecht



Technische Universiteit
Eindhoven



Eurandom



Technische Universiteit Delft



Universiteit Twente



Radboud Universiteit
Nijmegen

Fishe

Freudenthal Instituut

Op het omslag

Dit jaar wordt wereldwijd speciale aandacht besteed aan wiskunde die te maken heeft met onze planeet. In dit themanummer sluiten wij aan bij dit initiatief genaamd 'Mathematics of Planet Earth'. (Foto: Stereographic Projection, Aurélien Alvarez, Étienne Ghys, Jos Leys.)

Wiskunde en planeet Aarde

Gesteund door UNESCO wordt er dit jaar door wetenschappelijke organisaties, universiteiten en onderzoeksinstituten wereldwijd speciale aandacht besteed aan wiskunde die, in de ruimste zin, te maken heeft met onze planeet. Dit initiatief heeft de sprekende naam ‘Mathematics of Planet Earth’ gekregen, en wil de aandacht vestigen op de grote verscheidenheid aan dynamische processen op onze planeet. Denk hierbij aan geofysische processen (in de aardmantel, oceanen, en atmosfeer), biologische processen, en menselijke processen (zoals landbouw, transport, energieomzetting).

Wiskunde speelt een centrale rol in de wetenschappelijke inspanning om deze processen te begrijpen. Zonder dit begrip kunnen we de uitdagingen die met bovengenoemde processen samenhangen niet optimaal het hoofd bieden, terwijl recente ontwikkelingen juist laten zien dat dit voor onze planeet steeds urgenter wordt. De competitie om grondstoffen neemt toe ten gevolge van de bevolkingsgroei, ecologische en economische belangen staan steeds vaker op gespannen voet met elkaar, en klimaatverandering en haar mogelijke gevolgen vormen een steeds grotere zorg. ‘Mathematics of Planet Earth’ (MPE) is dan ook breed geïnterpreteerd en omvat vier hoofdthema’s:

- *A planet to discover*: oceans, meteorology and climate, mantle processes, natural resources, solar system.
- *A planet supporting life*: ecology, biodiversity, evolution.
- *A planet organized by humans*: political, economic, social and financial systems, organization of transport and communication networks, management of resources, energy.
- *A planet at risk*: climate change, sustainable development, epidemics, invasive species, natural disasters.

Op de website www.mpe2013.org is uitgebreide informatie te vinden over dit internationale initiatief. Onder de activiteiten die georganiseerd worden in het kader van MPE2013 bevinden zich workshops, speciale conferentiesessies en publiekslezingen.

In aansluiting op het MPE-initiatief is dit nummer van het *Nieuw Archief voor Wiskunde* gewijd aan ‘Wiskunde en planeet Aarde’. Gezien de zeer brede insteek van MPE hebben we er voor gekozen om de onderwerpen voor dit themanummer wat meer toe te spitsen, door ons hoofdzakelijk te beperken tot het eerste

thema ‘A planet to discover’, met een klein uitstapje naar ‘A planet supporting life’. De bijdragen in dit nummer hebben betrekking op oceanen, atmosfeer en klimaat, op seismiek, hydrologie en ecologie. De onderwerpen lopen uiteen van fytoplankton tot onweer en van El Niño tot zandbanken.

De wiskunde die aan de orde komt in de verschillende bijdragen is erg gevarieerd. Dynamische systemen en niet-lineaire partiële differentiaalvergelijkingen lopen als een rode draad door een groot deel van dit themanummer, maar de verschillende auteurs gaan daar op uiteenlopende manieren mee om. Aan de orde komen onder andere stabiliteitsanalyse, Hamiltoniaanse dynamica, numerieke methoden en Monte Carlo-simulaties, filtertheorie, statistiek van extreme waarden, inverse problemen en de theorie van hyperbolische behoudswetten.

Naast de elf artikelen in dit nummer staan ook de rubrieken ‘In de verdediging’ en ‘Het keerpunt’ in het teken van ‘Wiskunde en planeet Aarde’. In ‘In de verdediging’ staat een proefschrift centraal dat handelt over numerieke methoden voor de Navier–Stokes-vergelijkingen die gebruikt kunnen worden voor simulaties van windmolenparken. En in ‘Het keerpunt’ wordt een wiskundige geïnterviewd die vijftientig jaar geleden promoveerde bij Floris Takens op een proefschrift over cirkelafbeeldingen en sinds zijn promotie bij het KNMI werkt.

Alles bij elkaar biedt dit themanummer een rijke en gevarieerde staalkaart van overwegend Nederlands onderzoek waarin wiskunde en aardwetenschappen bijeen komen. Waarmee niet gezegd is dat hiermee een uitputtend overzicht is gegeven van onderzoek op dit snijvlak: het is onvermijdelijk dat er onderzoekers zijn wier werk uitstekend binnen het MPE thema past maar hier desalniettemin niet aan bod komen. We hopen dat er in toekomstige NAW-nummers ruimte zal zijn voor verdere artikelen rondom dit thema. Ook na 2013 zal planeet Aarde een rijke bron blijven van inspiratie en toepassingen voor wiskundig onderzoek. 

Daan Crommelin en Henk Schuttelaars, gastredacteuren

Research Group Computational Dynamics, CWI, resp. Faculteit EWI, TU Delft

Koninklijk Wiskundig Genootschap

❑ **Secretariaat**

Joke Blom, CWI
Postbus 94079, 1090 GB Amsterdam
wiskgenoot@wiskgenoot.nl
www.wiskgenoot.nl

❑ **Wintersymposium 2014**

Het traditionele KWG Wintersymposium voor docenten wiskunde in het voortgezet onderwijs en andere belangstellenden zal worden gehouden op zaterdag 11 januari 2014 in Utrecht. Doel van het symposium is contacten tussen wiskundedocenten en beroepsmatige beoefenaars van wiskunde te stimuleren, om zo leerlingen een beter beeld te geven van de rijkdom aan mogelijkheden van wiskunde. In 2014 is het thema 'Mathematics of Planet Earth'. Het vormt daarmee voor Nederlandse docenten wiskunde een afsluiting van dit wereldwijde thema voor 2013. De sprekers zijn: Ferdinand Verhulst (Universiteit Utrecht), Leo Maas (NIOZ, Universiteit Utrecht), Geert Lenderink (KNMI) en Carel Eijgenraam (Centraal Planbureau). www.wiskgenoot.nl/watbiedt

❑ **Proceedings KNAW digitaal**

De Proceedings KNAW 1895–1950 (*Indagationes Mathematicae*) zijn nu digitaal beschikbaar. www.dwc.knaw.nl/?s=proceedings

Platform Wiskunde Nederland

❑ **Bureau PWN**

Science Park 123, L013, 1098 XG Amsterdam
bureau@platformwiskunde.nl
pr@platformwiskunde.nl (publiciteit)
Tel: 020-5924006/06-51892525
www.platformwiskunde.nl

❑ **Publieksevenement in Utrecht**

Op zaterdag 21 december organiseert PWN in het Educatorium te Utrecht een publieksevenement met als thema 'Mathematics of Planet Earth', in het kader van het MPE 2013 jaar.

❑ **PWN Nieuwsbrief**

PWN publiceert eens in de drie maanden een PWN Nieuwsbrief, die op de website te vinden is. Actuele PWN-informatie wordt verder zowel via de website als via Twitter gegeven.

European Mathematical Society

❑ **Science Europe**

László Lovász (Eötvös Loránd Universiteit, Boedapest) en Karl Sigmund (Universiteit van Wenen) zijn benoemd tot leden van de 'Physical, Chemical and Mathematical Sciences Committee of Science Europe', de nieuwe associatie van de 'European Research Funding and Research Performing Organizations'. Meer informatie: www.scienceeurope.org/scientific-committees/Physical-sciences. Lovász en Sigmund zullen bij de Europese Unie de belangen van de wiskunde behartigen bij de toekenning van subsidies voor onderzoek. www.scienceeurope.org

❑ **EMS Newsletter**

Van de drie-maandelijkse EMS Newsletter is de elektronische versie van de meest recente uitgave beschikbaar op de EMS website: www.ems-ph.org/publications.php.

International Mathematical Union

❑ **IMU Newsletter**

De twee-maandelijkse IMU Newsletter is beschikbaar via www.mathunion.org/imu-net.

Recente publicaties van het KWG

❑ **Indagationes Mathematicae** (www.elsevier.com/locate/indag)

Special issue honoring Israel Gohberg; M.A. Kaashoek and H.W. Broer (eds.), Volume 23, issue 4 (gratis toegankelijk op www.sciencedirect.com/science/journal/00193577).

❑ **Epsilon Uitgaven** (www.epsilon-uitgaven.nl)

- 76. *Inleiding in de Statistiek*, F. Bijma, M. Jonker en A. van der Vaart, €32,00, 2013.
- 75. *An Invitation to Functional Analysis*, B. de Pagter en A.C.M. van Rooij, €25,00, 2013.
- 74. *L.E.J. Brouwer - van Mystiek tot Wiskunde*, D. van Dalen, €26,00, 2013.
- 73. *Honderd gevangenen en een gloeilamp*, H. van Ditmarsch en B. Kooij, €20,00, 2013.
- 48. *Wiskunde in Werking deel 1 – vectoren en matrices toegepast*, M. de Gee, €37,00, 2013 (geheel herziene druk).

❑ **Zebra-reeks** (Epsilon Uitgaven)

- 35. *Verborgen boodschappen – een inleiding in de cryptografie*, A. Aaten en C. Kraaikamp, €10,00, 2013.
- 34. *De Ster van de dag gaat op en onder – rekenen aan zonsopkomst en zonsondergang*, A. Goddijn, €10,00, 2012.

Manuscripten voor Epsilon Uitgaven kunt u mailen naar redactie@epsilon-uitgaven.nl.

Agenda

| Upcoming Events

september 2013

2–6 september 2013

16de algemene ontmoeting EWM

16de editie van de conferentie voor *European Women in Mathematics*. Lezingen door en voor vrouwen in de wiskunde.

plaats Hausdorff Center for Mathematics, Bonn

info europeanwomeninmaths.org

3–6 september 2013

Workshop on Stochastics and Neuroscience

Workshop over het gebruik van stochastiek in de neurowetenschap.

plaats Universiteit Utrecht

info eurandom.tue.nl/STAR

14 september 2013

Tweede Ig Nobel Night

Nacht voor onderzoekers en wetenschappers met Nederlandse Ig Nobelwinnaars, en live vertoning van de Ig Nobel Prize-ceremonie.

plaats Stadsgehoorzaal, Leiden

info ignobelnight.nl

15 september 2013

Wat doen satellieten in de ruimte?

Kinderlezing over de verschillende soorten satellieten en hoe ze werken.

plaats NEMO, Amsterdam

info kinderlezingen.nl

20 september 2013

Vernieuwingsimpuls Informatiebijeenkomst 2013

Informatiebijeenkomst voor onderzoekers die een Veni, Vidi of Vici willen aanvragen.

plaats NWO-gebouw, Den Haag

info nwo.nl

23 september 2013

Finale Nederlandse Wiskunde Olympiade

43 onderbouwleerlingen, 45 vierdeklassers en 32 vijfdeklassers zijn door naar de finale van de Nederlandse Wiskunde Olympiade 2013.

plaats TU Eindhoven

info wiskundeolympiade.nl

27 september 2013

Discovery Festival

Uitgaansnacht-met-inhoud vol wetenschap, muziek, kunst en experimenten.

plaats Amsterdam, Eindhoven en Rotterdam

info discoveryfestival.nl

oktober 2013

2–4 oktober 2013

Woudschoten conferentie

38ste editie van de Vlaams/Nederlandse conferentie over Scientific Computing.

plaats Conferentiecentrum Woudschoten, Zeist

info wsc.project.cwi.nl/conferentieN.php

3 oktober 2013

Bèta onder de Dom

Lezingen en werkgroepen voor docenten biologie, natuurkunde, scheikunde, wiskunde en NLT.

plaats Utrecht

info betasteunpunt-utrecht.nl

4 oktober 2013

Junior Wiskunde Olympiade

Junior versie van de wiskunde olympiade, klas 1 tot 3 van het voortgezet onderwijs.

plaats Vrije Universiteit, Amsterdam

info wiskundeolympiade.nl

5 oktober 2013

Dag van de leraar

Internationale dag van de leraar, zet uw onderwijzer in het zonnetje! Op deze dag wordt ook de 'leraar van het jaar' gekozen.

plaats overal

info dagvandeleraar.nl

7–9 oktober 2013

9th European Computer Science Summit

Internationale ontmoeting voor mensen werkend in de informatica.

plaats CWI, Amsterdam

info cwi.nl

Suggesties voor deze agenda zijn welkom.

Redacteur: Patrick Hafkenscheid

agenda@nieuwarchief.nl

 11 oktober 2013

Wiskundetoernooi 2013

 22ste editie van het Wiskundetoernooi, dit jaar in het teken van *Wiskunde en Sport*
plaats Radboud Universiteit Nijmegen

info ru.nl/wiskundetoernooi

14–18 oktober 2013

Workshop Noncommutative Geometry and Particle Physics

Workshop met als doel onderzoekers in de non-commutative meetkunde samen te brengen met wetenschappers in de deeltjesfysica.

plaats Lorentzcenter, Leiden

info lorentzcenter.nl

25–26 oktober 2013

Masterclass 'Diophantische Vergelijkingen'

Masterclass voor 5/6 vwo'ers over diophantische vergelijkingen.

plaats Utrecht

info betasteunpunt-utrecht.nl

november-december 2013

 4–6 november 2013

Workshop YEQT VII

 Workshop voor *Young European Queueing Theorists* heeft als thema 'Scheduling and priorities in queueing systems'.

plaats TU Eindhoven

info eurandom.tue.nl

5 november 2013

Afscheidcollege Odo Diekmann

Afscheidcollege prof.dr. Odo Diekmann, hoogleraar toegepaste analyse aan de Universiteit Utrecht.

plaats Academiegebouw, Utrecht

info math.uu.nl

9 november 2013

Studiedag Jaarvergadering NVvW 2013

Jaarlijkse studiedag en jaarvergadering van de nederlandse vereniging van wiskundeleraren. Thema van dit jaar: 'Krachtig vooruit met wiskunde!'

plaats Ichthus College, Veenendaal

info nvvw.nl

 11–13 november 2013

42ste bijeenkomst Stochastici

Jaarlijkse bijeenkomst in Lunteren voor stochastici uit Nederland.

plaats Lunteren

info eurandom.tue.nl/STAR

15 november 2013

Voorrondes wiskunde A-lympiade

Teamwedstrijd voor leerlingen met wiskunde A in hun pakket.

plaats eigen school

info fisme.science.uu.nl/alympiade

22 november 2013

CWI in Bedrijf 2013

Ontmoeting tussen het CWI en het bedrijfsleven waar het bedrijfsleven vragen kan stellen aan top-onderzoekers.

plaats CWI, Amsterdam

info cwi.nl/events/cwi-in-bedrijf-2013

25–29 november 2013

Workshop Generalizations of Symmetric Spaces

Workshop over symmetrische ruimtes belicht vanuit verschillende gebieden (meetkundig, analytisch, representatie-theoretisch, getal-theoretisch).

plaats Lorentzcenter, Leiden

info lorentzcenter.nl

27–28 november 2013

ICT.Open2013 congres

Het jaarlijkse congres over ICT- en Informatica-onderzoek in Nederland.

plaats Hotel Van der Valk, Eindhoven

info nwo.nl

januari 2014

 11 januari 2014

KWG Wintersymposium 2014

Jaarlijks symposium met dit jaar als thema 'Mathematics of Planet Earth'.

plaats Academiegebouw, Utrecht

info www.wiskgenoot.nl

 16–17 januari 2014

Panamacongres

Jaarlijkse conferentie over de ontwikkeling van reken- en wiskundeonderwijs.

plaats De Leeuwenhorst, Noordwijkerhout

info fisme.science.uu.nl/panama

20–22 januari 2014

13th Winter School on Mathematical Finance

Jaarlijkse winterschool over Mathematical Finance. Met lezingen en minicursussen.

plaats Lunteren

info mathfin.nl

20–31 januari 2014

Workshop Representing Streams II

 Workshop over *streams*; oneindige rijtjes symbolen die voorkomen in combinatoriek, getaltheorie en kansrekening.

plaats Lorentz Center, Leiden

info lorentzcenter.nl

Nieuws

| News

Deze rubriek is een kroniek van wiskundige activiteiten in Nederland. Toekomstige activiteiten worden aangekondigd en van voorbije activiteiten wordt verslag gedaan. Wilt u uw aankondiging of verslag in deze rubriek geplaatst zien? Stuur ons dan uw bijdrage van ± 350 woorden, zo mogelijk met illustratie. De redactie behoudt zich het recht voor berichten te weigeren of in te korten.

Redacteur: Patrick Hafkenscheid
 nieuws@nieuwarchief.nl

Planten goed in wiskunde

Planten werden de laatste weken door de wetenschap in het zonnetje gezet. Allereerst werd een mechanisme uit de kwantummechanica gevonden in planten. Wetenschappers hebben dit fenomeen beschreven in een artikel dat gepubliceerd is in *Science*. In het artikel wordt uitgelegd dat in de eerste stappen van fotosynthese de fotonen met enorme snelheid worden geabsorbeerd en worden getransporteerd naar reactiekernen waar de fotonen worden omgezet in chemische energie. Het uitkiezen van de beste route door de fotonen maakt hier gebruik van de kwantumeffecten van de deeltjes.

Een ander fenomeen dat plaatsvindt in planten is een handig trucje dat de planten uitvoeren om 's nachts te overleven. Overdag gebruiken planten zonlicht om koolstofdioxide om te zetten in suikers en zetmeel, deze stoffen gebruiken ze om niet te verhongeren. In de nacht, als er geen zonlicht is, gebruiken de planten hun zetmeelvoorraad om te overleven. Wetenschappers van het John Innes Centrum hebben nu laten zien dat planten precies berekenen hoeveel zetmeel ze nodig hebben door de nacht heen. Als de nacht is gevallen schat de plant hoe lang de nacht nog gaat duren en past de snelheid waarmee de zetmeel wordt opgebruikt zo aan dat er na de nacht 95 procent van de totale zetmeel is opgebruikt. Er vindt dus ook een zekere vorm van optimalisatie plaats. Een belangrijk gevolg hiervan is dat planten op een indirecte manier kunnen delen!

sciencemag.org

3D-model van een kamer via geluid

Blinde mensen kunnen soms op opmerkelijke wijze de kamer waarin ze zijn verkennen. Dit doen ze volledig met behulp van geluid. Vleermuizen en dolfinen gebruiken dezelfde techniek om te bepalen waar ze heen moeten. Onderzoekers van de École Polytechnique Fédérale de Lausanne hebben nu een computeralgoritme ontwikkeld dat een hele kamer kan beschrijven met behulp van maar(!) vier microfoons. Voor een simpele convexe kamer kan dit worden gedaan met een nauwkeurigheid van enkele millimeters. Opmerkelijk is dat de plaatsing van de microfoons niet uitmaakt voor het resultaat. Het algoritme maakt gebruik van zowel het directe als indirecte (weerkaatsing via een muur) geluid. Na de verwerking van de data is de uitvoer een signatuur van de muren. Het algoritme is getest met een kamer waarin een verplaatsbare muur stond en de resultaten waren erg precies.

De makers van het algoritme hebben ook al enkele toepassingen bedacht. Zo zouden architecten het programma kunnen gebruiken om concerthallen op een zodanige manier te ontwerpen dat het een specifiek geluid geeft. Een spannendere toepassing zou kunnen zijn dat als iemand aan het lopen is terwijl hij aan de telefoon is, het algoritme de kamer kan beschrijven. Dit zou informatie kunnen geven over waar de persoon zich bevindt.

eurekalert.org

Middeleeuwse tekst geen onzin

Het mysterieuze Voynich-manuscript, genoemd naar de boekhandelaar Wilfrid Voynich, is een manuscript dat in een onbekend handschrift is geschreven. Het manuscript dateert waarschijnlijk van rond de 15de eeuw, mogelijk uit het noorden van Italië. Het manuscript heeft al jaren cryptografen en handschriftkenners beziggehouden omdat niemand wist wat de tekens die gebruikt werden betekenden. Nu hebben wetenschappers van de Universiteit van Manchester met behulp van statistische computermodellen beredeneerd dat het wel degelijk om een taal gaat. Veel wetenschappers hebben het Voynich-manuscript nooit

serieus genomen, gezien de taal nergens anders voorkomt. Ze denken dat het een hele uitgebreide grap is geweest uit de middeleeuwen.

De computermodellen onderzoeken patronen van combinaties van woorden die in het boek voorkomen. Als een woord steeds in eenzelfde groep andere woorden verschijnt, betekent dit vaak dat er wel degelijk een betekenis aan vast zit. De statistieken die de wetenschappers hebben laten duidelijk dit soort patronen zien. Volgens de onderzoekers is het onmogelijk dat de mensen uit de middeleeuwen dit soort kennis hadden over taal. Daaruit concluderen ze dus dat het niet gaat om een grap.

Het Voynich-manuscript bevat behalve veel vreemde tekens ook afbeeldingen. Vooral veel illustraties van planten en diagrammen geven de indruk dat het om recepten of bepaalde berekeningen gaat. *plosone.org*



Gevecht om kleinste priemkloof

Er is een ware (vriendschappelijke) competitie aan de gang in de wereld van de getaltheorie. De slag gaat om priemkloven. Een priemkloof is een natuurlijk getal H zodat er oneindig veel priemgetallen zijn met verschil H . Of, symbolisch, $\liminf_{n \rightarrow \infty} p_{n+1} - p_n \leq H$, waarbij p_n het n -de priemgetal is. Op 14 mei 2013 heeft *Yitang Zhang* bewezen dat er in ieder geval een eindige priemkloof bestaat. Hij bewijst dat de kleinste priemkloof kleiner is dan $7 \cdot 10^7$. Vanaf dat moment barstte de strijd om de kleinste bovengrens los. De eerste twee weken waren het vooral Scott Morrison en Terrence Tao die het bewijs probeerden te optimaliseren. Ze brachten de bovengrens terug tot ongeveer $40 \cdot 10^7$. Begin juni waren de eerste optimalisaties onder de 10^7 . Momenteel is de laatste gecontroleerde bovengrens 12006.

De hoop is dat uiteindelijk kan worden bewezen dat de kleinste priemkloof 2 is, dit zou betekenen dat er oneindig veel priem tweelingen zijn, een antwoord op een oud vermoeden.

De actuele race voor de kleinste bovengrens wordt bijgehouden op *michaelnielsen.org* onder projectnaam 'Polymath8'. *Annals of Mathematics*

Freudenthal Instituut ondertekent JetNet-convenant

Recentelijk heeft het Freudenthal Instituut een convenant ondertekend dat samenwerking tussen bedrijven, onderwijs en overheid moet verbeteren. Het doel van dit zogenaamde JetNet, het Jongeren en Technologie Netwerk Nederland, is havo- en vwo-leerlingen een reëel beeld te geven van bèta en technologie en stimuleren hier verder in te leren.

JetNet werkt nauw samen met technologische bedrijven en basisscholen om een omgeving te creëren waarin de interesse en talenten van de leerlingen optimaal kunnen worden ontwikkeld. Zo zou je kunnen denken aan: gastlessen, bedrijfsbezoeken of wedstrijden.

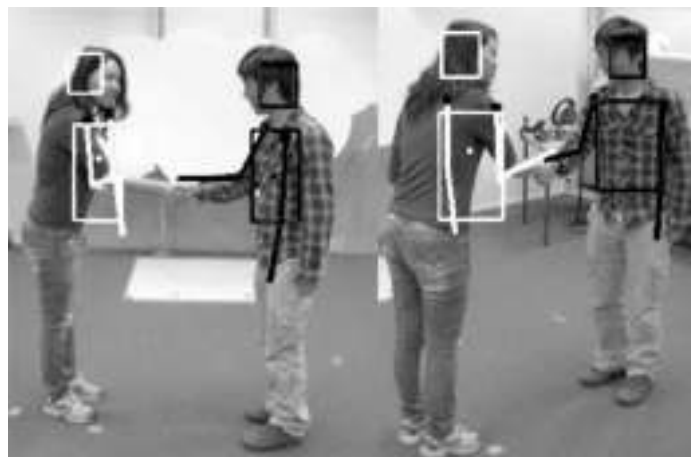
Voorbeelden van projecten die al lopen zijn de 'Techno Challenge', een initiatief van staalbedrijf Tata Steel waarin men op een goede manier lessen organiseert die door een combinatie van didactiekdeskundigen en vakdeskundigen zijn ontwikkeld. Verder heeft Tata Steel in het 'Training Centre' een lab gemaakt voor leerlingen uit het basisonderwijs met onder andere een 'Ontdekplek' waar scholieren spelenderwijs kunnen leren. Voor meer informatie over dit project bekijk *www.technochallenge.nl*. *fisme.science.uu.nl*

Handen schudden of slaan?

Kan een computer op een consistente manier bepalen of twee personen elkaar de hand gaan schudden of elkaar gaan slaan? Feifei Huo van de TU Delft heeft dit onderzocht in haar promotieonderzoek. Voor computers is dit vanouds een ingewikkeld probleem, vooral omdat er veel berekeningen in real time moeten worden uitgevoerd. Feifei heeft met name onderzoek gedaan naar hoe de algoritmes die menselijk bewegen analyseren kunnen worden versimpeld.

De eerste stap in Feifei's algoritmes is het vinden van het hoofd en het torso van de personen in beeld, gevolgd door het vinden van de armen. Hierna worden de bewegingen van de personen in beeld geanalyseerd. Door middel van een leerproces leerde de computer het verschil tussen een aantal handelingen, waaronder handen schudden, jezelf voorstellen, wijzen, slaan, zwaaien of duwen.

Applicaties van deze technieken zijn vrij divers, zo zou je kunnen denken aan de aansturing van spelletjes waarbij verschillende acties met het lichaam verschillende acties in het spel besturen. Ook kan het in de gezondheidszorg worden gebruikt om ouderen thuis in de gaten te houden. *ewi.tudelft.nl*



Google-leergang onderwijsvernieuwing wiskunde

Met de komst van de nieuwe examenprogramma's voor wiskunde in het voortgezet onderwijs in 2015 zullen er flink wat dingen veranderen. Google heeft daarom Platform Bèta Techniek de financiële middelen gegeven voor het opstarten van een leergang om wiskundesecties te helpen aan de eisen van de examenprogramma's te voldoen en zo nodig hun eigen invulling er aan te geven.

Het voornaamste doel van de leergang is om de wiskundesecties samen te laten komen en gezamenlijk kennis te laten maken met de examenprogramma's, maar ook met hoe wiskunde in het bedrijfsleven en de wetenschap een plek heeft. Om dit voor elkaar te krijgen zullen behalve deskundigen van SLO en het Freudenthal Instituut ook experts

uit het bedrijfsleven en uit de wetenschap de materie behandelen en verrijken. Het traject kan worden opgedeeld in vier onderdelen,

- de praktische toepassing van wiskunde,
- de nieuwe examenprogramma's voor wiskunde,
- inhoudelijke verdieping, en
- ontwikkeling van lesmateriaal/modules.

De bijeenkomsten zullen per regio gaan plaatsvinden, per regio kunnen 25 docenten zich inschrijven. Inhoudelijk zal de stof het volgende omvatten:

- Bijeenkomst 1: nieuwe examenprogramma's, hoofdonderwerp wiskundige denkactiviteiten,
- werken aan lesmateriaal.
- Bijeenkomst 2: inleiding, hoofdonderwerp grote datasets, werken aan lesmateriaal.
- Bijeenkomst 3: hoofdonderwerp analytische meetkunde, werken aan lesmateriaal.
- Bijeenkomst 4: hoofdonderwerp authentieke toepassingen, werken aan lesmateriaal.
- Terugkomdag: presentatie lesmateriaal.

Voor de hoofdonderwerpen worden steeds twee personen uitgenodigd: iemand uit de wetenschap en iemand uit het bedrijfsleven.

leergangwiskunde.nl

Vijf Nederlandse medailles op IMO 2013

Nederland heeft bij de Internationale Wiskunde Olympiade in Colombia een prestatie van formaat neergezet. Teamlid Jeroen Winkel (16) wist tot de beste 10 procent van de wereld door te dringen. Hij en Jeroen Huijben (17) ontvingen op zaterdag 27 juli een zilveren medaille. Daarnaast behaalde Nederland nog drie bronzen medailles en een eervolle vermelding en eindigde daarmee op de 25ste plaats. De Internationale Wiskunde Olympiade vond van 21 tot 28 juli plaats in Santa Marta, Colombia. Op elk van de twee wedstrijddagen kregen de leerlingen drie zeer pittige wiskundeopgaven voor hun kiezen. Voor elke opgave waren zeven punten te behalen. Jeroen Winkel wist op de uitzonderlijk moeilijke opgave 6 maar liefst drie punten te scoren en hoefde daarmee slechts 19 van de 528 andere deelnemers voor zich te laten.

De resultaten van de afzonderlijke leerlingen zijn als volgt: Jeroen Winkel (16 jaar, Nijmegen, 5 vwo, Stedelijk Gymnasium Nijmegen): 30 punten, zilver; Jeroen Huijben (17 jaar, Goirle, 6 vwo, Theresia Lyceum Tilburg): 27 punten, zilver; Peter Gerlagh (16 jaar, Tilburg, 5 vwo, Theresia Lyceum Tilburg): 23 punten, brons; Michelle Sweering (16 jaar, Krimpen aan den IJssel, 5 vwo, Erasmiaans Gymnasium Rotterdam): 21 punten, brons; Ragnar Groot Koerkamp (18 jaar, Opijnen, 6 vwo, Cambium College Zaltbommel): 19 punten, brons; Djurre Tijsma (18 jaar, Leeuwarden, 6 vwo, Christelijk Gymnasium Beyers Naudé Leeuwarden): 13 punten, eervolle vermelding. Toppresteerder Jeroen Winkel: "Dit jaar heb ik jammer genoeg net een gouden medaille gemist. Maar volgend jaar heb ik nog een kans en ga ik zeker voor goud!" Ook Peter Gerlagh en Michelle Sweering kunnen volgend jaar opnieuw een gooi naar de prijzen doen.

www.wiskundeolympiade.nl

Slingerklok maakt een comeback

Het 350 jaar oude probleem van de dubbele slingerklok maakt een comeback! Als eerste onderzocht door Huygens, heeft de beweging van een slinger die aan een andere slinger vast zit wiskundigen al eeuwen beziggehouden. Huygens merkte als eerste op dat het mogelijk

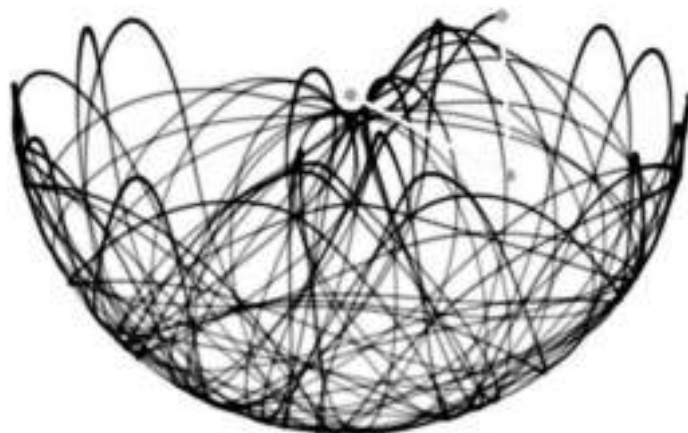
was dat de twee slingers andere richtingen op konden bewegen. Nog steeds is het erg lastig om een expliciete formule voor de beweging op te schrijven.

Nu hebben onderzoekers van de Universiteit van Pittsburgh de expliciete formules voor de beweging gebruikt om bepaalde interactie tussen biologische verschijningen te verklaren. Deze interactie wordt ook wel *indirecte koppeling* genoemd.

De modellen die het effect nu gebruiken, zijn bijvoorbeeld modellen voor ziekten waarin neuronen niet meer 'uit' gaan, zoals epilepsie. Een ander voorbeeld is in de biologie, waarin bacteriën van buitenaf worden beïnvloed om de groei te synchroniseren.

In de modellen gebruiken de wetenschappers genen die ritmisch aan- en uitgaan. Als ze aan staan maken ze een proteïne aan die de genen stimuleert om aan te gaan. In het verleden werd gedacht dat uiteindelijk alle genen synchroon gaan lopen, maar met dit model kunnen de wetenschappers laten zien dat ze ook volledig asynchroon kunnen lopen.

science20.com



Khan Academie in handen van gebruikers

Binnenkort zal de Khan Academie verder gaan als stichting, waardoor het online leerplatform onafhankelijk wordt van ontwikkelaar Perceptum. Dit biedt meer mogelijkheden tot groei voor het platform en vergroot de inbreng van de gebruikers. De stichting zal worden geleid door een nog aan te stellen bestuur. Daarnaast kunnen ambassadeurs en vrijwilligers zich melden om bij te dragen aan het project en zich verbinden met deze innoverende leeromgeving.

De reden dat de Khan Academie een stichting wordt, heeft te maken met het ideaal van Perceptum van het bieden van gratis online onderwijs. Het bedrijf vindt dat deze adaptieve en gedifferentieerde manier van onderwijs voor iedereen toegankelijk moet zijn. Toch zijn er de benodigde kosten voor het bouwen en onderhouden van dit platform en daarom wil de organisatie overgaan op het werken met ambassadeurs en donateurs met affiniteit voor het onderwijs.

De Khan Academie gaat daarnaast binnenkort verder onder een nieuwe naam. De initiatiefnemers van de online leeromgeving geloven in de filosofie van crowdsourcing, hetgeen betekent dat de gebruikers informatie kunnen delen, verbeteren en aanvullen. De gebruikersgemeenschap heeft de touwtjes in handen, het bestuur zal de stichting slechts in goede banen leiden. Met deze denkbeelden slaat het platform een andere weg in dan de Amerikaanse Khan Academy. Terwijl die zich richt op het intern ontwikkelen en organiseren van lesmateriaal, richt de Nederlandse Khan Academie zich op de kracht van het

publiek. Omdat het platform vanwege de associatie met de Khan Academy haar visie niet voldoende kan uitdragen, heeft Perceptum besloten een naamswijziging door te gaan voeren. Iedereen kan meedenken over een nieuwe naam en zich aanmelden voor een bestuursfunctie op de website van Perceptum.

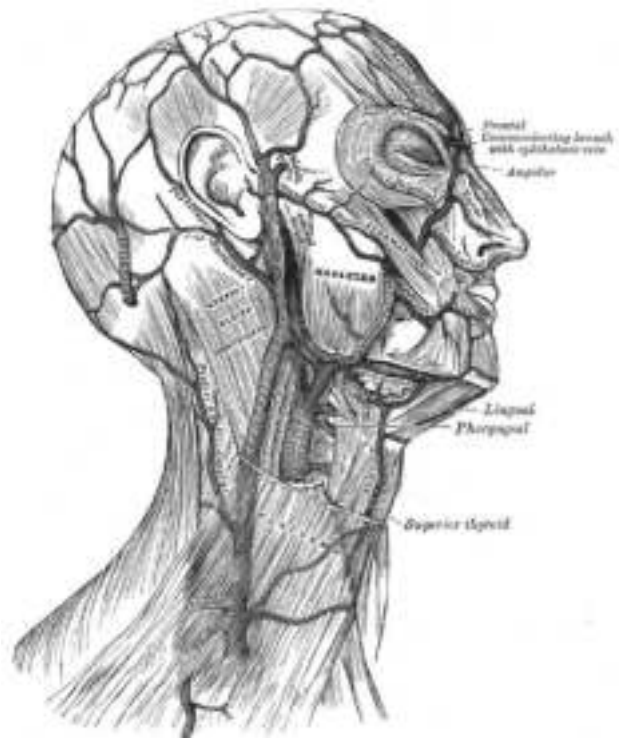
perceptum.nl

Mysterieuze wiskunde-graffiti

In Boscombe, een buitenwijk van het stadje Bournemouth in Zuid-Engeland is een wel heel aparte vorm van graffiti gevonden. Op een zwarte muur zijn verschillende wiskundige symbolen geklad. Journalisten van de Daily Echo, de plaatselijke krant, en de inwoners waren nogal geïnteresseerd in wat de betekenis was. Ze waren zo geïntrigeerd dat ze een oproep deden aan de lezers van de krant of iemand misschien de betekenis wist. Al eerder was er sprake van een grote hoeveelheid graffiti in de buurt, het meeste met name politiek en beledigend bedoeld, waarschijnlijk dat dit 'kunstwerk' daarom zoveel aandacht kreeg.

Uiteindelijk heeft wiskundeleraar Owen Elton het 'mysterie' opgelost. Het bleek een collectie van willekeurige elementen van de wiskunde te zijn zonder enige samenhang. Onder de formules stond de formule van een eerste-orde-differentiaalvergelijking, waarvan de oplossing grappig genoeg niet een van de grafieken was op de muur. Elton denkt dat het om een grapje gaat, dit concludeert hij met name uit de conclusie "NO VACANCIES" rechtsonder op de muur. Helaas dus geen geniale uitwerking van een open probleem.

Daily Echo, Bournemouth



Jonathan Corrier

Bloed nieuwste manier van identificeren

Vergeet de irisscan of de vingerafdrukken, de nieuwste manier van identificeren is via de adertjes in het gezicht. Bij deze manier vanuit de biometrie wordt er een warmtescan gemaakt van het gezicht en worden de aders afgebeeld in de computer. Het schijnt dat deze plaats van de aders bijna onmogelijk is om te imiteren. Dit maakt het natuurlijk een ideale identificatietechniek.

Wetenschappers van de Jadavpur Universiteit in India leggen in de *International Journal of Computational Intelligence Studies* uit dat de aders onder de huid van het gezicht net zo uniek zijn als de iris of vingerafdrukken. Deze oude technieken zijn te omzeilen door bijvoorbeeld rubberen vingerhoedjes of contactlenzen. Het is echter praktisch onmogelijk om deze aders te simuleren en zo je identiteit te verhullen.

Ook is deze manier van scannen passiever omdat er alleen maar een warmte-scan gemaakt hoeft te worden, dus er hoeft niet in een klein apparaat gekeken te worden of een vingerafdruk gemaakt te worden. De wetenschappers claimen in hun artikel dat ze met 97 procent zekerheid personen kunnen identificeren met hun techniek. Dit is in principe genoeg voor hoge beveiligingstoepassingen, helemaal in combinatie met andere vormen van beveiliging.

eurekalert.com

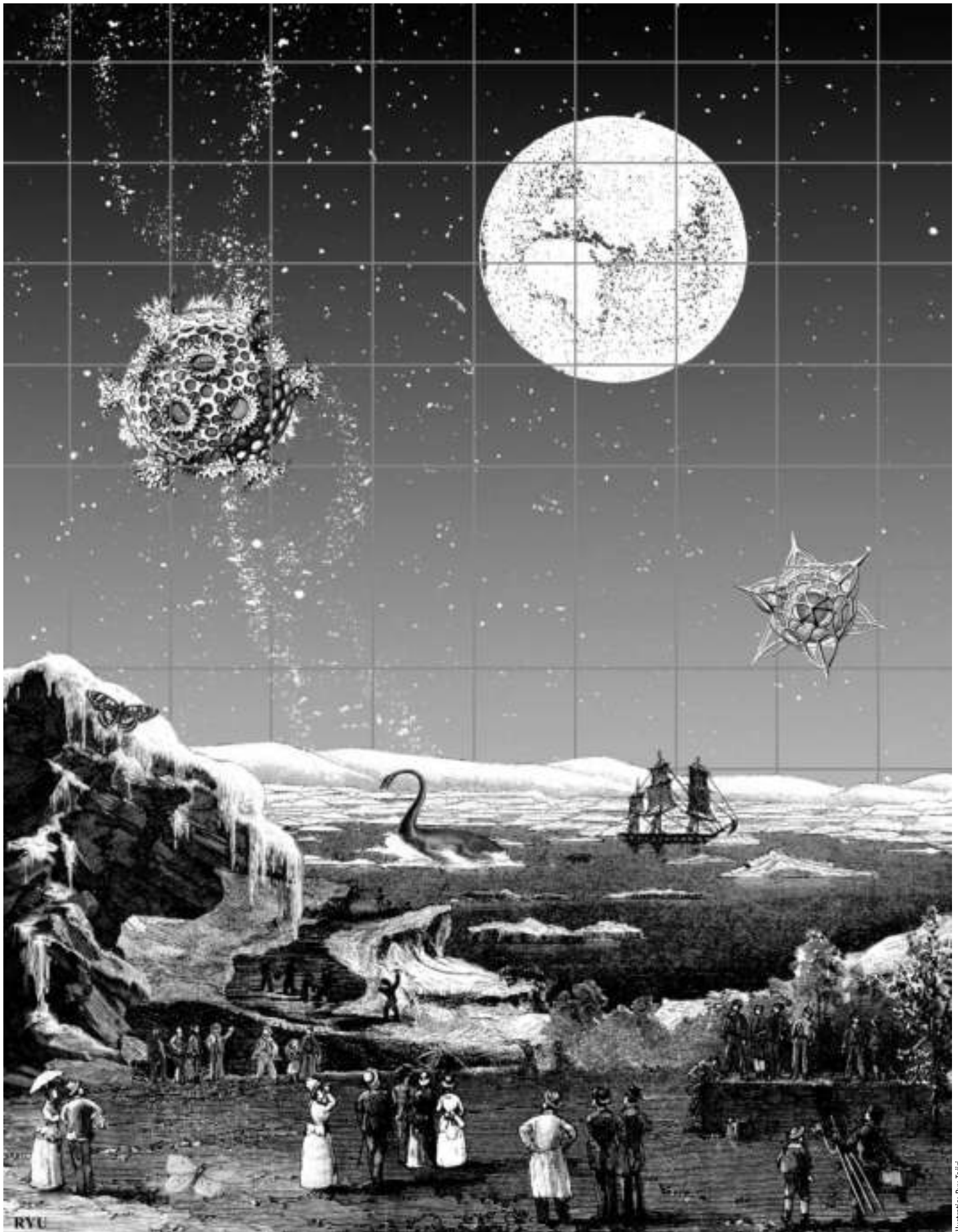
Derde universiteit Amsterdam

De gemeente Amsterdam heeft begin april een competitie opgezet met als doel een nieuwe kennisinstelling te creëren in de stad. De gemeente wil door middel van dit kennisinstituut de economische structuur versterken. Er kwamen uiteindelijk dertien voorstellen binnen. Een onafhankelijke jury heeft de voorstellen anoniem beoordeeld en uiteindelijk vijf uitgekozen die doorgaan naar de volgende ronde. Onder de inzendingen zaten ook voorstellen van prestigieuze buitenlandse universiteiten zoals Colombia en MIT. De hoofdpartijen van de geselecteerde vijf voorstellen zijn:

- Amsterdam City Technology: instituut voor 'city-directed' technologie, met zowel onderwijs als onderzoek (Universiteit van Amsterdam, Vrije Universiteit, Hogeschool van Amsterdam, Columbia University).
- The Amsterdam Institute of Advanced Metropolitan Solutions: instituut voor het ontwikkelen van urban solutions (TU Delft, Wageningen Universiteit en Massachusetts Institute of Technology).
- The Amsterdam School of Health and Technology: richt zich op onderwijs en onderzoek naar gezond leven in een stedelijke omgeving (Amsterdam Institute for Global Health and Development en Duke University).
- The Center for Urban Tech Acceleration: steunt ondernemers en onderzoekers bij het uitbreiden van hun onderneming op gebied van stedelijke technologie (THNK, OMA en UNstudio).
- Nexuslabs Amsterdam: netwerk-organisatie die multinationals, ondernemers, onderzoekers, durfkapitaal en overheid bijeen brengt ten behoeve van duurzame stedelijke oplossingen (Stichting Nexus-Labs).

De tweede ronde zal bestaan uit het verder uitwerken van de plannen met als deadline eind september. Najaar 2013 zal de gemeente een besluit nemen en beginnen met de constructie van een van de voorstellen.

amsterdam.nl



Alef Sterk

Faculteit Elektrotechniek, Wiskunde en Informatica
Universiteit Twente
a.e.sterk@gmail.com

Renato Vitolo

Exeter Climate Systems
University of Exeter, UK
r.vitolo@exeter.ac.uk

Henk Broer

Johann Bernoulli Instituut
Rijksuniversiteit Groningen
h.w.broer@rug.nl

Het onvoorspelbare venijn van de staart

De statistische theorie van extreme waarden biedt een handvat voor de studie van meteorologische extremen zoals orkanen en hittegolven. De klassieke theorie voor extreme waarden richt zich op onafhankelijke stochasten, later uitgebreid naar afhankelijke (bijvoorbeeld tijdsgecorreleerde) stochasten. Een recente ontwikkeling is de uitbreiding van de theorie naar extremen in deterministische dynamische systemen. In dit artikel gaan Alef Sterk, Renato Vitolo en Henk Broer in op deze ontwikkeling.

Weersextremen, zoals orkaan Katrina of windstorm Kyrill, kunnen grote schade veroorzaken. Accurate inschattingen voor de kansverdeling en de voorspelbaarheid van dergelijke extremen vormen een grote uitdaging voor zowel de verzekeringsindustrie als meteorologische instituten. In dit artikel worden kort twee nieuwe ontwikkelingen besproken betreffende de statistische eigenschappen en de voorspelbaarheid van extremen in deterministische systemen, zoals modellen die voor weersvoorspelling worden gebruikt.

Kansverdelingen voor extremen

De klassieke Extreme Waardestelling [3] beschrijft de limietkansverdeling van grote waarden uit een stochastisch proces. Beschouw een rij $\{X_i\}_{i=1}^{\infty}$ van onafhankelij-

ke, gelijkverdeelde stochasten. Zij $M_n = \max\{X_1, \dots, X_n\}$ het n -de partiële maximum (de 'grote waarden' van de rij) en neem aan dat voor geschikt gekozen rijen (a_n) en (b_n) de herschaalde variabele $a_n(M_n - b_n)$ convergeert naar een verdeling G :

$$\lim_{n \rightarrow \infty} P(a_n(M_n - b_n) \leq z) = G(z).$$

De Extreme Waardestelling zegt dan dat G behoort tot de GEV (Generalized Extreme Value) parametrische familie met plaats-, schaal- en staartparameters (μ, σ, ξ) :

$$G(z) = \exp \left(- \left[1 + \xi \left(\frac{z - \mu}{\sigma} \right) \right]^{-1/\xi} \right), \quad (1)$$

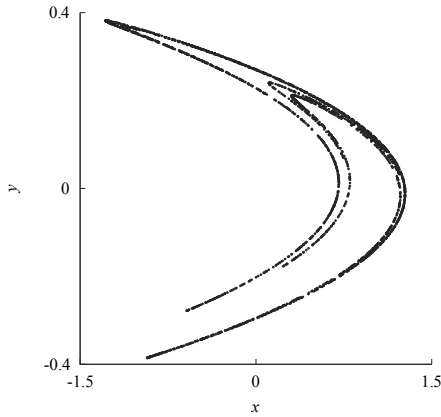
met $\mu \in (-\infty, \infty)$ en $\sigma \in (0, \infty)$. Voor toepas-

singsdoeleinden is de zogenaamde staartparameter $\xi \in (-\infty, \infty)$ de belangrijkste parameter omdat deze de kans op extremen ('the tail thickness') bepaalt. De vorm van de limietverdeling is onafhankelijk van het 'moederproces' $\{X_i\}_{i=1}^{\infty}$, net zoals voor de Centrale Limietstelling: in deze zin is de Extreme Waardestelling universeel.

Ter illustratie beschouwen we een rij $\{X_i\}_{i=1}^{\infty}$ van onafhankelijke stochasten met een uniforme verdeling op het interval $[0, 1]$. Wegens de onafhankelijkheid geldt voor alle $z \leq 1$ dat

$$\begin{aligned} P(M_n \leq z) &= P(X_1 \leq z, X_2 \leq z, \dots, X_n \leq z) \\ &= P(X_1 \leq z)P(X_2 \leq z) \cdots P(X_n \leq z) \\ &= z^n. \end{aligned}$$

Hieruit blijkt ook de noodzaak tot herschalen van de maxima M_n . Immers, zonder herschaling zou de kansverdeling ontaarden in de Dirac-delta op $z = 1$. Met de keuze $a_n = n$



Figuur 1 De Hénon-aantrekker

en $b_n = 1$ volgt

$$\begin{aligned} \lim_{n \rightarrow \infty} P(a_n(M_n - b_n) \leq z) \\ = \lim_{n \rightarrow \infty} \left(\frac{z}{n} + 1 \right)^n = e^z, \end{aligned}$$

en hierin herkennen we de verdeling (1) met parameters $(\mu, \sigma, \xi) = (-1, 1, -1)$. Andere keuzes voor (a_n) en (b_n) zijn ook mogelijk; in het algemeen veranderen μ en σ hierdoor, maar ξ is eenduidig bepaald.

Leadbetter [12–13] heeft voorwaarden geformuleerd waaronder de kansverdeling (1) ook geldt in het geval dat de X_i afhankelijk zijn. Deze voorwaarden komen neer op het volgende:

- Correlaties tussen de X_i convergeren voldoende snel naar nul.
- Overschrijdingen van de X_i boven grenswaarden clusteren niet te sterk.

Voor precieze definities wordt de lezer verwezen naar bovengenoemde werken.

Extremen in deterministische systemen

In het afgelopen decennium heeft het onderzoek naar extremen zich uitgebreid naar het terrein van deterministische systemen. In het algemeen bestaat een dergelijk systeem uit een toestandsruimte M (meestal $M \subset \mathbb{R}^n$), een tijdsverzameling T en een evolutieoperator

$$\Phi : T \times M \rightarrow M, \quad (t, x) \mapsto \Phi_t(x),$$

met de eigenschappen

$$\Phi_0 = \text{Id}_M, \quad \Phi_t \circ \Phi_s = \Phi_{t+s},$$

zie bijvoorbeeld [1]. Denk bijvoorbeeld aan systemen beschreven door geïtereerde afbeeldingen ($T \subset \mathbb{Z}$) of differentiaalvergelijkingen ($T \subset \mathbb{R}$).

Het idee is om langs de evolutie, begin-

nend in $x \in M$, een scalaire observatie $\omega : M \rightarrow \mathbb{R}$ te berekenen. Hieruit volgt een tijdreeks

$$X_t = \omega(\Phi_t(x)), \quad t \in T, \quad (2)$$

waarvan de statistiek van extremen kan worden bestudeerd. Dit is voornamelijk zinvol voor chaotische systemen die een Sinai–Ruelle–Bowen-invariante maat hebben: hierdoor kan X_t in (2) als een stochastische variabele worden beschouwd [2, 18]. Voor de geldigheid van de Extreme Waardstelling in deze context moet aan de eerder genoemde voorwaarden van Leadbetter worden voldaan. In recente literatuur zijn deze voorwaarden geverifieerd voor diverse klassen dynamische systemen en observaties, zie bijvoorbeeld [6–10]. Deze studies beperkten zich voornamelijk tot observaties van de vorm

$$\omega(x) = f(\text{dist}(x, \tilde{x})), \quad f : [0, \infty) \rightarrow \mathbb{R}, \quad (3)$$

waarin $\text{dist}(\cdot, \cdot)$ een geschikte metriek op de toestandsruimte M is. Verder wordt in 3 aangenomen dat $\tilde{x}$ een dichtheidspunt is van de invariante maat van het systeem, hetgeen betekent dat de relatieve maat van elke ϵ -omgeving van $\tilde{x}$ nadert tot 1 als ϵ tot 0 nadert. Dit soort observaties zijn echter minder relevant voor fysische toepassingen.

Observaties in geofysische modellen

De studie naar extremen in dynamische sys-

temen vindt vooral toepassing in de context van geofysische modellen. Dergelijke modellen worden doorgaans afgeleid uit behoudswetten voor grootheden zoals massa, impuls en warmte [11]. Hieruit volgen partiële differentiaalvergelijkingen die de evolutie van velden zoals druk, temperatuur en windsnelheid beschrijven. Discretisatie geeft een dynamisch systeem op een eindigdimensionale toestandsruimte $M = \mathbb{R}^n$ in de vorm van een stelsel van gewone differentiaalvergelijkingen. Typische grootheden om de extremen van te bestuderen zijn bijvoorbeeld druk of vorticeiteit, welke gemodelleerd kunnen worden met observaties van de vorm

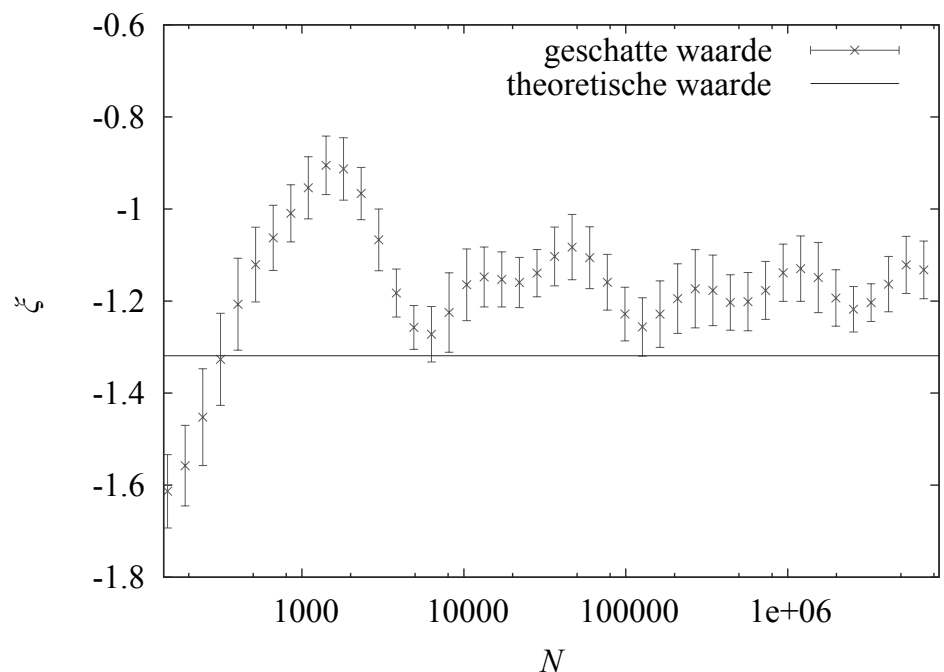
$$\omega(x) = v^\top x \quad (4)$$

met $v \in \mathbb{R}^n$, of kwadratische windsnelheid, welke gemodelleerd kan worden door

$$\omega(x) = x^\top E x, \quad (5)$$

waarbij E een $n \times n$ semi-positief definitieve matrix is.

Recent onderzoek van de auteurs [14] heeft zich gericht op het geval van chaotische dynamische systemen met observaties van de vorm (4) en (5). In dit geval blijkt de statistiek van extreme waarden cruciaal af te hangen van de meetkunde van de onderliggende vreemde aantrekker van het model in de



Figuur 2 Schattingen voor de staartparameter ξ als functie van de bloklength N berekend met behulp van de blokmaximum-methode voor de Hénon-afbeelding.

zin dat de staartparameter ξ gerelateerd kan worden aan de dimensies van lokale stabiele en instabiele variëteiten. Voor enkele systemen kunnen dergelijke relaties analytisch bewezen worden; voor de meeste systemen zijn de resultaten echter nog experimenteel van aard.

Ter illustratie beschouwen we de afbeelding van Hénon gegeven door

$$H: \mathbb{R}^2 \rightarrow \mathbb{R}^2, \quad H(x, y) = (1 - ax^2 + y, bx),$$

met parameters $(a, b) = (1.4, 0.3)$. Voor deze klassieke parameterwaarden zijn de evoluties chaotisch, zie Figuur 1. Als observatie kiezen we $\omega(x, y) = x$: dit is duidelijk niet van de vorm (3). In [14] is gepostuleerd dat de staartparameter ξ voldoet aan

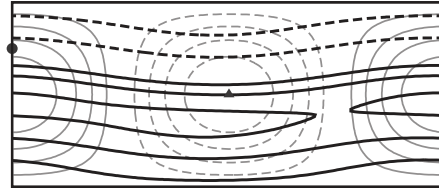
$$-\frac{1}{\xi} = \dim(A) - \frac{1}{2}, \quad (6)$$

waar $\dim(A)$ de Hausdorff-dimensie van de Hénon-aantrekker is. Dit vermoeden is getoetst door ξ te schatten met behulp van de zogenaamde ‘blokmaximum-methode’. In deze methode wordt de tijdreeks (2) opgedeeld in voldoende grote blokken en van elk blok wordt het maximum berekend. Onder de aanname dat deze maxima de verdeling (1) hebben, kan de parameter ξ worden geschat met behulp van standaard methoden uit de statistiek zoals maximum likelihood.

Figuur 2 toont schattingen van ξ als functie van de blok lengte N . Merk op dat deze schattingen grote schommelingen vertonen; hierdoor is het geven van een precieze schatting zeer lastig. Voor zeer grote N benaderen de schattingen de theoretisch voorspelde waarde. In het algemeen zijn zeer lange tijdreeksen nodig om ξ te kunnen schatten met de blokmaxima. Een interessante vraag voor verder onderzoek is dan ook of relaties zoals (6) algemeen geldig zijn en zo als basis kunnen dienen voor schattingsmethoden die minder data vergen. Hoewel de relatie (6) nog een vermoeden is voor de Hénon-afbeelding, kunnen soortgelijke formules wel bewezen worden voor systemen als de Solenoïde of de afbeelding van Thom [14].

Voorspelbaarheid van extremen

Kansverdelingen voor extremen zijn vooral belangrijk voor de verzekeringsindustrie om bijvoorbeeld voldoende kapitaal te reserveren om vergoedingen bij stormschade te kunnen uitkeren. Voor meteorologische instituten is het echter ook van belang om de voor-



Figuur 3 Een westelijk en een geblokkeerd stromingspatroon. Stroomlijnen zijn dik/zwart; orografie is in grijs aangegeven: gestippelde en doorgetrokken lijnen geven respectievelijk negatieve en positieve hoogten aan, dat wil zeggen het dal en de berg. De driehoek markeert de locatie met de laagste waarde van de orografie; de stip markeert de locatie halverwege het hoogste punt van de orografie.

spelbaarheid van extremen te kennen. Sinds het werk van Edward Lorenz [15] in de jaren zestig van de vorige eeuw weten we dat evoluties van niet-lineaire systemen gevoelig kunnen afhangen van de gekozen begintoestand. Het was deze ontdekking die tot het besef heeft geleid dat het weer maar een beperkte tijd vooruit kan worden voorspeld. De vraag is nu of extremen van meteorologische grootheden slechter of juist beter voorspelbaar zijn dan niet-extremen.

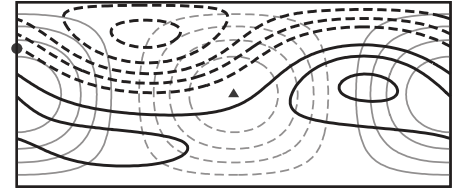
In chaotische systemen drijven evoluties beginnend in verschillende begintoestanden exponentieel snel uit elkaar. Voor twee begintoestanden x en y geldt

$$\text{dist}(\Phi_\tau(x), \Phi_\tau(y)) \propto e^{\lambda\tau} \text{dist}(x, y).$$

Hierin is λ een maat voor de exponentiële groei van fouten in de begintoestand en deze hangt af van zowel de begintoestand x als de voorspellingstijd τ . In [16] is onderzocht of de groefactor λ groter of juist kleiner is voor begintoestanden $x \in M$ die leiden naar extreme waarden van de observatie $\omega(\Phi_\tau(x))$ na voorspellingstijd τ .

Als voorbeeld beschouwen we een model voor grootschalige atmosferische stromingen langs orografie (bergen) op gemiddelde breedtegraden. Fysische beschouwingen leiden tot een vereenvoudigd model gegeven door een partiële differentiaalvergelijking voor de stroomfunctie [11]. Door de stroomfunctie te ontwikkelen in een afgekapte Fourierreeks met tijdsafhankelijke coëfficiënten volgt een stelsel van zes gewone differentiaalvergelijkingen die de dynamica op de grootste ruimtelijke schalen beschrijven [17].

Het model werd gebruikt om de relatie tussen orografie en atmosferische regimes te analyseren [4–5]. Voor specifieke keuzes van de parameters van het model (waaronder de hoogte van de orografie) wordt de dynamica gekarakteriseerd door intermitterende overgangen tussen een westelijke straalstroom en een geblokkeerde stroom, zie Figuur 3. Deze dynamica kan worden verklaard door de aanwezigheid van een Shil'nikov-achtige



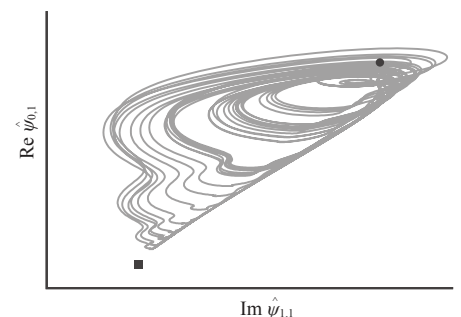
vreemde aantrekker in de 6-dimensionale toestandsruimte van het model, zie Figuur 4.

We kiezen twee observaties door de windsnelheid te berekenen op twee geografische locaties: ω_{zij} en ω_{dal} geven respectievelijk de windsnelheid halverwege het hoogste punt (aan de zijkant van de orografie) en het laagste punt van de orografie. Deze locaties zijn in Figuur 3 weergegeven met respectievelijk een stip en een driehoek). Voor 10^6 begintoestanden $x \in \mathbb{R}^6$ berekenen we de groefactor $\lambda = \lambda(x, \tau)$ en bestuderen we de verdeling van de groefactoren voor begincondities die na τ tijdseenheden leiden tot een overschrijding van ω_{zij} of ω_{dal} boven een vooraf gekozen grenswaarde q :

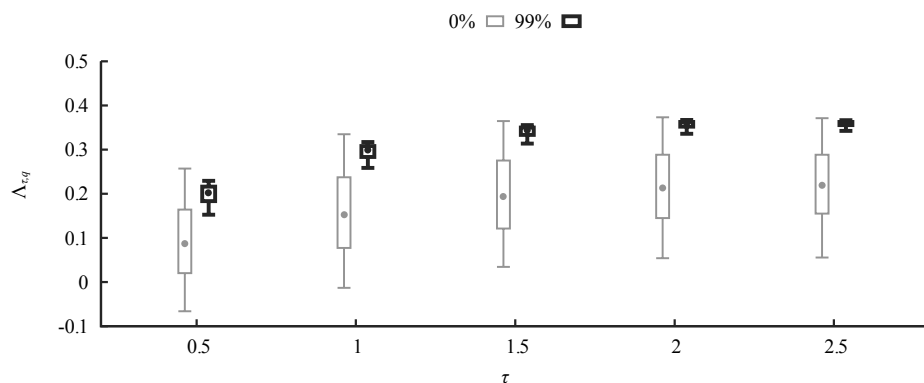
$$\Lambda_{\tau,q} = \{ \lambda(x, \tau) \mid \omega(\Phi_\tau(x)) \geq q \}.$$

Als groefactoren typisch groter zijn voor hogere waarden van q dan betekent dit dat extremen minder voorspelbaar zijn.

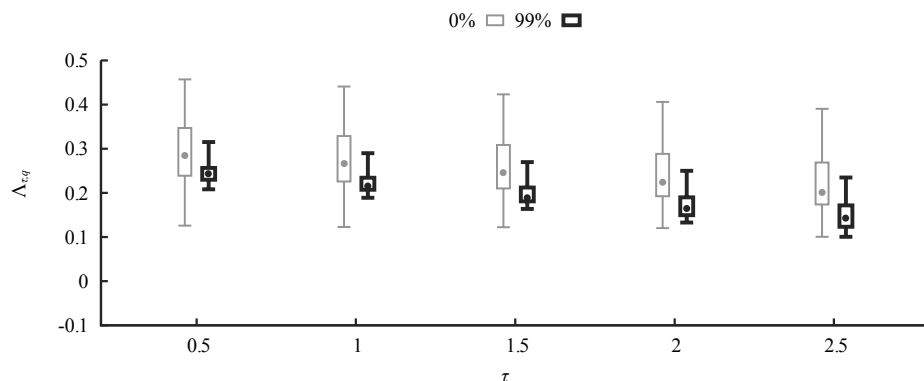
Figuur 5 toont boxplots van $\Lambda_{\tau,q}$ als functie van de voorspellingstijd τ voor ω_{zij} . Voor elke waarde van τ zijn twee boxplots weergegeven: voor grenswaarde $q = 0$ (dat wil zeggen alle begincondities) en q gelijk aan het 99ste percentiel van de windsnelheid (dat wil zeggen alleen de begincondities die naar hele sterke windsnelheden leiden). Het is duidelijk dat in dit geval de begincondities die



Figuur 4 Een Shil'nikov-achtige aantrekker in de 6-dimensionale toestandsruimte van een atmosferisch model. Er vinden intermitterende overgangen plaats tussen twee punten in de toestandsruimte (aangeduid met een stip en een vierkant); deze punten corresponderen met een westelijke en een geblokkeerde atmosferische stroming zoals weergegeven in Figuur 3.



Figuur 5 Verdelingen van de verzameling van groeifactoren $\Lambda_{\tau,q}$: boxplots als functie van de voorspellingstijd τ voor twee ondergrenzen q . De lengte van de doos geeft de interkwartielafstand, de lijnen strekken zich uit tot het maximum en minimum en stippen geven de mediaan aan. Voor elke τ -waarde zijn twee boxplots weergegeven: voor $q=0$ en het 99ste percentiel van de windsnelheid. Begintoestanden die leiden naar extreme windsnelheden hebben systematisch grotere groeifactoren.



Figuur 6 Zoals Figuur 5, maar dan voor de windsnelheid berekend in het laagste punt van de orografie. Begintoestanden die leiden naar extreme windsnelheden hebben systematisch lagere groeifactoren.

leiden naar grote windsnelheden geassocieerd zijn met grotere groeifactoren: dit betekent dat hoge windsnelheden slechter voorspelbaar zijn in dit punt van de orografie. Figuur 6 toont hetzelfde diagram voor de observatie ω_{dal} : nu hebben begincondities die leiden naar hoge windsnelheden juist lagere groeifactoren, hetgeen betekent dat hoge windsnelheden beter voorspelbaar zijn in dit punt van de orografie. Een belangrijke conclusie uit [16] is dat het kennelijk niet mogelijk is om algemene uitspraken te kunnen doen over de voorspelbaarheid van extreme waarden. De voorspelbaarheid hangt af van de dynamica van het systeem, de observatie waarvan we de extremen willen bestuderen en de voorspellingstijd τ .

Tot slot

De studie van extreme waarden in dynamische systemen is een onderzoeksgebied dat zich in snel tempo aan het uitbreiden is. Enerzijds geven deze studies aanleiding tot interessante wiskundige vragen, zoals de relatie tussen de staartparameter van de Extreme Waardeverdeling van een chaotisch dynamisch systeem en de dimensie van de vreemde aantrekker waarop de dynamica plaatsvindt. Anderzijds zijn er vele maatschappelijk relevante toepassingen mogelijk, bijvoorbeeld studies naar de voorspelbaarheid van meteorologische extremen zoals orkanen of windstormen.

Referenties

- 1 H.W. Broer, F. Takens. *Dynamical Systems and Chaos*. Springer, 2011.
- 2 C. Bonatti, L. Diaz, M. Viana. *Dynamics Beyond Uniform Hyperbolicity*. Springer, 2005.
- 3 S. Coles. *An Introduction to Statistical Modeling of Extreme Values*, Springer, 2001.
- 4 J.G. Charney, J.G. DeVore. *Multiple flow equilibria in the atmosphere and blocking*, J. Atmos. Sci. 36, pp. 1205–1216, 1979.
- 5 D.T. Crommelin, J.D. Opsteegh, F. Verhulst. *A mechanism for atmospheric regime behavior*, J. Atmos. Sci. 61, pp. 1406–1419, 2004.
- 6 A.C.M. Freitas, J.M. Freitas. *On the link between dependence and independence in extreme value theory for dynamical systems*, Statistics and Probability Letters 78, pp. 10088–1093, 2008.
- 7 A.C.M. Freitas, J.M. Freitas, M. Todd. *Extreme value laws in dynamical systems for non-smooth observations*, J. Stat. Phys. 142, pp. 108–126, 2010.
- 8 C. Gupta. *Extreme-value distributions for some classes of non-uniformly partially hyperbolic dynamical systems*, Ergodic Theory and Dynamical Systems 30, pp. 757–771, 2010.
- 9 G. Haiman. *Extreme values of the tent map process*, Statistics and Probability Letters 65, pp. 451–456, 2003.
- 10 M.P. Holland, M. Nicol, A. Török. *Extreme value theory for non-uniformly expanding dynamical systems*, Transactions of the AMS, pp. 661–688, 2012.
- 11 J.R. Holton. *Introduction to Dynamic Meteorology*. Elsevier Academic Press, 2004.
- 12 M.R. Leadbetter, G. Lindgren, H. Rootzén. *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag, 1980.
- 13 M.R. Leadbetter. *Extremes and local dependence in stationary sequences*, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 65, pp. 291–306, 1983.
- 14 M.P. Holland, R. Vitolo, P. Rabassa, A.E. Sterk, H.W. Broer. *Extreme value laws in dynamical systems under physical observables*, Physica D 241, pp. 497–513, 2012.
- 15 E.N. Lorenz. *Deterministic non-periodic flow*, J. Atmos. Sci. 20, pp. 130–141, 1963.
- 16 A.E. Sterk, M.P. Holland, P. Rabassa, H.W. Broer, R. Vitolo. *Predictability of extreme values in geophysical models* Nonlinear Processes in Geophysics 19, pp. 529–539, 2012.
- 17 H.E. de Swart. *Analysis of a six-component atmospheric spectral model: chaos, predictability and vacillation*, Physica D 36, pp. 222–234, 1989.
- 18 L.-S. Young. *What are SRB measures, and which dynamical systems have them?*, J. Stat. Phys. 108, pp. 733–754, 2002.

Arnold Heemink

Faculteit EWI

TU Delft

a.w.heemink@tudelft.nl

Peter Jan van Leeuwen

Department of Meteorology

University of Reading, UK

p.j.vanleeuwen@reading.ac.uk

Voorspellen met behulp van data-assimilatie

Voor het maken van voorspellingen van bijvoorbeeld het weer zijn behalve een numeriek model ook waarnemingen van groot belang. Onder meteorologen staat het systematisch combineren van modellen en waarnemingen bekend als data-assimilatie. Moderne technieken voor data-assimilatie maken gebruik van ideeën uit de optimale besturingstheorie en de filtertheorie, en het ontwikkelen van dergelijke technieken is een actief onderzoeksterrein waar zowel wiskundigen als meteorologen en oceanografen aan werken. Arnold Heemink en Peter Jan van Leeuwen bespreken in dit artikel de voornaamste hedendaagse data-assimilatie methoden.

Numeriek wiskundige modellen worden tegenwoordig steeds vaker gebruikt bij het berekenen van voorspellingen. Voorbeelden hiervan zijn voorspellingen van het weer, de waterstanden langs de Nederlandse kust of luchtvervuiling in steden. De resultaten van de modelberekeningen zijn echter beïnvloed door tal van onzekerheden. Bijvoorbeeld door onnauwkeurige begin- en of randvoorwaarden of door onzekerheden in de waarden van de modelparameters. Door gebruik te maken van waarnemingen kunnen de modelresultaten worden verbeterd. Dit wordt data-assimilatie genoemd. Deze term is vele jaren geleden al geïntroduceerd door de meteorologen. Meteorologen zeggen vaak dat zij het weer van morgen niet goed kunnen voorspellen om de simpele reden dat het weervan vandaag niet overal goed bekend is. Veel fouten in de kortetermijnmodelvoorspellingen zijn het gevolg van fouten die op het moment van voorspellen al aanwezig waren. Dit geeft het belang aan van goede meetinformatie én van goede data-assimilatietechnieken om de voorspelfouten te minimaliseren.

Tot het eind van de negentiger jaren was de meest gebruikte data-assimilatietechniek optimale interpolatie. Hierbij werd de bere-

kende modeltoestand gecorrigeerd op basis van waarnemingen van die toestand en aannames omtrent de statistiek van de modelfout en de meetfout. Optimale interpolatie heeft een zeer principiële nadeel. De gecorrigeerde modeltoestand is niet consistent met het model omdat de procesdynamica niet gebruikt wordt bij het bepalen van de correcties. De gecorrigeerde modeltoestand voldoet dan niet aan de modelvergelijkingen, wat zelfs kan leiden tot instabiliteiten.

Data-assimilatie kan gezien worden als een invers probleem. Op basis van de uitvoer van het model op de meetlocaties worden de modelfouten gereconstrueerd. Met informatie over het gevolg moet dus de oorzaak worden opgespoord. De algoritmen die voor nauwkeurige data-assimilatieschema's gebruikt worden, zijn sterk geïnspireerd door de optimale besturingstheorie en de filtertheorie. De laatste aanpak resulteert uiteindelijk in het Ensemble Kalman Filter. Anders dan optimale interpolatie zijn deze methoden in staat om het inverse probleem netjes op te lossen, waardoor de modeltoestand ook na correctie blijft voldoen aan de modelvergelijkingen. Het gebruik van deze data-assimilatiemethoden is daardoor niet slechts

een marginale verbetering ten opzichte van een traditionele aanpak als optimale interpolatie. Het is een wezenlijke stap vooruit.

Data-assimilatie kan ook worden geformuleerd op basis van het Theorema van Bayes. Voor het geval dat modelonzekerheden benaderd kunnen worden met behulp van een Gaussische kansdichtheid resulteert dit ook weer in het Ensemble Kalman Filter. Als de Gaussische benadering niet op gaat, krijgen we het Particle Filter.

4D variationele data-assimilatie

Beschouw het volgende numeriek wiskundige model:

$$\begin{aligned} x(t_{i+1}) &= M_i[x(t_i), p], \\ x(t_0) &= x_0(p), \quad i = 0, 1, 2, \dots, K, \end{aligned} \quad (1)$$

waarbij $x(t_i)$ de toestandsvector is op tijdstip t_i en p de te schatten parametervector. De niet-lineaire functie M_i stelt het numerieke schema voor. $M_i[x(t_i), p]$ is een representatie van een computercode die op een ingewikkelde wijze $x(t_{i+1})$ uitrekent bij gegeven input $x(t_i)$ en p . $x(t_i)$ bestaat uit alle variabelen in alle roosterpunten van het model. De dimensie van $x(t_i)$ kan daarom zeer groot zijn ($10^6 - 10^8$). De parameters kunnen ruimteafhankelijke fysische parameters zijn in het model, parameters in de randvoorwaarden, maar ze kunnen ook de volledige beginconditie bevatten. Ook de dimensie van p kan dus heel groot zijn.

Er wordt vervolgens aangenomen dat de waarnemingen beschikbaar zijn in de volgen-

Minimalisatie met het geadjungeerde model

Het principe van de geadjungeerde aanpak is eenvoudig. Beschouw bijvoorbeeld het volgende niet-lineaire model:

$$x(t_{i+1}) = M_i[x(t_i)], \quad x(t_0) = p, \quad i = 0, 1, \dots, K, \quad (2)$$

en de situatie dat er meetinformatie beschikbaar is van $x(t_K)$. We willen nu de beginvoorwaarde p schatten via het minimaliseren van de doelfunctie

$$J(x(t_K)) = J(M_{K-1}[M_{K-2}[M_{K-3}[\dots[M_0[p]]\dots]]). \quad (3)$$

We kunnen nu eenvoudig het effect δJ uitrekenen van een kleine verstoring δp in de parameter op de waarde van de doelfunctie:

$$\begin{aligned} \delta J &= \left[\frac{\partial J}{\partial x(t_K)} \right]^T \cdot \frac{\partial M_{K-1}}{\partial x(t_{K-1})} \cdot \dots \cdot \frac{\partial M_0}{\partial x(t_0)} \cdot \delta p \\ &= \left[\frac{\partial J}{\partial x(t_K)} \right]^T \cdot M(t_K, t_{K-1}) \cdot \dots \cdot M(t_1, t_0) \cdot \delta p, \end{aligned} \quad (4)$$

met $M(t_i, t_{i-1})$ het gelineariseerde model.

We kunnen dit herschrijven als een inwendig product:

$$\delta J = \left\langle \frac{\partial J}{\partial x(t_K)}, M(t_K, t_{K-1}) \cdot \dots \cdot M(t_1, t_0) \delta p \right\rangle. \quad (5)$$

Bij veel parameters is dit geen efficiënte wijze om het effect van een verstoring δp op de doelfunctie te berekenen. Bij iedere δp moet een volledige simulatie van het gelineariseerde model worden uitgevoerd.

Maar we kunnen vergelijking (5) ook schrijven als

$$\delta J = \left\langle M^*(t_i, t_0) \cdot \dots \cdot M^*(t_K, t_{K-1}) \cdot \frac{\partial J}{\partial x(t_K)}, \delta p \right\rangle \equiv \left\langle \frac{\partial J}{\partial p}, \delta p \right\rangle, \quad (6)$$

waarbij $M^*(t_i, t_{i-1})$ de geadjungeerde operator is van $M(t_i, t_{i-1})$. Aangezien $M(t_i, t_{i-1})$ een matrix is, is $M^*(t_i, t_{i-1})$ niets anders dan de getransponeerde van deze matrix. Vergelijking (6) is een zeer efficiënte uitdrukking voor de gradiënt van de doelfunctie. Met slechts één berekening van het geadjungeerde model terugwaarts in de tijd kan deze gradiënt berekend worden, ongeacht het aantal parameters.

den verkregen. Deze methode wordt 4D-Var genoemd. Een groot praktisch nadeel van variationale data-assimilatie is dat het geadjungeerde model beschikbaar moet zijn. Dit is bij zeer complexe modelsystemen een grote programmeerinspanning, vaak enkele jaren programmeerwerk. Er zijn geadjungeerde compilers beschikbaar, die als invoer de code van een voorwaarts model vragen en dan als output de code van het gelineariseerde model of van het geadjungeerde model produceren. Deze compilers werken echter alleen goed bij relatief eenvoudige modellen.

Een fundamenteel probleem is dat het optimalisatieprobleem niet convex is en er dus een risico is dat het algoritme strandt in een lokaal minimum. Voorts is het algoritme zeer rekenintensief en kunnen we slechts een beperkt aantal iteraties uitvoeren en zullen dan tevreden moeten zijn met het resultaat. Ondanks deze fundamentele problemen heeft 4D-Var zijn grote waarde in de praktijk bewezen en wordt het op grote schaal met succes toegepast.

Een alternatief voor de standaard 4D-Var aanpak is *incremental 4D-Var*. Hierbij wordt een vereenvoudigd model gebruikt voor de assimilatie. Dit vereenvoudigde model wordt gelineariseerd met als referentie de beste voorspelling van het oorspronkelijke model. Vervolgens wordt de minimalisatie uitgevoerd met dit vereenvoudigde en gelineariseerde model. Het uiteindelijke resultaat levert een nieuwe voorspelling op. Deze is echter niet optimaal omdat gebruik is gemaakt van het vereenvoudigde model. Door de procedure nu opnieuw uit te voeren met de nieuwe voorspelling als referentie, kan het resultaat nog verbeterd worden. Convergentie is hierbij helaas niet gegarandeerd. Incremental 4D-Var is de meest gebruikte methode voor het berekenen van de dagelijkse weersvoorspellingen.

Ze recent is de aandacht verschoven naar een aanpak analoog aan incremental 4D-Var, waarbij in plaats van het ontwikkelen van een nieuw vereenvoudigd numeriek model gebruik wordt gemaakt van *Proper Orthogonal Decomposition* om een vereenvoudigd model te verkrijgen. Met deze generieke modelreductie-aanpak hoeft er dus geen apart vereenvoudigd numeriek model ontwikkeld te worden. Deze aanpak is recent bijvoorbeeld gebruikt bij het schatten van de bodemtopografie van het nieuwe stormvloedvoorspelmodel van Deltares.

De bodemgegevens van een stormvloedmodel worden meestal verzameld ten behoeve van de scheepvaart en de bodemkaarten die daaruit worden afgeleid, hebben de nei-

de vorm:

$$y_i^0 = H_i[x(t_i)] + \epsilon_i, \quad i = 1, 2, \dots, K, \quad (7)$$

waarbij y_i^0 de waarnemingen zijn op tijdstip t_i . De functie H_i is de relatie tussen de metingen en de toestandsvector.

De onbekende parameters kunnen geschat worden door de volgende doelfunctie te minimaliseren:

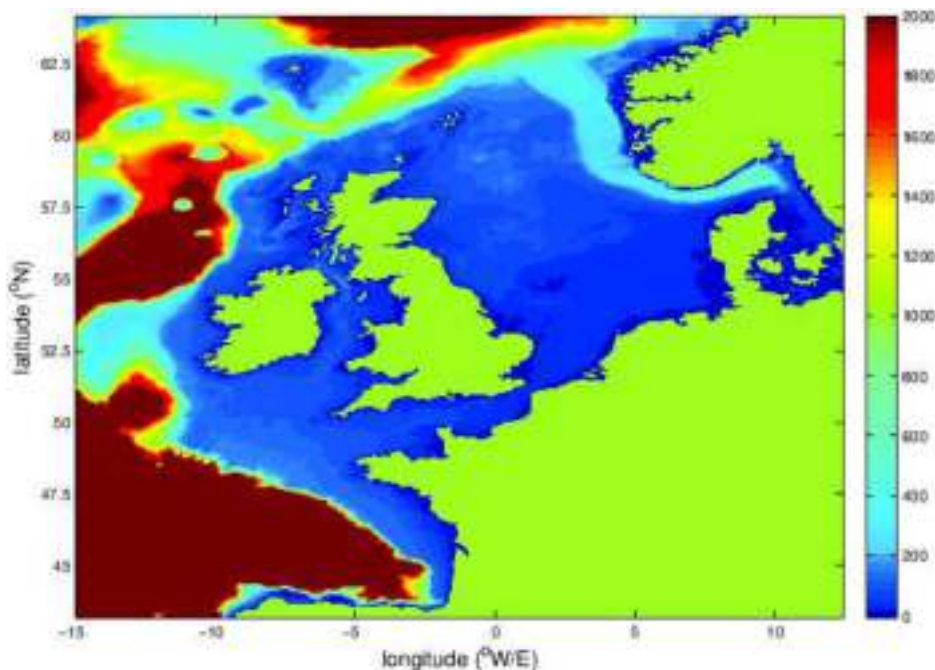
$$\begin{aligned} J(p) &= \frac{1}{2} (p - p_0)^T P_0^{-1} (p - p_0) \\ &+ \frac{1}{2} \sum_{i=1}^K (H_i[x(t_i)] - y_i^0)^T \\ &\quad \cdot R_i^{-1} (H_i[x(t_i)] - y_i^0). \end{aligned} \quad (8)$$

Hierbij is de eerste term een regularisatie-term, terwijl de tweede uitdrukking de afstand

van het model tot de beschikbare waarnemingen representeert. De modelvergelijking (1) is bij het minimaliseren van (8) een nevenvoorwaarde.

Het bovenstaande probleem is een standaard optimalisatieprobleem. Via de introductie van Lagrange-multiplicatoren of geadjungeerde toestanden kunnen de nevenvoorwaarden aan de doelfunctie worden toegevoegd. Vervolgens kan via variatierekening een eenvoudige uitdrukking van de gradiënt van de doelfunctie worden verkregen. Hierbij is dan wel de oplossing nodig van het geadjungeerde model van het gelineariseerde model (zie het kader).

Door de bovenstaande aanpak te combineren met een gradiëntmethode voor het minimaliseren van de doelfunctie, kan er een efficiënte data-assimilatiemethode wor-



Figuur 1 De bodemtopografie van het domein van het stormvloedvoorspelmodel

ging wat te ondiep te zijn. Als deze kaarten vervolgens worden gebruikt voor het bepalen van de geometrie van het numerieke stormvloedmodel zullen de getijgolven in dit model zich te langzaam voortplanten. Door gebruik te maken van de vele meetgegevens die van het getij beschikbaar zijn, kan de bodemtopografie in het model gecorrigeerd worden net zolang totdat de golfvoortplanting in het model correct is. In Figuur 1 is het domein van het nieuwe stormvloedmodel en de bodemtopografie voor calibratie weergegeven. In Figuur 2 zijn de *root-mean-square errors* in een aantal meetstations voor en na calibratie weergegeven. Bij dit calibratie-experiment zijn twintig correctieparameters geschat. De rekentijd die nodig was om dit inverse probleem op te lossen was equivalent met slechts elf evaluaties van de doelfunctie (=model simulaties). Dus behalve dat het geadjungeerde model niet nodig is, is de aanpak door deze toepassing ook zeer efficiënt gebleken.

Ensemble data-assimilatie

Data-assimilatie kan ook beschreven worden in termen van waarschijnlijkheidsdichtheden, of 'probability density functions', kortweg pdf's. Dit werkt als volgt: de onzekerheid in de modelvoorspellingen geven we weer met een pdf $p(x)$. Deze pdf is het resultaat van onzekerheden in de begintoestand van de voorspelling, in de modelvergelijkingen (die numerieke discretisaties van niet volledig bekende fysische wetten zijn), en in de randvoorwaarden. Data-assimilatie wordt nu gezien als het aanpassen van deze 'a-priori

pdf' voor de waarnemingen, resulterende in de 'a-posteriori pdf', de pdf van de mogelijke modeltoestanden gegeven die waarnemingen $p(x|y)$. In wiskundige notatie vinden we

$$p(x|y) = \frac{p(y|x)}{p(y)} p(x). \quad (9)$$

Deze relatie, het Theorema van Bayes, vormt de algemene basis voor data-assimilatie. Het geeft weer hoe onze kennis van het systeem, gegeven via de a-priori pdf $p(x)$, wordt gemodificeerd door de waarnemingen y , tot de a-posteriori pdf $p(x|y)$. De prior wordt vermenigvuldigd met de zogenaamde likelihood $p(y|x)$, die voor iedere modeltoestand x aangeeft hoe waarschijnlijk de waarnemingen y zijn.

Ten eerste valt op dat data-assimilatie nu te interpreteren is als een vermenigvuldigingsprobleem: we moeten de a-priori pdf vermenigvuldigen met de zogenaamde likelihood $p(y|x)$ om de a-posteriori pdf te vinden. De noemer $p(y)$ kan gezien worden als een normalisatiefactor omdat de waarnemingen al bekend zijn.

Ondanks het belang van het hebben van een formele oplossing voor het data-assimilatieprobleem, is de praktijk weerbarstig. Het probleem is dat de modellen die gebruikt worden in de geowetenschappen typisch zeer hoge dimensies hebben, en alleen al voor het opslaan van de a-priori pdf hebben we computers nodig met astronomisch grote geheugens. We zijn dus gedwongen benaderingen te aanvaarden. Een veelgebruikte benadering is dat de pdf Gaussisch is, zodat

we alleen de gemiddelde modeltoestand en de bijbehorende covariantie nodig hebben. Dit leidt onmiddellijk naar het zogenaamde Kalman Filter. Helaas is het voor de hoogdimensionale systemen niet mogelijk de covariantie op te slaan, dus we moeten extra aanpassingen toestaan. Een zeer populaire is het Ensemble Kalman Filter. Bij deze methode wordt de pdf weergegeven als een ensemble van modeltoestanden. Iedere modeltoestand evolueert volgens de modelvergelijkingen naar de tijd waarop we waarnemingen hebben. Daar berekenen we het gemiddelde van de Gaussische pdf als het gemiddelde van de ensembleleden, en ook de covariantie wordt berekend gebruikmakende van het ensemble van modeltoestanden.

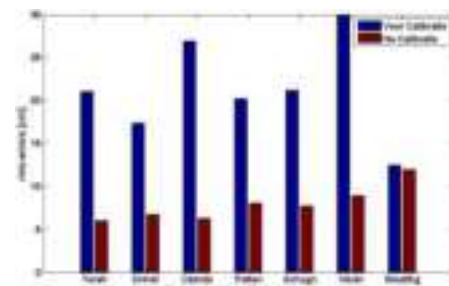
Deze methode is erg succesvol en wordt tegenwoordig gebruikt in zeer hoogdimensionale systemen. Figuur 3 geeft een voorbeeld van het gebruik van het Ensemble Kalman Filter in een model van de Atlantische Oceaan. Het model heeft meer dan een miljoen variabelen op ieder tijdstip, en we assimileren satellietwaarnemingen van de temperatuur van het zeewater en de hoogte van het zeeoppervlak. Die laatste waarnemingen geven informatie over het drukveld in de oceaan, welke zeer belangrijk is in het beïnvloeden van stromingspatronen.

Een nieuwe ontwikkeling is het loslaten van de restrictie tot Gaussische pdf's. De methode, Particle Filter genoemd, gebruikt weer een ensemble van modeltoestanden en gebruikt het Theorema van Bayes meer direct. In wiskundige termen, de a-priori pdf wordt weergegeven als een ensemble van modeltoestanden x_i , als

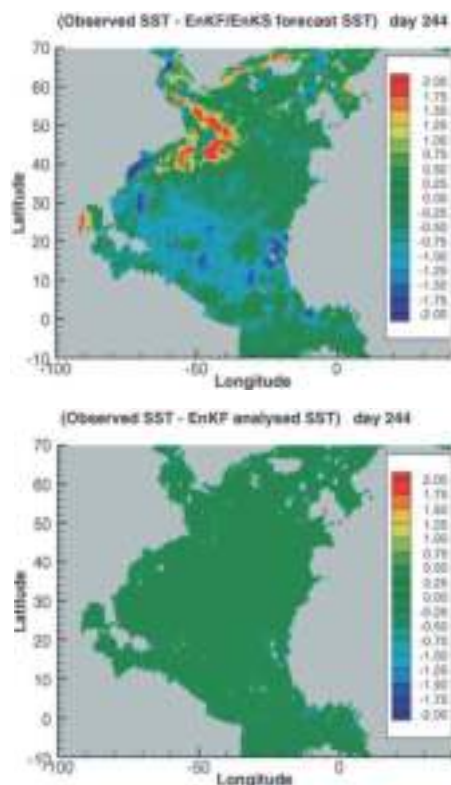
$$p(x) = \frac{1}{N} \sum_{i=1}^N \delta(x - x_i). \quad (10)$$

Als we dit direct gebruiken in het Theorema van Bayes vinden we

$$p(x|y) = \sum_{i=1}^N w_i \delta(x - x_i), \quad (11)$$



Figuur 2 De RMS error van de modelresultaten voor en na calibratie

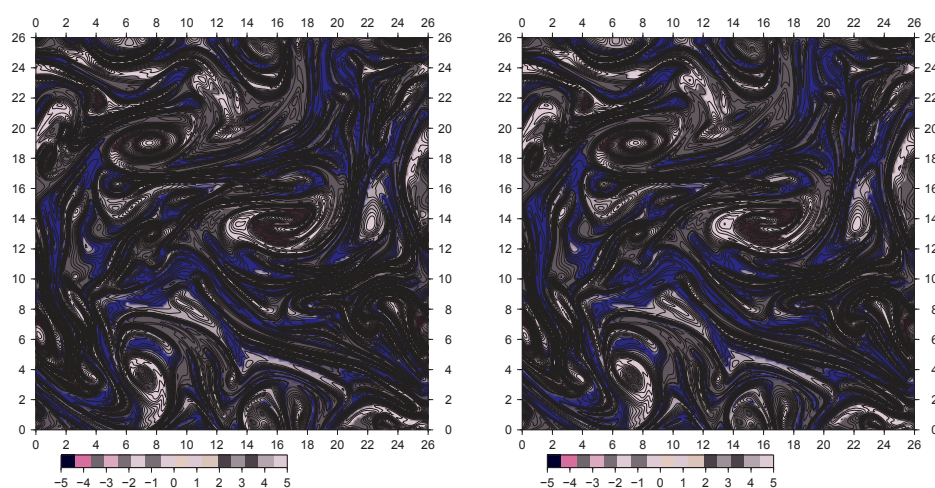


Figuur 3 Voorbeeld van succesvolle assimilatie in de Atlantische Oceaan. Het bovenste plaatje geeft het verschil tussen een 7-daagse voorspelling voor zeewatertemperatuur (SST) van het model en de waarnemingen weer op dag 244. Verschillen van 2 graden Celsius zijn zichtbaar, bijvoorbeeld voor de Amerikaanse kust. Het onderste plaatje geeft het verschil tussen de waarnemingen en het geassimileerde model weer op dag 244. De verschillen zijn minimaal, wat aangeeft dat bijna alle informatie in de waarnemingen opgenomen is in het model.

waarbij de gewichten w_i gegeven zijn als

$$w_i = \frac{p(y|x_i)}{\sum_j p(y|x_j)}, \quad (12)$$

wat laat zien dat de waarde van de gewichten w_i afhangt van hoe waarschijnlijk modeltoestand x_i is als de waarnemingen gegeven zijn als y . Hoe dichter modeltoestand x_i bij de waarnemingen is, hoe groter zijn gewicht. De formule voor de a-posteriori pdf laat zien dat de oplossing van het data-assimilatieprobleem een gewogen ensemble is. Bijvoorbeeld voor het vinden van de gemiddelde toestand nemen we het gewogen



Figuur 4 Snapshot van het vortiteitsveld van de 'waarheid' (links) en de schatting van het Particle Filter (rechts). Merk op dat de verschillen minimaal zijn, wat aangeeft dat het Particle Filter zeer goed werkt.

gemiddelde over de ensembleleden, met gewichten w_i .

Het Particle Filter als hierboven weergegeven is niet erg effectief voor systemen met dimensies groter dan 10. Het probleem is dat de gewichten te veel uit elkaar liggen, met typisch een van de gewichten dicht bij 1, en al de andere gewichten zeer dicht bij nul. Het gewogen ensemble bestaat dan effectief uit een modeltoestand, en alle informatie over de pdf is verloren. Gelukkig is de ontwikkeling van efficiëntere Particle Filters in volle gang. Er bestaan nu methoden die efficiënt zijn voor systemen met zeer grote dimensie. Een voorbeeld is het zogenaamde Equivalent-Weight Particle Filter. In dit filter worden de modeltoestanden zo aangepast dat ze bijna gelijke gewichten hebben, terwijl we toch het data-assimilatieprobleem netjes oplossen.

Figuur 4 geeft een voorbeeld van het toepassen van het Equivalent-Weights Particle Filter in een turbulente sterk niet-lineair systeem van cyclonale en anticyclonale wervels zoals die voorkomen in de oceaan. Dit voorbeeld is van een test van de methode waarbij de waarnemingen verkregen zijn van een modelrun met verstoorde begincondities en realisaties van modelfouten. De dimensie van het systeem is 65.000, en in dit voorbeeld is het vortiteitsveld waargenomen op

ieder roosterpunt iedere 50 modeltijdstappen. De decorrelatietijd van het vortiteitsveld is 25 modeltijdstappen, wat aangeeft dat dit een moeilijk en sterk niet-lineair data-assimilatieprobleem is. Figuur 4 laat zien dat de waarheid en de schatting van het gemiddelde van de ensembleleden van het Particle Filter zeer dicht bij elkaar liggen. Naast andere statistische technieken geeft dit vertrouwen in de data-assimilatiemethode.

Tot slot

Numerieke modellen worden steeds beter en steeds vaker gebruikt voor praktische toepassingen. Er wordt ook steeds meer meetinformatie ingewonnen. Het gevolg is dat het gebruik van data-assimilatiemethoden de afgelopen tijd zeer sterk is toegenomen. Deze ontwikkeling is nog verder versneld doordat er inmiddels enkele goede ensembletechnieken beschikbaar zijn die vrij eenvoudig kunnen worden gekoppeld aan bestaande numerieke modelsystemen. Van veel ensemblealgoritmen is ook open source code te downloaden (zie www.openda.org) zodat het benodigde programmeerwerk heel beperkt is. Daarnaast zijn er ook nog vele wetenschappelijke uitdagingen om de methoden efficiënter en nauwkeuriger te maken, met name voor sterk niet-lineaire problemen.

Referenties

- 1 M.U. Altaf, M. Verlaan en A.W. Heemink, Efficient identification of uncertain parameters in a large-scale tidal model of the European continental shelf by proper orthogonal decomposition, *Int. J. Numer. Meth. Fluids* 68 (2012), 422–450.
- 2 K. Brundal, J.M. Brankart, G. Halberstadt, G. Evensen, P. Brasseur, P.J. van Leeuwen, E. Dombrowsky en J. Verron, A demonstration of ensemble-based assimilation methods with a layered OGCM from the perspective of operational ocean forecasting systems, *Journal of Marine Systems* 40–41 (2003), 253–289.
- 3 P. Courtier, J.N. Thepaut en A. Hollingsworth, A strategy for operational implementation of 4D-VAR, using an incremental approach, *Quarterly Journal of the Royal Meteorological Society* 120 (1994), 1367–1387.
- 4 G. Evensen, *Data Assimilation: The Ensemble Kalman Filter*, Second Edition, Springer, 2009.
- 5 F.X. Le Dimet en O. Talagrand, Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects, *Tellus* 38 (1986), 97–110.
- 6 P.J. van Leeuwen, Particle Filtering in Geophysical Systems, *Monthly Weather Review* 137 (2009), 4089–4114.
- 7 P.J. van Leeuwen en M. Ades, Efficient fully non-linear data assimilation for geophysical fluid dynamics, *Computers and Geosciences* 55 (2013), 16–27.

Theo Gerkema

NIOZ, Den Burg

theo.gerkema@nioz.nl

Leo Maas

NIOZ, Den Burg

IMAU, Universiteit Utrecht

leo.maas@nioz.nl

Wiskundige aspecten van interne golven

Het bestaan van golven aan het zeeoppervlak is wel bekend. Van het voorkomen van golfbewegingen onder het oppervlak is men zich vaak veel minder bewust. De amplitudes van deze golfbewegingen kunnen toch behoorlijk groot zijn (tot wel honderd meter). Vanwege hun transport van warmte en impuls, en van zuurstof, plankton en mineralen, zijn interne golven van belang voor de algemene oceaancirculatie en voor de mariene voedselketen. Theo Gerkema en Leo Maas laten aan de hand van een combinatie van fysische en mathematische methoden zien dat deze golven zich heel anders gedragen dan de golven aan het vrije oppervlak. Dit gedrag wordt sterk beïnvloed door de dichtheidsstructuur in de verticale richting en de rotatie van de aarde, en resulteert erin dat interne golven gestuurd worden naar heel bepaalde locaties, in zogenaamde golf aantrekkers.

In het inwendige van de oceaan zijn golfbewegingen alom aanwezig; deze worden *interne golven* genoemd. Ze hebben grote verticale uitwijkingen (orde 10 tot 100 m) en perioden variërend van minuten tot dagen. De terugdrijvende krachten verantwoordelijk voor de golfbewegingen zijn van tweeërlei aard: ten eerste zorgen de verticale dichtheidsverschillen voor een effectieve zwaartekracht; ten tweede geeft de dagelijkse rotatie van de aarde rond haar as aanleiding tot een afbuigende kracht, de zogenaamde corioliskracht.

Deze bepalen niet alleen het bestaan van interne golven, maar ook hun ongebruikelijke wiskundige eigenschappen. Ze maken de wiskundige behandeling weliswaar soms lastig, maar bevatten ook de sleutel tot de ongebruikelijke eigenschappen van de golven zelf. Zo zijn dus de wiskundige en fysische eigenaardigheden nauw verstrengd. In het hiernavolgende zetten we dit nader uiteen.

Symmetrie-brekers

Allereerst is het van belang een beeld te krijgen van het medium waarin de interne golven zich voortplanten. Figuur 1a toont een

voorbeeld van de structuur van de temperatuur in de Stille Oceaan, gaande van de Fiji-eilanden (zuidelijk halfrond) tot aan de Beringstraat (noordelijk halfrond). Duidelijk is te zien dat op grotere diepten, vanaf 1000 m onder het oppervlak, de temperatuur tamelijk homogeen is en steeds heel laag (minder dan 5°C), ook in de bovenste paar honderd meter neemt de temperatuur scherp toe naarmate men dichterbij het oppervlak komt. Een ander voor dit verhaal belangrijk aspect, is het verloop van de diepte. We zien dat de oceaانبodem verre van vlak is, maar vol zit met bergruggen en onderzeese heuvels (die een vulkanische oorsprong hebben). Terwijl dus het temperatuurverloop min of meer beschouwd kan worden als een op een stapeling van horizontale lagen, breekt de oceaانبodem deze horizontale symmetrie.

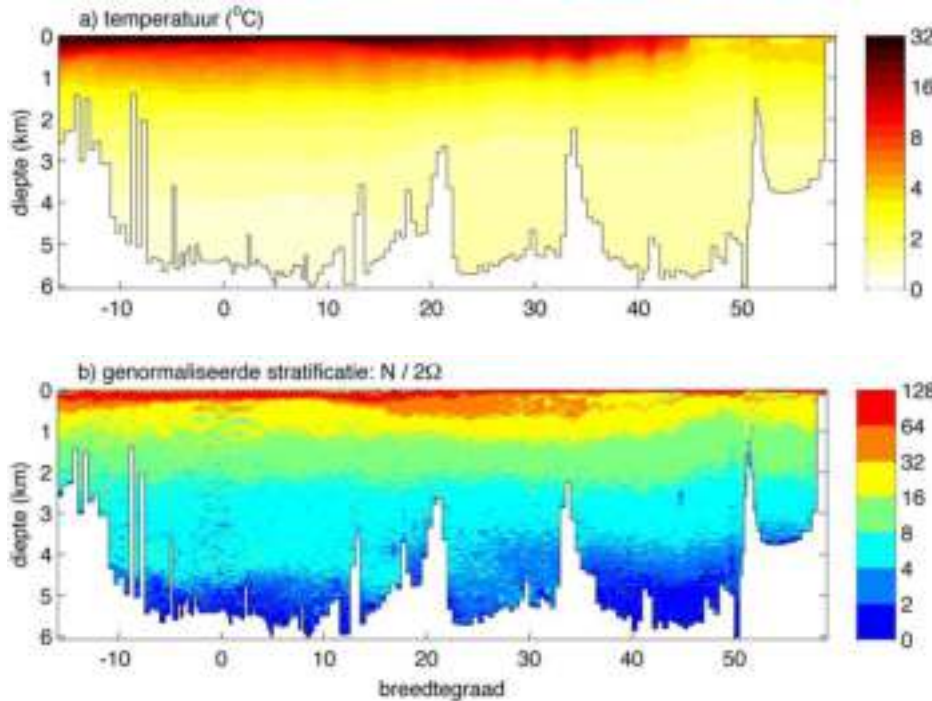
In de oceanografie en meteorologie is het gebruikelijk de verticale gelaagdheid uit te drukken in termen van de zogenaamde Brunt-Väisälä-frequentie (stabiliteitsfrequentie) N , hierna kortweg aangeduid als BV-frequentie. Deze wordt bepaald door de verticale dichtheidsgradiënt; hierbij is de dicht-

heid ρ , in het geval van oceaanwater, een functie van de toestandsvariabelen temperatuur, zoutgehalte en druk (p). De precieze definitie luidt:

$$N^2 = g^2 \left(\frac{d\rho}{dp} - \frac{1}{c_s^2} \right). \quad (1)$$

Hierin is g de valversnelling en c_s de geluidssnelheid in zeewater (zelf eveneens een functie van temperatuur, zoutgehalte en druk). Deze laatste term hangt samen met de samen-drukbaarheid van zeewater, en is vooral van belang in de diepere lagen, waar de eerste term (de verticale gradiënt van dichtheid) gering is. De verdeling van N in de oceaan is te zien in Figuur 1b; N omspannt twee orden van grootte, is gering in de diepere lagen en zeer hoog in de warme dunne bovenlaag.

In Figuur 1b is N gedeeld door een constante 2Ω , waarbij $\Omega = 7.292 \times 10^{-5}$ radianen per seconde, de hoeksnelheid van de aarde (dat wil zeggen: 2π gedeeld door de *siderische* dag, 23h 56 min 4s, de tijd waarin zich ten opzichte van de sterren één omwenteling voltrekt). De aardas zelf (zwarte pijl in Figuur 2) breekt lokaal de horizontale symmetrie. Bezien vanuit een positie op een zekere breedte ϕ , is de oriëntatie van de aardas noch loodrecht op, noch parallel aan het aardoppervlak. Deze symmetriebreking manifesteert zich in een van de componenten van de corioliskracht, de kracht die bewegende deeltjes op de roterende aarde doet afbuigen in een richting loodrecht op hun snelheid (voor een nadere uiteenzetting hierover verwijzen we naar [3]).

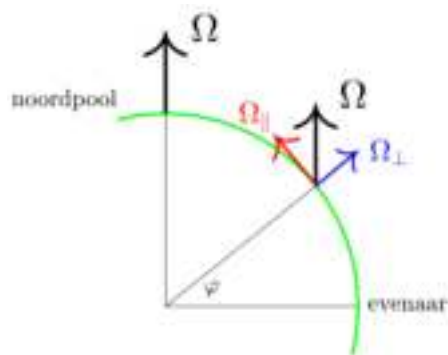


Figuur 1 Een voorbeeld van de structuur van het inwendige van de oceaan: een zuid-noordsectie in de Stille Oceaan, lopend van de Fiji-eilanden (links) tot de Beringstraat (rechts). In (a), temperatuur; in (b) de geschaalde BV-frequentie, een maat voor de gravitationele stabiliteit van de waterkolom. (Data uit WOCE sectie P14)

De symmetrie in de overwegend horizontale gelaagdheid en de verticale zwaartekracht wordt dus op twee manieren gebroken: door de lokaal ‘dwarse’ aardas en door de hellende bodem. Wiskundig betekent dit dat het probleem van de propagatie van interne golven zich niet simpel laat scheiden in een vergelijking voor de horizontale en verticale propagatie afzonderlijk, iets wat zonder de symmetriebreking wel mogelijk zou zijn. Deze niet-separeerbaarheid verhoogt de wiskundige complexiteit van het probleem. Toch staat dit een oplossing niet altijd in de weg, zoals hieronder zal blijken.

Aardas

We bekijken hier het probleem van de corioliskracht afzonderlijk, en nemen aan dat



Figuur 2 Een doorsnede van een deel van de aardbol, met in zwart de rotatievector, en een ontbinding daarvan in een loodrechte en parallelle component op breedtegraad ϕ .

de bodem horizontaal is. De beweging van het water wordt beschreven door de tweede wet van Newton, toegepast op vloeistoffen, tezamen met de wetten voor massa- en energiebehoud. Om de behandeling te vereenvoudigen, maken we een aantal benaderingen: wrijving wordt verwaarloosd en ook wordt aangenomen dat de golven voldoende klein van amplitude zijn om als lineair behandeld te kunnen worden. Tenslotte veronderstellen we, dat de BV-frequentie N enkel in verticale richting varieert; Figuur 1b geeft een indicatie hoe deze afhankelijkheid er in de oceaan uitziet.

Voor sinusoidale golven van frequentie ω kan men voor het ruimtelijke deel van de stroomfunctie, $\psi(x, z) \exp(-i\omega t)$, dan de volgende vergelijking afleiden:

$$A \frac{\partial^2 \psi}{\partial x^2} + 2B \frac{\partial^2 \psi}{\partial x \partial z} + C \frac{\partial^2 \psi}{\partial z^2} = 0. \quad (2)$$

Hierin is x de horizontale coördinaat en z de verticale. Wiskundig interessant en ook ongebruikelijk is de aanwezigheid van de gemengde afgeleide — de tweede term — die wijst op een verstrengeling van horizontale en verticale afhankelijkheden. Anders gezegd, de vergelijking is *niet-separeerbaar*; we kunnen de oplossing niet schrijven als een product van reële functies van de onafhankelijke variabelen, $\psi = f(x)g(z)$.

Wat is de oorsprong van deze kruisterm?

Dit is direct te zien aan de definitie van de coëfficiënten:

$$A = N^2 - \omega^2 + \tilde{f}^2; \quad B = f\tilde{f}; \quad C = f^2 - \omega^2,$$

met $f = 2\Omega \sin \phi$ en $\tilde{f} = 2\Omega \cos \phi$. Deze laatste, $\tilde{f}$, wordt in de geofysische stromingsleer meestal verwaarloosd (de zogenaamde ‘Traditionele Benadering’), en we zien dat het effect van deze benadering is dat de kruisterm in (2) verdwijnt; het probleem wordt daardoor separeerbaar.

We zijn hier echter geïnteresseerd in wat er gebeurt wanneer men deze term *niet* verwaarloost. Hoewel (2) in haar volledige vorm niet-separeerbaar is, kunnen we toch gemakkelijk tot een oplossing komen. Wanneer we namelijk substitueren

$$\psi = \Psi(z) \exp i(kx + \delta z), \quad (3)$$

met $\delta = -kB/C$ en (vooralsnog willekeurig) golfgetal k , dan krijgen we

$$\Psi'' + k^2 \left[\frac{B^2 - AC}{C^2} \right] \Psi = 0, \quad (4)$$

waarbij het accent de afgeleide naar z weergeeft. Merk op dat deze vergelijking geheel reëel is. Wanneer we dus het reële deel nemen van $\psi(x, z) \exp(-i\omega t)$, krijgen we

$$\psi = \Psi(z) \cos(kx + \delta z - \omega t).$$

Het niet-separeerbare karakter, de verstrengeling van x en z , toont zich in het argument van de cosinus.

Wat betekent dit concreet voor de golven zelf? We zullen kijken naar twee aspecten: de dispersie-relatie en de verticale structuur van de golven.

Wat betreft het eerste, zullen we veronderstellen dat N een constante is. Hoewel onrealistisch (zie Figuur 1b), heeft dit het voordeel dat we het probleem geheel analytisch kunnen oplossen. We leggen als randvoorwaarde op dat de stroomfunctie nul moet zijn aan bodem en oppervlak ($z = -H, 0$), wat neerkomt op de natuurlijke eis dat daar de verticale (dat wil zeggen loodrechte) snelheid nul moet zijn. De oplossing van (4) luidt dan: $\Psi = \sin n\pi z/H$ met $n = 1, 2, 3, \dots$, waarbij tevens moet gelden:

$$k^2 = \left(\frac{n\pi}{H} \right)^2 \frac{C^2}{B^2 - AC}, \quad (5)$$

waarmee het verband is gelegd tussen het

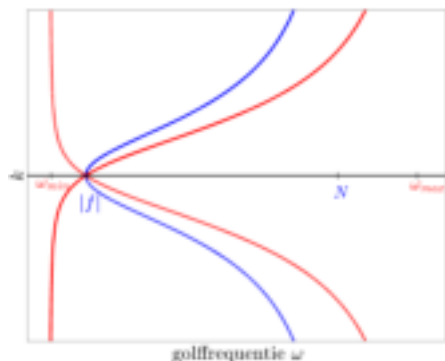
golfgetal k en de golffrequentie ω (bevat in A en C) — de dispersie-relatie.

Hoe ziet dit verband eruit? Duidelijk is dat k naar oneindig gaat (dat wil zeggen de golven worden *kort*) wanneer $B^2 - AC$ naar nul nadert. Deze laatste bevat de gegeven constanten van het medium (N , f en $\tilde{f}$) en de golffrequentie ω . Wegens het kwadratische voorkomen van ω in A , zullen er twee waarden van de frequentie zijn waarvoor $B^2 - AC \rightarrow 0$. Dit is geïllustreerd met de rode lijn in Figuur 3: wanneer de frequentie ω haar maximum nadert, gaat het golfgetal, in absolute zin, naar oneindig; maar dit is óók het geval wanneer de frequentie haar minimum nadert. Dit is ongebruikelijk, want doorgaans vallen in de geofysische stromingsleer lage frequenties samen met *kleine* golfgetallen, niet met grote! De sleutel voor dit vreemde gedrag ligt in de kruis-term met de gemengde afgeleide. Stellen we namelijk B nul dan valt C weg uit de noemer van (5); golfgetallen gaan dan alleen nog naar oneindig wanneer $\omega = N$, voor hoge frequenties (zie blauwe lijn in Figuur 3).

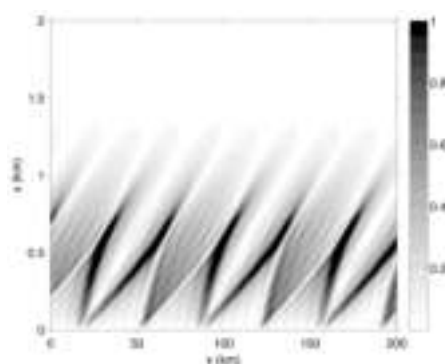
We zien hier dat de niet-separeerbaarheid een singulier karakter heeft: voor eindige B , hoe klein ook, vinden we steeds hogere golfgetallen naarmate de frequentie haar minimum nadert; stellen we $B = 0$, dan gaat ineens het golfgetal naar nul bij de laagste frequentie.

Voor de oceaan heeft dit mogelijk interessante implicaties, want een van de kernvragen uit de fysische oceanografie is hoe de energie van de lage frequenties en grote ruimteschalen naar de kleine ruimteschalen gaat waar de dissipatie plaatsvindt. In de hierboven beschreven eigenschappen zou deze overgang al in de golfeigenschappen zelf bevat zijn!

Tenslotte vormen we ons een beeld van die golfpropagatie. We kijken daartoe naar een iets realistischer geval van N en nemen deze nu niet constant maar laten N^2 lineair met



Figuur 3 Dispersierelatie met de niet-traditionele component (rood), en zonder (blauw). Ontleend aan [2].



Figuur 4 Amplitude van een golfoplossing voor een sub-inertiale frequentie (dat wil zeggen $\omega \leq |f|$) bestaande in het onderste, zwak gestratificeerde deel van de waterkolom.

z variëren, afnemend met de diepte. De oplossing van (4) kan dan in termen van Airy-functies worden verkregen. Voor lage frequenties krijgen we dan een golf die gebonden is aan de bodem en boven een bepaalde hoogte exponentieel verval, zie Figuur 4. Het patroon geeft de plaatsen weer waar verticale snelheden hoog zijn en de golfenergie plant zich daarlangs voort. Duidelijk is de symmetriebreking ten opzichte van de verticaal een gevolg van de kruis-term met B . Deze term speelt in dit geval een nog fundamentele rol: voor $B = 0$ zou de hier afgebeelde golf, waarvan de frequentie lager is dan $|f|$, zelfs helemaal niet bestaan!

Hellende bodems

Een alternatief voor het zoeken naar gesloten-vorm oplossingen, waarin de oplossing een expliciete afhankelijkheid van de onafhankelijke variabelen — hierboven x en z — kent, bestaat in het algoritmisch oplossen van de vergelijking. Hier zullen we deze methode toelichten voor het geval de afmeting van het uniform-gelaagde ($N = \text{const}$) vloeistofbasin zo klein is dat rotatie van de aarde verwaarloosbaar is ($f = \tilde{f} = 0$ en in het bijzonder $B = 0$). Denk bijvoorbeeld aan het relatief kleine, door temperatuurverschillen nagenoeg uniform-gelaagde Loch Ness, een fjord dat we hierna als ‘kanaal’ zullen typeren. Aan het einde van de ijstijd bezorgden terugtrekkende gletsjers dit kanaal in dwarsrichting een trapezoidale vorm. Omwille van de eenvoud zullen we naar zo’n kanaal kijken waar de bodem, $z = -H(x)$, voor $-L < x < 0$ een constante diepte, H_0 , heeft, die voor $0 \leq x < L$ lineair afneemt, $z = H_0(x/L - 1)$. Hier is H_0 de maximale diepte en L de halve kanaalbreedte. Het ruimtelijk patroon van vrije golven van frequentie ω wordt beschreven door een stroomfunctie $\psi(x, z)$ die nul moet zijn op alle randen.

Als we de horizontale coördinaat x schalen met L en de verticale coördinaat z met $D \equiv L(N^2/\omega^2 - 1)^{-1/2}$, dan vereenvoudigt (2) tot

$$\psi_{xx} - \psi_{zz} = 0. \quad (6)$$

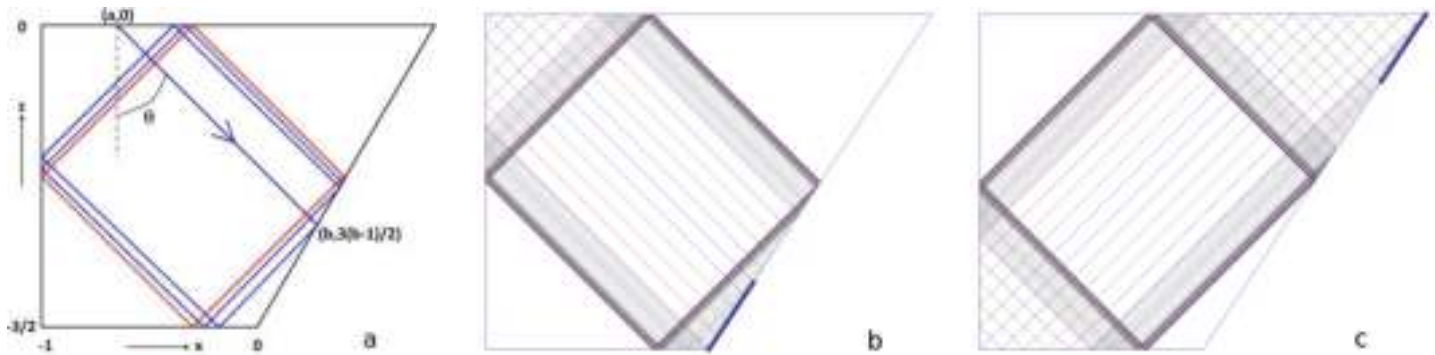
Dit probleem lijkt samen met de eerdergenoemde randvoorwaarde zo schraal dat men wellicht denkt dat er geen niet-triviale oplossingen zijn. Dit is echter wel degelijk het geval, en wel voor bijna iedere frequentie $\omega < N$. Het probleem hangt nog van één parameter af. Dit is de met D geschaalde kanaaldiepte H_0 , ofwel dimensieloze diepte $\tau = H_0/D = (N^2/\omega^2 - 1)^{1/2} H_0/L$. Volledigheidshalve luiden de randvoorwaarden in het trapezoidale kanaal dan ook $\psi = 0$ aan het oppervlak, $z = 0$, aan de verticale zijwand, $x = -1$, aan de bodem $z = -\tau$, voor $-1 < x < 0$ en, voor $0 \leq x < 1$, op de hellende bodem, $z = \tau(x - 1)$.

Een algemene oplossing die voldoet aan de oppervlakterandvoorwaarde is formeel te schrijven als

$$\psi = F(x + z) - F(x - z). \quad (7)$$

‘Formeel’ omdat $F(\xi)$ een nog onbekende functie van karakteristieke coördinaten $x \pm z$ is, wiens precieze vorm nadere bepaling vereist, gebaseerd op de onderliggende geometrische structuur van de karakteristieken. In de gestrekte verticale coördinaat (z) bewaren karakteristieken altijd dezelfde hoek van 45 graden ten opzichte van de zwaartekracht. Als twee karakteristieken (golfstralen) daarom parallel, van boven af invallen op een hellende bodem (Figuur 5a) zullen zij na weerkaatsing nog steeds parallel zijn, maar zal hun afstand kleiner worden (focusering). Als de karakteristieken in omgekeerde richting worden gevolgd zal deze afstand juist toenemen (defocusering). Het blijkt dat gesloten gebieden — waar een vloeistof zich noodzakelijkerwijs altijd in bevindt — door focusering gedomineerd worden. Dit impliceert dat golfstralen uiteindelijk bijna altijd een periodieke baan zullen bereiken, een *golfaantrekker* genoemd (weergegeven door de rode lijn in Figuur 5a). Dit geometrisch aspect van de oplossing is, zoals we zullen zien, ook in het stroomfunctieveld zichtbaar.

De algoritmische oplosmethode [5] bestaat daarin dat we de functionele afhankelijkheid van $F(\xi)$ niet pogen te raden maar, bescheidener, eerst veronderstellen dat we diens waarde op een of andere wijze in één



Figuur 5 (a) Trapezoïdaal bassin ter diepte $\tau = 3/2$ met half karakteristiek web startend in $(a, 0)$ (blauwe lijn) en golfaantrekker (rood) van een intern golfveld in een uniform-gelaagde vloeistof. (b) en (c) Fundamentele intervallen (dik lijnstuk op hellende bodem) en daarvandaan geconstrueerde karakteristieke webben.

punt kennen. Laten we dat punt aan het oppervlak kiezen als $\mathbf{x}_0 = (a, 0)$, zodat we veronderstellen dat $F(a)$ bekend is, zeg $F(a) = F_*$, zie Figuur 5a. De algemene oplossing laat ons toe te concluderen dat op de naar rechtsonder hellende lijn $z + x = a$, $F(x + z)$ vanzelfsprekend ook constant is. Als we nu het snijpunt van deze lijn en de bodem bepalen, gegeven door $\mathbf{x}_1 = (b, 3/2(b - 1))$, met $b = (2a + 3)/5$, dan kunnen we ook de specifieke, naar links-onder hellende reflecterende karakteristiek bepalen die door dat punt gaat. Op die karakteristiek, $x - z = 3/2 - b/2$, is F alweer constant, namelijk $F(3/2 - b/2)$. Omdat het snijpunt op de rand ligt moet ψ daar nul zijn. Derhalve volgt uit (7) dat $F(3/2 - b/2) = F(a) = F_*$. Met andere woorden, F is *invariant* onder reflectie van die karakteristiek aan de wand. Aangezien dit geldt voor *alle* daaropvolgende en ook voor alle eerdere reflecties, dus voor de gehele verzameling reflectiepunten $\{\dots, \mathbf{x}_{-1}, \mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \dots\}$, concluderen we dat voor het gehele *karakteristieke web* — het web van via reflecties aan elkaar verbonden karakteristieken, de *karakteristieke baan* — $F(X)$ invariant is en wel gelijk aan F_* . In Figuur 5a is de helft van het web geplot. De andere helft volgt door de karakteristiek vanuit $(a, 0)$ naar linksonder te volgen. Tegelijkertijd zien we dat dit ene web niet het gehele kanaal bedekt. Maar hoeveel webben zijn daarvoor nodig, waar kunnen we die web-invariant per web kiezen, en wat is de betekenis van die web-invariant eigenlijk?

In een trapezoïdaal kanaal blijken op de rand twee intervallen te liggen, *fundamentele intervallen* genaamd, waarvan de randen worden bepaald door zogenaamde *kritische punten* van de domeinrand. Dit zijn punten waar de raaklijn aan de rand onder eenzelfde hoek met de verticaal staat als de karakteristieken (vanwege de schaling, hier dus een hoek van 45 graden). De bijzondere aard van deze punten volgt uit het feit dat een invallende karakteristiek (een *kritische* karakteristiek)

gedwongen is terug te keren volgens hetzelfde pad, in plaats van, als gewoonlijk, zich in een richting daar loodrecht op voort te zetten. Voor het trapezoïdale domein vallen deze kritische punten samen met de vier hoekpunten.

Wanneer in beide fundamentele intervallen ieder punt als startpunt $\mathbf{x}_0$ van een karakteristiek web gekozen wordt, dan zullen de aldus verkregen karakteristieke webben gezamenlijk het hele kanaal op een unieke manier bedekken. Op de hellende bodem geven de in Figuur 5b en 5c doorgetrokken dikke lijnelementen deze fundamentele intervallen weer. Voor een kanaal van dimensieloze diepte $\tau = 3/2$ zijn dat de intervallen $x \in (0, 1/5)$ en $(4/5, 1)$. De karakteristieke webben die vanuit beide fundamentele intervallen bereikt worden zijn complementair aan elkaar. Zij naderen allen de voor dit geval vierkante golfaantrekker, en doen dat, opmerkelijkerwijs, voor beide flanken van het web, $x_{\pm\infty}$, met dezelfde kloksgewijze oriëntatiezin.

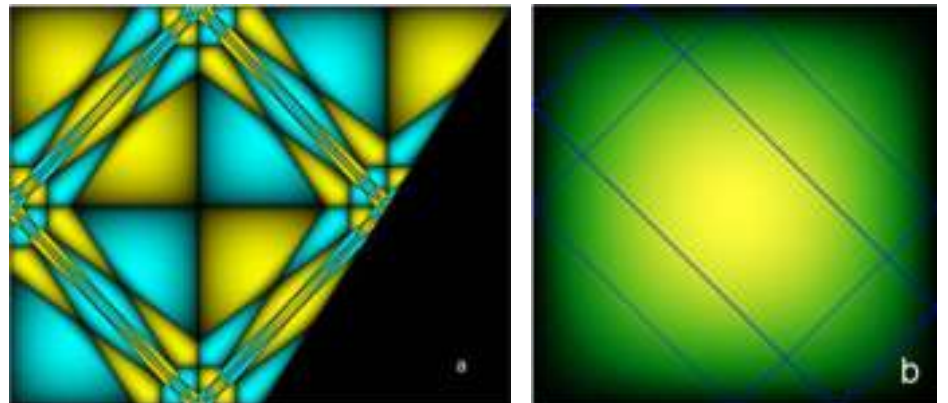
Ieder punt in het kanaal ligt op het kruispunt van precies twee karakteristieke lijnen. Als daarom $F(\xi)$ op beide fundamentele intervallen (willekeurig) wordt voorgeschreven, dan is de stroomfunctie uniek bekend. Door in deze gebieden voor $F(\xi)$ bijvoorbeeld een halve cosinus te kiezen — in het voorbeeld hierboven $F(x) = \cos(5\pi x)$ — construeren we

het stroomfunctiepatroon zoals getoond in Figuur 6a.

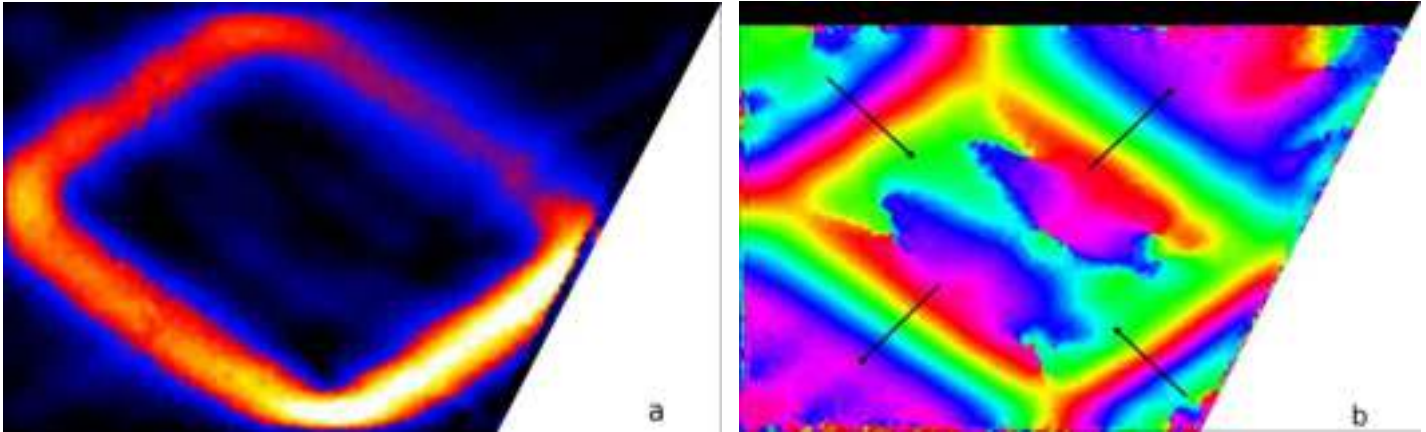
Bij constructie is dit een vrije, maar staande golfoplossing. Omdat door de reële keuze van $F(\xi)$, het ruimtelijke patroon reëel is, zal deze oplossing in de tijd slechts ‘knippen’. Dit lijkt de betiteling van de periodieke baan als ‘golfaantrekker’ niet te rechtvaardigen. Wanneer we op een fundamenteel interval F echter *complex* kiezen (hetgeen een randforcering van interne golven voorstelt) zullen deze golven de golfaantrekker daadwerkelijk benaderen.

Maar wat stelt F nu precies voor? De fysische betekenis van F volgt uit het feit dat storingsdruk, $p(x, z)$, net als stroomfunctie ψ moet voldoen aan vergelijking (6). Diens oplossing, $p = F(x + z) + F(x - z)$, wordt gegeven in termen van precies dezelfde functie F . Dit identificeert F als een partiële druk.

Merk op dat we ‘stroomfunctie’ zelf niet kunnen meten. Wel echter diens afgeleide, de snelheid. Hoewel ψ begrensd is, laat Figuur 6a zien dat bij nadering van de vierkante golfaantrekker diens gradiënt alsmaar toeneemt. Dit suggereert dat deze golven zich rond de aantrekker zullen manifesteren als intense gelokaliseerde beweging. Figuur 7 toont experimentele bevestiging van dit vermoeden in de hierboven besproken configuratie.



Figuur 6 Stroomfunctiepatroon in (a) trapezoïde, diepte $\tau = 3/2$; (b) vierkant. Lijnen in (b) zijn enkele periodieke webben.



Figuur 7 Experimenteel bepaalde (a) amplitude en (b) fase van verstoringen van het dichtheidsveld in een interne golf, die ontstaan door de bak horizontaal te schudden [4]. Deze verstoringen volgen het snelheidsveld (de gradiënt van de stroomfunctie). In (b) geven de pijlen de richting aan waarin de fase in de tijd loopt.

Golfaantrekkers treden niet alleen op in een kanaal van trapezoidale vorm, maar voor bijna elke bodemvorm die de reflectiesymmetrie van kaatsende golfstralen ‘voldoende breekt’. Met dat laatste voorbehoud bedoelen we dat sommige domeinvormen, zoals de rechthoek, aan de dominantie van focussering weten te ontsnappen (zie Figuur 6b). Voor rationale dieptes $\tau = 2m/n$, ($m, n \in \mathbb{N}$) is de stroomfunctie simpelweg in separeerbare vorm te vinden,

$$\psi = \sin(n\pi(x+1)/2) \sin(m\pi z/\tau). \quad (8)$$

In dat geval is iedere reflectie van twee parallelle golfstralen neutraal: er treedt noch focussering noch defocussering op, althans, als de zijdes van de rechthoek tenminste loodrecht op (of parallel aan) de zwaartekracht staan (in [1] wordt een tegenvoorbeeld gegeven). Voor de niet-gekanтелеde rechthoek ligt iedere golfstraal op zijn eigen periodieke web en bijgevolg kent zo’n vloeistofdomein oneindig veel periodieke karakteristieke webben (zie Figuur 6b). In andere uitzonderlijke gevallen kan focussering geneutraliseerd worden door een precies even sterke mate van defocussering, zodanig dat iedere golfstraal weer op zijn ei-

gen periodieke web ligt (bijvoorbeeld in een elliptisch domein waarvan één van de symmetrieassen loodrecht op de zwaartekracht staat). Maar in het algemeen zullen golfaantrekkers de respons domineren en kunnen stroomfunctie en druk met behulp van de hierboven geschetste geometrische methode geconstrueerd worden.

De hier besproken singuliere oplossing in een trapezoidaal bekken voor $\tau = 3/2$ is interessant omdat voor dit speciale geval de stroomfunctie ook analytisch bekend is [6]:

$$\psi = \sum_{m=1}^{\infty} a_{2m} \sin\left(4m\pi \frac{(x+1)}{3}\right) \cdot \sin\left(4m\pi \frac{z}{3}\right), \quad (9)$$

met Fourier-coëfficiënten

$$a_{2m} = \frac{2m(-1)^{m+1}}{\pi} \sum_{n=0}^{\infty} \sin\left(\frac{m\pi}{2b_n}\right) \cdot \left(\frac{1}{b_n^2 - m^2} - \frac{1}{(5b_n)^2 - m^2}\right),$$

waar

$$b_n = \frac{3}{2} 5^n.$$

De individuele termen in de Fourier-reeks (9) voldoen afzonderlijk aan de ruimtelijke golfvergelijking en aan de randvoorwaarden op het horizontale deel van de bodem, de verticale zijwand en het oppervlak. Ze zijn in wezen stuk voor stuk identiek aan de separeerbare modi in de rechthoek (8). Maar hun koppeling en daarmee hun *niet*-separeerbaarheid (en de bepaling van de a_{2m}) wordt door de laatste randvoorwaarde op de schuine helling bepaald. Deze is te schrijven als een functionaalvergelijking die op algebraïsche wijze het bestaan van de eerder geometrisch afgeleide fundamentele intervallen bevestigt.

Conclusie

Dichtheidsgelaagdheid en rotatie verlenen de oceaan een anisotroop karakter. Als gevolg hiervan gedragen interne golven in de oceaan zich heel anders dan oppervlaktegolven. De anisotropie leidt ertoe dat de ruimtelijke vergelijking met randvoorwaarden in het algemeen geen separeerbare oplossingen kent. Dit heeft verreikende gevolgen voor de dispersie van deze golven en voor hun scheve propagatie. Binnen de kaders van deze theorie leidt dit uiteindelijk bijna altijd tot hun onderschepping door golfaantrekkers. $\leftarrow$

Referenties

- 1 J. Bajars, J. Frank en L.R.M. Maas, On the appearance of internal wave attractors due to an initial or parametrically excited disturbance, *J. Fluid Mech.* 714, 2013, 283–311.
- 2 T. Gerkema en V.I. Shrira, Near-inertial in the ocean: beyond the ‘traditional approximation’, *J. Fluid Mech.* 529, 2005, 195–219.
- 3 T. Gerkema en L. Gostiaux, A brief history of the Coriolis force, *Eur. Phys. News* 43/2, 2012, 14–17.
- 4 J. Hazewinkel, P. van Breevoort, S.B. Dalziel en L.R.M. Maas, Observations on the wave number spectrum of an interval wave attractor in a two-dimensional, *J. Fluid Mech.* 598, 2008, 373–382.
- 5 L.R.M. Maas en F.-P.A. Lam, Geometric focusing of internal waves, *J. Fluid Mech.* 300, 1995, 1–40.
- 6 L.R.M. Maas, Exact analytic self-similar solution of a wave attractor field, *Physica D – Nonlinear Phenomena* 238, 2009, 502–505.

Henk Schuttelaars*Afdeling Toegepaste Wiskunde
TU Delft
h.m.schuttelaars@tudelft.nl***Huib de Swart***IMAU
Universiteit Utrecht
h.e.deswart@uu.nl*

Dynamica van bodempatronen in ondiepe zeeën

Door de wisselwerking van stroming en sedimenttransport kunnen op de bodem van ondiepe zeeën allerlei patronen ontstaan. Voorbeelden zijn de zandribbels die zichtbaar zijn op het strand bij laag tij, zandbanken voor de kust, en het stelsel van geulen en platen in de Waddenzee. Henk Schuttelaars en Huib de Swart doen onderzoek naar de dynamica van dergelijke patronen aan de hand van geïdealiseerde modellen. In dit artikel leggen ze uit hoe zulke modellen geformuleerd zijn, en hoe de analyse van de modellen inzicht biedt in het ontstaan van bodempatronen.

Wanneer je vanaf de duintoppen zoals bij Egmond aan Zee, naar de overgang kijkt tussen zee en strand, dan blijkt de kustlijn vaak meer structuur te hebben dan je vermoedt. Zo vertoont de kustlijn tussen Hoek van Holland en Den Helder golven met golflengten tussen de 2 en 5 km. Soms word je ook verrast door veel kleinschaliger ruimtelijke patronen bij deze overgang, die wel eenvoudig zijn op te merken. Een mooi voorbeeld hiervan is te zien in Figuur 1. Deze bodemvormen worden 'beach cusps' genoemd,



Figuur 1 Een voorbeeld van beach cusps, overgenomen uit [1]

de lengte van deze cusps varieert van een aantal meter tot wel 60 meter. Als je nu van het duin afdaalt en richting zee loopt, is de kans groot dat je, als het laag water is, uitgestrekte velden met zogeheten zandribbels aantreft. Zandribbels zijn dikwijls georganiseerd in langgestrekte golfpatronen met een hoogte van enkele centimeters en toppen die ongeveer 10 centimeter van elkaar afliggen (zie Figuur 2). Interessant is dat de golfpatronen ook 'defecten' vertonen, waar het patroon wezenlijk driedimensionaal is.

Andere veelvoorkomende bodempatronen zijn vaak moeilijker waarneembaar. Soms kunnen we deze patronen op een indirecte manier waarnemen. In de brandingszone bijvoorbeeld, zien we dat golven vaak op specifieke locaties breken. Het is bekend dat golven breken als de waterdiepte van dezelfde orde is als de amplitude van de golf. De intensiteit van brekende golven is dus gecorreleerd met de lokale waterdiepte, de precieze relatie tussen breking en waterdiepte wordt op dit moment onderzocht. Tegenwoordig zijn op verscheidene locaties langs de kust videosystemen geïnstalleerd die de brekende golven fotograferen. Dit geeft een

beeld van de dynamische veranderingen in de kustmorfologie, zie bijvoorbeeld de beelden op <http://argus-data.wldelft.nl/sites/egmond/2013>. Bestudering van deze beelden laat zien dat er verschillende patronen zijn waar te nemen. Soms is er alleen een kust-uniforme brekerbank te zien. Op andere beelden zijn er duidelijke ruimtelijk-ritmische patronen te zien. Veranderingen in deze patronen doen zich voor op een tijdschaal van uren tot dagen.



Figuur 2 Een voorbeeld van zandribbels

Foto: A. Falqués



Figuur 3 Satellietfoto van het Friesche Zeegat (links) en het Pinkegat (rechts)

Andere bekende bodempatronen waarmee we allemaal wel bekend zijn van bijvoorbeeld het wadlopen of een overtocht per veerboot naar Ameland, zijn de geulplaatssystemen in de Waddenzee. In deze bekens zien we vaak een fractaal netwerk van geulen, met tussen de geulen ondiepten, de platen. Een voorbeeld zien we op een satellietfoto van het Friesche Zeegat en Pinkegat, zie Figuur 3. Aan de zeewaartse kant van het Friesche Zeegat is één breed kanaal te zien dat zich vervolgens in meerdere kanalen splitst. Het Pinkegat heeft twee geulen aan de zeewaartse kant, die zich vervolgens in meerdere geulen splitsen. Uit waarnemingen blijkt dat deze geulen niet stationair zijn, maar in de tijd min of meer periodiek gedrag vertonen met een typische tijdschaal van 20 jaar (elders 7 tot 100 jaar).

In de volgende paragraaf wordt kort geschetst hoe we deze patronen kunnen bestuderen met behulp van een gekoppeld stelsel niet-lineaire partiële differentiaalvergelijkingen. Voor een voorbeeld wordt dit stelsel vergelijkingen expliciet uitgewerkt. Dit stelsel vergelijkingen wordt daarna in meer detail bestudeerd en in de laatste paragraaf bediscussieerd.

Morfodynamische Modelleren

Het is intuïtief duidelijk dat, om bovengenoemde bodempatronen (en vele andere, niet genoemde bodenvormen zoals megaribbels, zandgolven, zandbanken, et cetera) met een

mathematisch-fysisch model te bestuderen, zo'n model de wisselwerking tussen de beweging van het water, het transport van sediment en de verandering van de bodem moet beschrijven. Deze interacties kunnen met behulp van een complex-numeriek model worden bestudeerd, waarin alle mogelijke fysische processen worden meegenomen en de meest geavanceerde formuleringen voor bijvoorbeeld zandtransport worden gebruikt. Het nadeel van dit type modellen is dat de modelresultaten vaak moeilijk zijn te analyseren, dominante processen moeilijk zijn te identificeren en parametergevoeligheden slechts in beperkte mate te onderzoeken (door lange rekentijden per numeriek experiment en het grote aantal parameters).

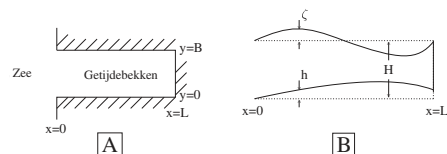
Een ander type modellen, de zogenaamde geïdealiseerde modellen, is juist ontwikkeld om de interacties en parametergevoeligheden in detail te bestuderen. Dit betekent wel dat alleen die processen in de mathematisch-fysische beschrijving worden meegenomen die essentieel zijn om specifieke bodenvormen te verklaren. Zo'n model kan dan ook niet direct worden toegepast op andere bodenvormen! Vaak kunnen we met dit type modellen ook alleen geïdealiseerde geometrieën bestuderen.

In deze bijdrage zullen we een model bespreken voor een getijdenbekken, zoals het Friesche Zeegat, dat is gepubliceerd in [4, 10]. De echte geometrie is vereenvoudigd tot een rechthoekig bekken, zie Figuur 4. In deze fi-

guur is L de lengte van het bekken ($\mathcal{O}(10)$ km), B de breedte ($\mathcal{O}(1)$ km) en H de diepte ($\mathcal{O}(10)$ m) aan de zeewaartse kant. De zeeniveauijking ten opzichte van gemiddeld niveau wordt gegeven door ζ ($\mathcal{O}(1)$ m), h geeft de bodemuitwijking ten opzichte van de diepte aan de zeewaartse kant van het bekken. Verder is x de coördinaat in de longitudinale richting, y in de laterale richting. De dominante forcering van de waterbeweging is het getij, andere mogelijke forceringen worden verwaarloosd (zoals windgolven, dichtheidsverschillen, windforcering). Aangezien we de geul-plaatpatronen in dit bekken willen bestuderen, zijn we geïnteresseerd in bodenvormen die van de orde van de bekkenlengte zijn. Deze bodenvormen zijn veel groter dan de typische waterdiepte, zodat we ons kunnen beperken tot de ondiepwatervergelijkingen, een limiet van de Navier–Stokes vergelijkingen. Als we de vergelijkingen schalen door typische orde-grootten voor de verschillende variabelen te nemen, zien we dat de niet-lineaire interacties in leidende orde verwaarloosbaar zijn. Als we ook Coriolis-effecten verwaarlozen, wordt de waterbeweging beschreven door de volgende dimensieloze vergelijkingen:

$$\begin{aligned}\zeta_x &= 0, \\ \zeta_y &= 0, \\ \zeta_t + [(1-h)u]_x + [(H-h)v]_y &= 0, \\ v_{xt} - u_{yt} &= -F_{2x}^b + F_{1y}^b.\end{aligned}\quad (1)$$

Hierin zijn t de met de getijfrequentie ($\sim 10^{-4} \text{ s}^{-1}$) geschaalde tijd, en x en y de met L geschaalde ruimtelijke coördinaten. Subscripts geven partiële afgeleiden aan. De grootheden u en v representeren geschaalde snelheden in respectievelijk de x - en y -richting, de vector F^b representeert de bodemwrijvingseffecten. Deze effecten parametriseren we als $F^b = (F_1^b, F_2^b) = r\tilde{u}(1+h)$, de bodemwrijving is dus evenredig met de dieptegemiddelde snelheid en neemt toe als de waterdiepte afneemt (h toeneemt). Als randvoorwaarden schrijven we de getij-uitwijking aan de open rand voor, in deze bijdrage beperken we ons tot één getijcomponent (het dubbeldagse getij, met een periode van 12,5



Figuur 4 Boven-aanzicht (A) en zij-aanzicht (B) van het gemodelleerde bekken

uur). Verder eisen we dat er geen flux van water door de vaste wanden is.

Het transport van sediment in het water wordt beschreven met een advectie-diffusievergelijking. Uit schalingsargumenten volgt dat transport ten gevolge van advectieve processen in een kort bekken verwaarloosbaar is ten opzichte van diffusieve sediment-fluxen. Dit resulteert in de volgende dimensioneloze concentratievergelijking:

$$aC_t = (u^2 + v^2) - C + a\mu[C_{xx} + C_{yy}], \quad (2)$$

waarbij C de dieptegeïntegreerde concentratie is, a (~ 0.01) de verhouding van de typische tijdschaal voor een deeltje om naar de bodem te zakken en de getijdetijdschaal en μ ($\sim 10^{-3}$) de dimensioneloze horizontale diffusiecoëfficiënt. Het sediment wordt opgewerveld door de schuifspanningen aan de bodem, deze *erosieterm* is evenredig met $u^2 + v^2$. Ten gevolge van de zwaartekracht zinkt het sediment weer naar de bodem, deze *depositieterm* is evenredig met $-C$. Als randvoorwaarden leggen we op dat erosie en depositie in balans zijn op de open rand, en dat er geen diffusief sedimenttransport is door de gesloten randen.

Het bodemniveau zelf verandert natuurlijk ten gevolge van erosie en depositie van sediment: waar sediment wordt geërodeerd zal het bodemniveau h afnemen, ten gevolge van depositie zal h toenemen. De typische tijdschaal waarop het geul-plaatstelsel zich ontwikkelt (maanden tot een jaar), is veel langer dan de getijperiode. Door deze observatie kunnen we de middelingsmethode toepassen: we kijken niet naar de bodemveranderingen tijdens een getijcyclus, maar naar de bodemverandering gemiddeld over een getijcyclus (zie bijvoorbeeld [8]). Als we nu getijmiddeling aangeven met $\langle \cdot \rangle$, volgt uit de behoudswet voor sediment dat de bodemevolutie wordt beschreven door de volgende vergelijking:

$$h_\tau = -\langle (u^2 + v^2) - C \rangle + \lambda \langle \nabla^2 h \rangle, \quad (3)$$

waarbij $\tau = \delta t$, met δ de verhouding van de hydrodynamische en morfodynamische tijdschaal. Merk op dat de bodemdiepte h toeneemt ten gevolge van depositie en afneemt ten gevolge van erosie. Verder modelleert de laatste term aan de rechterkant van (3) het transport ten gevolge van bodemhellingstermen: sediment rolt makkelijker van een hel-

ling af dan er tegenop. De randvoorwaarden voor dit systeem worden gegeven door de eis dat de bodemdiepte aan de zeewaartse kant niet verandert, en dat er geen sedimenttransport is ten gevolge van de bodemhellingstermen door de gesloten wanden.

Oplosmethode

De vergelijkingen en randvoorwaarden uit de vorige sectie vormen, gegeven een initiële conditie, een zogenaamd morfodynamisch model. Dit model wordt vaak met behulp van tijdsintegratie opgelost: gegeven een bodemniveau kunnen we de bijbehorende snelheden uitrekenen. Deze snelheden resulteren in erosie van sediment, dat door diffusieve processen wordt getransporteerd en vervolgens door de zwaartekracht weer naar de bodem zinkt. Hierdoor verandert de bodem, waarna we weer vooraan in de loop beginnen. Dit gaat net zolang door totdat de bodem getijgemiddeld niet meer verandert.

Breedtegemiddelde evenwichtsoplossing

In plaats van zo'n morfodynamische integraal te doorlopen, kunnen we ook direct op zoek gaan naar evenwichtsoplossingen. Als eerste stap proberen we een oplossing van vergelijkingen (1)–(3) te vinden, zodanig dat de waterbeweging, sedimentconcentraties en bodem niet meer veranderen. Zo'n oplossing is redelijk eenvoudig te vinden als we het systeem middelen over de breedte. Dit betekent dat de laterale snelheid v nul is en de andere grootheden niet van de laterale coördinaat y afhangen. Als we het transport ten gevolge van bodemhellingstermen verwaarlozen (door $\lambda = 0$ te kiezen), dan reduceert het stelsel vergelijkingen tot

$$\begin{aligned} \zeta_t + [(1-h)u]_x &= 0, \\ \zeta_x &= 0, \\ aC_t - a\mu C_{xx} &= u^2 - C, \\ h_\tau &= -\langle u^2 - C \rangle, \end{aligned} \quad (4)$$

met als randvoorwaarden dat $\zeta = \cos t$, $h = 0$ op de zeewaartse rand ($x = 0$) en $(1-h)u = 0$, $\mu \langle C_x \rangle = 0$ op de gesloten rand ($x = 1$). Verder willen we geen diffusieve grenslagen voor het tijdsvariërende deel van de sedimentconcentratie C op $x = 0$ en $x = 1$. Uit de tweede vergelijking in (4) volgt nu dat $\zeta = \cos t$ voor alle $x \in [0, 1]$, ofwel een ruimtelijk uniform getij. Het horizontale snelheidsveld wordt dan gegeven door $u(x, t) = \frac{(x-1)}{1-h} \sin t$. Door op te merken dat $a \ll 1$ kan de bijbehorende sedimentconcentratie C worden opgelost,

wat resulteert in de volgende bodemevolutievergelijking:

$$h_\tau = \frac{a\mu}{2} \left(\frac{x-1}{1-h} \right)_{xx}^2. \quad (5)$$

Voor een evenwichtsoplossing h_{eq} eisen we dat $h_\tau = 0$. Hieruit volgt dat $h_{eq} = x$, ofwel dat de evenwichtbodem een constant hellende bodem is. Dit resultaat blijkt behoorlijk goed te kloppen met waarnemingen van breedte-gemiddelde bodemprofielen in de Waddenzee, zie bijvoorbeeld [2], waarin ook een uitgebreide versie van dit model wordt geanalyseerd. In [7] wordt de dynamische ontwikkeling van zo'n bekken onder de invloed van relatieve zeespiegelstijging bestudeerd, volgens dit model kan er voldoende sediment geïmporteerd worden om de huidige zeespiegelstijging bij te houden. Een snellere zeespiegelstijging resulteert in het verdwijnen van de ondiepe delen, zie ook [6].

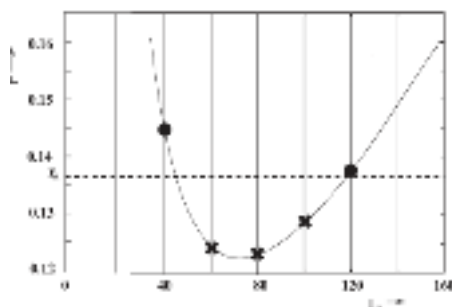
Lineaire stabiliteitsanalyse

Deze evenwichtsoplossing blijkt stabiel te zijn voor kleine verstoringen. Dit volgt uit een *lineaire stabiliteitsanalyse*, waarbij de evenwichtsbodem een klein beetje verstoord wordt:

$$h(x, \tau) = h_{eq} + \Re \{ h'(x) e^{i\omega\tau} \}. \quad (6)$$

Substitueer deze uitdrukking in (5) en lineariiseer de gevonden uitdrukking, dat wil zeggen dat alleen termen lineair in h' niet worden verwaarloosd. Het resulterende eigenwaardeprobleem kan expliciet opgelost worden, alle eigenwaarden ω zijn reëel en negatief, zie [9] voor details.

Om het ontstaan van geul-plaatpatronen te kunnen modelleren, kunnen we ons natuurlijk niet beperken tot het breedte-gemiddelde systeem van vergelijkingen. Bovenstaande analyse is niettemin essentieel omdat de breedte-gemiddelde evenwichtsoplossing ook een evenwichtsoplossing is van het tweedimensionale systeem van vergelijkingen (1) (merk op dat er een zeer kleine correctie in het bodemprofiel is ten gevolge van de bodemhellingsterm in vergelijking (3), zie [2] voor details). De constant hellende bodem blijkt *niet* meer lineair stabiel te zijn als we verstoringen met een structuur in de laterale richting toelaten. Om dit te laten zien verstoren we de breedte-gemiddelde evenwichtsbodem met kleine verstoringen die een laterale structuur hebben. Deze laterale structuur wordt volledig bepaald door de randvoor-



Figuur 5 Neutrale stabiliteitscurves voor de eerste (getrokken lijn) en tweede (gestippelde lijn) longitudinale eigenmodus. Voor $r < r_c$ zijn alle bodemverstoringen stabiel ($\omega < 0$). Voor $r = r_c$ zien we dat drie eigenmodi lineair instabiel zijn (aangegeven met een kruis), de andere eigenmodi (waarvan er een aantal zijn aangegeven met een gevulde cirkel), zijn stabiel. De parameter $l_n = \pi L/B = 40$.

waarden, de toegestane bodemverstoringen zijn te schrijven als

$$h'(x, y, \tau) = \Re \{ h_n(x) \cos(l_n y) e^{\omega_n \tau} \},$$

met het lateraal golfgetal $l_n = n\pi L/B$ voor $n = 0, 1, 2, \dots$; voor de structuur van de snelheids- en concentratieverstoringen, zie [10]. Substitueer de perturbaties in de niet-lineaire vergelijkingen, en lineariseer deze vergelijkingen. Het resulterende gegeneraliseerde eigenwaardeprobleem moet numeriek opgelost worden, en resulteert in de eigenwaarden ω_n (die hier altijd reëel blijken te zijn) en eigenfuncties $h_n(x)$ (er zijn bij elk golfgetal n natuurlijk oneindig veel longitudinale eigenmodi).

Figuur 5 laat op een compacte manier de stabiliteit van de verschillende verstoringen zien: kies een verstoring met een toegestaan lateraal golfgetal l_n (de variabele op de horizontale as), deze zijn aangegeven door de verticale lijnen in de figuur. Voor kleine waar-

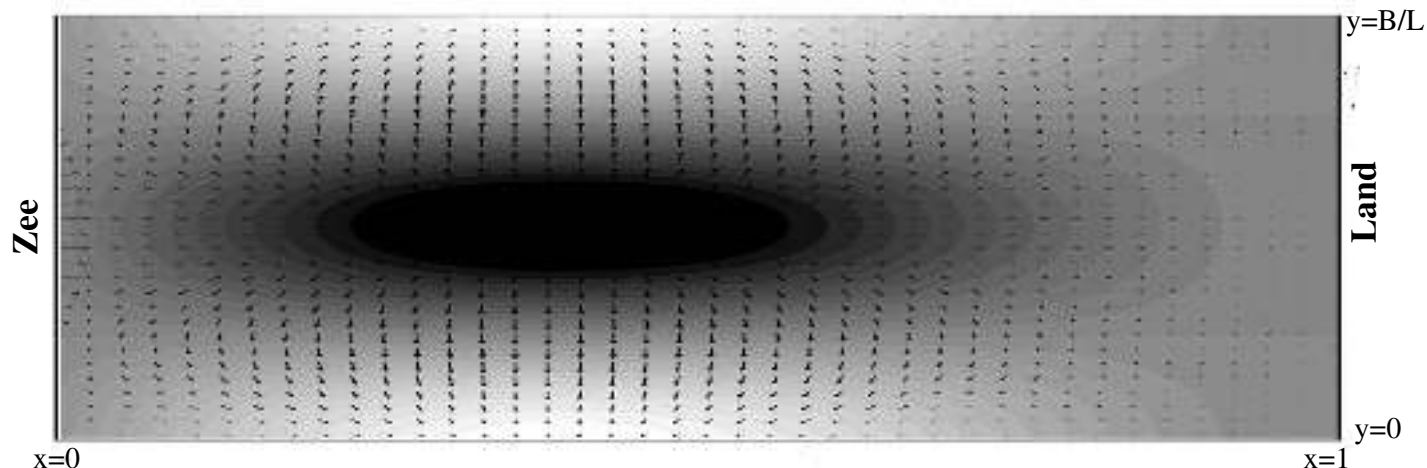
den van de wrijvingsparameter r (de variabele op de verticale as) neemt de amplitude van de verstoring af ($\omega < 0$). Als we r laten toenemen, verandert de groeisnelheid ω van teken, dit betekent dat verstoring niet meer uitdempt maar exponentieel gaat groeien. Wanneer dat het geval is, geldt dat de breedte-gemiddelde evenwichtsbodem lineair instabiel is. De getrokken lijn in Figuur 5, de *neutrale curve*, geeft de overgang aan tussen het parametergebied waarin oplossingen stabiel en instabiel zijn. De kleinste wrijvingswaarde waarvoor de constant hellende bodem instabiel wordt, wordt de *kritische* wrijvingswaarde genoemd, het bijbehorende golfgetal het kritische golfgetal. De ruimtelijke structuur van de bijbehorende bodemverstoring, getoond in Figuur 6, blijkt een duidelijk geulplaatpatroon te zijn.

Niet-lineaire evenwichtsooplossingen

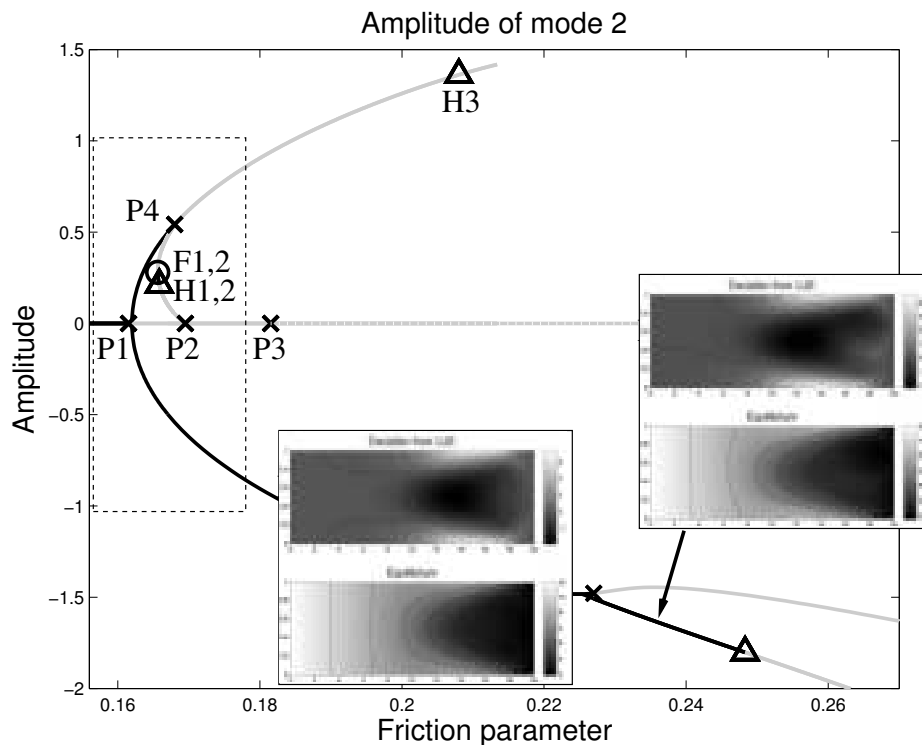
Uit de vorige paragraaf volgt dat, als de wrijvingsparameter groter is dan een kritische waarde r_c , de amplituden van sommige verstoringen gaan groeien. Niet-lineaire interacties zijn dan niet meer verwaarloosbaar, en het oorspronkelijke niet-lineaire stelsel van vergelijkingen moet worden geanalyseerd. We verwachten dat er voor deze waarden van de wrijvingsparameter niet-triviale evenwichten zullen bestaan. Dit zijn evenwichten met een structuur in de laterale richting. Deze evenwichten kunnen worden gevonden door een bekend evenwicht te *continueren* door de waarde(n) van modelparameter(s) langzaam te veranderen (zie [4] voor een gedetailleerde beschrijving). In dit geval zullen we de sterkte van de bodemwrijving r als continueringsparameter gebruiken. We starten met een wrijvingsparameter $r < r_c$, het evenwicht is dan bekend (de constant hel-

lende bodem). We gaan nu langzaam r groter maken, tot het bekende evenwicht instabiel wordt. Bij die waarde van r waarbij het reële deel van ten minste één van de eigenwaarden door nul gaat, treedt een *bifurcatie* op: het aantal evenwichten kan toe- of afnemen, de stabiliteit van de evenwichten kan veranderen, en (quasi-)periodieke oplossingen kunnen ontstaan of verdwijnen. Eventuele nieuwe evenwichten en/of periodieke oplossingen kunnen vervolgens ook weer worden gecontinueerd. Op deze manier vinden we een *bifurcatiediagram*, een voorbeeld is gegeven in Figuur 7. Zwarte lijnen in dit diagram geven aan dat evenwichten stabiel zijn, grijze lijnen duiden instabiele evenwichten aan. In deze figuur zien we dat voor kleine r de amplitude van de verstoring nul is, en dat het constant hellende bodemprofiel stabiel is. Als $r \sim 0,163$ verliest dit breedte-gemiddelde profiel zijn stabiliteit en zien we dat er twee nieuwe stabiele evenwichten ontstaan. Dit type bifurcatie wordt een *pitchfork bifurcatie* genoemd, hier aangegeven met *P1*. Als we deze nieuwe evenwichten volgen door r te variëren, treden er nieuwe bifurcaties op (zoals andere pitchfork bifurcaties, aangegeven met een *P*, folds (*F*) en Hopf-bifurcaties (*H*)). Bij Hopf-bifurcaties ontstaan of verdwijnen periodieke oplossingen. Het bifurcatiediagram laat zien dat er bij dezelfde parameterwaarden meerdere (stabiele) evenwichten mogelijk zijn. Verder kunnen we ook periodieke oplossingen verwachten, welke in dit model te interpreteren zijn als migrerende geulen en platen.

Het bovenste figuur in de linker inzet in Figuur 7 laat de niet-triviale evenwichtsbodem zien voor $r \sim 0,2$, de onderste figuur in deze inzet het verschil ten opzichte van de constant hellende bodem: door de interne dynamica



Figuur 6 Meest instabiele bodemverstoring ($n=2$), behorend bij de neutrale stabiliteitscurve in Figuur 5. Donkere (lichte) tinten geven negatieve (positieve) bodemverstoringen weer, pijltjes laten de richting van het getijgemiddeld sedimenttransport zien.



Figuur 7 Bifurcatie diagram voor een bekken met een lengte van 20 km en een breedte van 1 km

organiseert het systeem zich in een kanaal in het midden van het bekken met aan weerszijde een plaat. Voor $r = 0,23$, rechter inzet, zien we dat het kanaal zich nogmaals splitst: voor $x < 14$ km is er één kanaal, verder ach-

terin het bekken zijn er twee kanalen te zien, gescheiden door een plaat.

Discussie en conclusies

Een systematische analyse van een sys-

teem gekoppelde niet-lineaire differentiaalvergelijkingen resulteert in ruimtelijke patronen die kwalitatief het waargenomen geul-plaatsysteem in bijvoorbeeld Waddengebieden reproduceren. Deze ruimtelijke patronen zijn een voorbeeld van morfodynamische zelforganisatie [5], dat wil zeggen dat de positieve terugkoppelingen binnen het model zelf resulteren in de waargenomen patronen. Ook veel van de in de inleiding genoemde observaties kunnen worden beschreven met behulp van morfodynamische terugkoppelingen. Het is dan wel zaak om de voor die observaties relevante processen in het mathematische model correct mee te nemen.

Er zijn nog vele uitdagingen om deze terugkoppelingen correct te representeren binnen een geïdealiseerd mathematisch model, denk hierbij bijvoorbeeld aan de invloed van klimaatverandering op de geul-plaatpatronen, de interactie tussen verschillende getijdcomponenten, extreme gebeurtenissen en de interacties tussen sediment met verschillende korrelgroottes. Het is van belang om deze processen op een systematische manier te kunnen analyseren, inzicht in het bestaan van meervoudige evenwichten en hun domein van attractie is essentieel in het nemen van optimale beslissingen met betrekking tot bijvoorbeeld gaswinning in de Waddenzee.

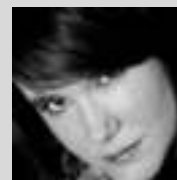
Referenties

- 1 R. Almar, G. Coco, K.R. Bryan, D.A. Huntley, A.D. Short en N. Senechal, Video observations of beach cusp morphodynamics, *Marine Geology* 254 (2008), 216–223.
- 2 M.C. ter Brake en H.M. Schuttelaars, Modeling equilibrium bed profiles of short tidal. On the effect of the vertical distribution of suspended sediment and the influence of the boundary conditions, *Ocean Dynamics* 60 (2009), 183–204.
- 3 M.C. ter Brake en H.M. Schuttelaars, Channel and shoal development in a short tidal embayment; an idealized model study, *J. Fluid Mech.* 677 (2011), 503–529.
- 4 M.C. ter Brake, Tidal embayments: modelling and understanding their morphodynamics, Proefschrift, TU Delft, 2011.
- 5 G. Coco en A.B. Murray, Patterns in the sand: From forcing templates to self-organization, *Geomorphology* 91 (2007), 271–290.
- 6 D.M.P.K. Dissanayake, R. Ranasinghe en J. A. Roelvink, The morphological response of large tidal inlet/basin systems to relative sea level rise, *Climatic Change* 113 (2012), 253–276.
- 7 S.M. van Leeuwen, Tidal inlet systems: bottom pattern formation and outer delta development. Proefschrift, Universiteit Utrecht, 2012.
- 8 J.A. Sanders en F. Verhulst, (1985). Averaging methods in nonlinear dynamical systems. Springer, New York.
- 9 H.M. Schuttelaars en H.E. de Swart, An idealized long-term morphodynamic model of a tidal embayment., *Eur. J. Mech., B/Fluids* 15 (1996), 55–80.
- 10 H.M. Schuttelaars en H.E. de Swart, Initial formation of channels and shoals in a short tidal embayment, *J. Fluid Mech.* 386 (1999) 15–42.

Ionica Smeets

wetenschapsjournalist, Leiden

i@ionica.nl



Het keerpunt van Jan Barkmeijer

Als je buiten om je heen kijkt, daar gebeurt het

Wie past er beter in een themanummer over *Mathematics of Planet Earth* dan een wiskundige van het Koninklijk Nederlands Meteorologisch Instituut (KNMI)? Jan Barkmeijer (1959) belandde als scholier per ongeluk op de open dag bij wiskunde. Dezelfde man die hem die dag inspireerde om wiskunde te studeren, vroeg hem jaren later of het KNMI niets voor hem was. Inmiddels werkt hij daar al ruim twintig jaar aan modellen en voorspelbaarheid.

Hoe beland je per ongeluk op een open dag van wiskunde?

“Als scholier was ik erg gecharmeerd van elementaire deeltjes. Dus ging ik met mijn vader naar de open dag van natuurkunde in Groningen. Daar vertelde een professor dat alle studenten in hun eerste jaar verplicht moesten meedoen aan het practicum. Ik had helemaal geen zin om daar tijd aan te besteden, ik had slechte ervaringen met practica. In de pauze liep ik weg en belandde bij de Faculteit Wiskunde waar ook een open dag was. Daar gaf iemand een praatje over zuivere wiskunde, hij was zo enthousiast over zijn vak. Ik ben gewoon blijven hangen en werd verliefd op de wiskunde. Die man was Floris Takens en ik ben later bij hem afgestudeerd en gepromoveerd.”

Was je tijdens je studie al bezig met meteorologie?

“Nee, maar ik was wel altijd geïnteresseerd in toepassingen. In die tijd werd wiskunde voor het eerst gebruikt in de neurologie om epileptische aanvallen te voorspellen. Toen snuffelde ik aan vakken bij psychologie, met het vage idee dat ik daar iets mee kon doen. En bij mijn afstuderen verzoon ik ook een soort showelement. Ik werkte aan een vermoeden over een globale aantrekker in een dynamisch systeem. Dat systeem sterft uit. Maar als je twee van die systemen koppelt, dan komen ze onder bepaalde voorwaarden tot leven. Stephen Smale had dat bewezen voor $\mathbb{R}^4$ en ik kreeg het voor elkaar in $\mathbb{R}^2$. Een van mijn huisgenoten studeerde chemi-

sche techniek en we bouwden samen een mooie demonstratie van dit soort systemen met een chemische oscillator. Na mijn afstudeerpraatje bij wiskunde trokken we een witte jas aan en staken we over naar het chemisch lab om de proefopstelling te laten zien.”

Toch nog een soort practicum dus. Wat wilde je na je afstuderen gaan doen?

“Ik overwoog om een eigen bedrijf te beginnen met die scheikundige huisgenoot. Maar toen ik een promotieplek bij wiskunde kreeg aangeboden, verwaterde dat idee. Op dat moment gonsde er iets rond in dat gebied van de wiskunde en er kwam een eerste geldstroom beschikbaar voor

dat onderwerp. Het ging over eendimensionale afbeeldingen en de vraag hoe je bepaalde resultaten voor intervalafbeeldingen kunt voortzetten naar cirkelafbeeldingen. Het was een nieuw, opwindend onderwerp, ik was nieuwsgierig en wilde daar mijn tanden wel inzetten. Zo’n afgesloten wereld waarover je kunt nadenken zonder dat je gestoord wordt door mensen van buiten, dat is mooi.”

Wilde je na je promotie weg van de universiteit?

“Nee, maar het was lastig om een vaste aanstelling in Nederland te krijgen. Er waren wel aanbiedingen uit Brazilië en Amerika. Ik twijfelde wat ik zou doen. Toen belde Floris Takens me. Bij een bijeenkomst op de KNAW was hij aan de praat geraakt met Henk Tennekes, destijds hoofd onderzoek op het KNMI. Tennekes was net bezig een nieuwe groep te starten die zou werken aan voorspelbaarheid. Het onderzoek had raakvlak-



Jan Barkmeijer

ken met dynamische systemen en het bood kans op een vaste baan. Was dat misschien iets voor mij? Dat was het inderdaad. Dit was een nieuw vakgebied, waar de theorie nog moest worden opgebouwd. Zo kon ik een link houden met waar ik elf jaar in getraind was bij wiskunde.”

Waar denk je inmiddels aan bij ‘Mathematics of Planet Earth’?

“Als je buiten om je heen kijkt, daar gebeurt het. Als je dat kunt beschrijven met een goed model en slim gebruik van de waarnemingen, dan ben je al een heel eind. Maar de atmosfeer is chaotisch en dat maakt het moeilijk.”

Aan wat voor problemen werk je nu?

“Aardig is een promotieonderzoek waar we de afgelopen vier jaar mee bezig waren. Bij weersverwachtingen is het heel belangrijk om een zo goed mogelijke schatting te hebben van de atmosfeer. Daarvoor maken we een model dat de fysica van de atmosfeer beschrijft: daarin zitten Navier–Stokes-vergelijkingen, maar ook fysica die processen als convectie en het ontstaan van wolken beschrijft en nog veel meer. Je wilt naast dat model natuurlijk ook waarnemingen gebruiken van satellieten, radar en weerballonnen. Soms is een waarneming in tegenpraak met het model, hoe moeten we dan de waarneming met het model combineren? Nemen we alleen de meetgegevens, alleen het model, of een mengsel van die twee? Dit heet data-assimilatie. De laatste jaren wordt die data-assimilatie steeds moeilijker doordat de resolutie van de weermodellen toeneemt. We werken nu op een schaal van een kilometer. Op die schaal spelen andere processen, en die kunnen zorgen voor foutengroei; de chaos neemt toe. Roel Stappers promoveerde in juni op een nieuwe manier om model en data te combineren.”

Zijn er grote verschillen tussen de weersvoorspelling van verschillende landen? Is Nederland meer bezig met regen dan bijvoorbeeld Spanje?

“Spanje kan ook hevige onweersbuien krijgen. Neerslag is voor ieder land belangrijk, het kan leiden tot enorme schade of zelfs doden. Denk maar aan de recente overstromingen in Centraal-Europa. Elk land heeft daar pijnlijke ervaringen mee, dus dat wil je allemaal goed modelleren. We werken ook veel samen. Weermodellen voor de korte termijn, tot een dag of twee, zijn moeilijk te ontwikkelen. Daarom werken we in een consorti-

um van Europese landen. Sommige landen zoals Duitsland en Engeland gebruiken hun eigen model. Er bestaan wel regionale prioriteiten. Op onze afdeling werken we intensief aan het begrijpen hoe mist ontstaat boven de Noordzee. Soms zegt het model dat daar mist hangt, terwijl dat niet zo is. Voor Schiphol is het bijvoorbeeld heel belangrijk om dat goed te weten. In principe speelt die mist vanaf de Noordzee alleen voor Denemarken en Nederland, maar in het consortium zijn er toch meer landen geïnteresseerd in een oplossing. Misschien ontdekken we wel een foutje in het model, of komt er een nieuwe techniek uit om meetresultaten op een slimme manier in het model te brengen.”

Is die tien dagen ook echt het maximum van wat je kunt voorspellen?

“Dat hangt af van de resolutie en het weerfenomeen waarin je geïnteresseerd bent. Als je kijkt naar het ontstaan van regenbuien op de kilometerschaal, dan is de termijn eerder in uren. Je kunt een weersverwachting daarna niet deterministisch meer interpreteren. Het is noodzakelijkerwijs een kansverwachting.”

Wat betekent het nu precies als de weerman zegt dat er morgen twintig procent kans op regen is?

“Even denken, dit moet ik weten. (stilte...) Als je een willekeurige plek in Nederland kiest, dan heb je twintig procent kans dat het daar die dag regent. Dat is het. Ik heb in Engeland gewerkt aan Ensemble, het programma om de weersverwachting in een reeks van simulaties door te rekenen. Op die manier krijg je toch een idee wat Moeder Aarde ons zal geven. Je kunt door de rekentijd maar een beperkt aantal simulaties doen.”

Jullie doen dus niet één voorspelling, maar een reeks simulaties?

“Precies, twee keer per dag draait in Engeland dat Ensemble, waarbij je niet één keer de weersverwachting doorrekent, maar vijftig keer met elke keer een kleine verstoring in de beginconditie en iets andere fysica in het model. We hebben hiervoor eerst de gevoelige plekken op aarde bepaald, waar een kleine verandering echt een andere weersverwachting oplevert. Wiskundig gezien is dat allemaal niet zo interessant, het zijn de eigenvectoren van een matrix. Die eigenvectoren kun je dan interpreteren als een verstoring van je beginconditie. Daarnaast zit er ook onzekerheid in het model zelf, we hebben niet de pretentie dat we precies kunnen

beschrijven hoe wolken ontstaan. Op basis van vijftig simulaties over tien dagen kun je een kansverwachting maken. Als je in de onzekerheden de extremen meeneemt en de voorspellingen blijven desondanks dicht bij elkaar, dan kun je zeggen dat de voorspelling goed uit zal komen. Op het 8-uur-journaal zie je gelukkig zo nu en dan de pluim die de kansverdeling laat zien. Dan zie je bijvoorbeeld op de x -as de dagen, op de y -as de temperatuur en dan de resultaten van die vijftig simulaties. Hoe breed die simulaties uitwaaien, zegt iets over de onzekerheid van de weersverwachting. Op dezelfde manier kun je ook klimaatruns visualiseren. Je wilt ook een kansverdeling voor hoe het klimaat over honderd jaar is.”

Hoe sta je als meteoroloog tegenover het klimaatdebat?

“Het gaat langzaam de goede kant op. Inmiddels is toch wel het besef doorgedrongen dat de veranderingen in het klimaat mens-geïnduceerd zijn. Dat is nu toch wel duidelijk. Een koude zomer is slechts een rimpeling in de vijver.”

Kun je jullie modellen ook voor andere vakgebieden gebruiken?

“De modellen zijn zeer specifiek, maar onze technieken niet. Ik heb wel eens contact gehad met economen. Je zou een kansverwachting kunnen maken bij de modellen van het Centraal Planbureau. Je zou onze technieken om onzekerheid zichtbaar te maken daar kunnen gebruiken. Ik zie het al voor mij: een kansverwachting voor het Nederlandse begrotingstekort.”

Maar is dat dan ook een chaotisch systeem?

“Daar ben ik wel van overtuigd. De wiskunde leert dat je niet zoveel moeite hoeft te doen om een chaotisch systeem te krijgen. In die zin is het wel vreemd dat daar niet meer rekening mee wordt gehouden. Als je al kijkt hoe lang het duurt voor het werk van Lorenz zich echt verspreidde. Het is moeilijk voor mensen om om te gaan met onzekerheid. Ensemble bestaat nu twintig jaar, maar nog steeds is het moeilijk om zo’n kansverwachting te laten gebruiken door bedrijven. Mensen hebben behoefte aan een deterministische uitspraak: ze willen weten wat ze moeten doen.”



Heeft u een goede suggestie voor een (Nederlandse) wiskundige met een bijzonder keerpunt in zijn of haar carrière? Stuur dan een e-mail naar i@ionica.nl.

Brenny van Groesen

Toegepaste Wiskunde, Universiteit Twente, Enschede
LabMath-Indonesia, Bandung, Indonesië
e.w.c.vangroesen@utwente.nl

Frits van Beckum

Ede
frits@vanbeckum.nl

Watergolven

Brenny van Groesen en Frits van Beckum schetsen in deze bijdrage een variationele formulering voor de beweging van het wateroppervlak. De variationele formulering behoudt de Hamiltoniaanse structuur van de dynamische vergelijkingen voor het oppervlak, en is daarom heel geschikt voor het modelleren en simuleren van oppervlaktegolven. Ter illustratie behandelen de auteurs voorbeelden uit hun eigen onderzoek aan de Universiteit Twente dat sinds de negentiger jaren in nauwe samenwerking met Indonesische instituten en studenten is uitgevoerd.

De vijfde plaats van Nederland in de lijst van kenniseconomielanden is voor een deel te danken aan de maritieme industrie waarin we internationaal koploper zijn. Dit wordt mede mogelijk gemaakt door de goede infrastructuur, waardoor wetenschappelijke vragen en resultaten direct worden uitgewisseld met bedrijven/industrie. Het maritiem instituut MARIN in Wageningen vervult hierin als een spin in het web een belangrijke rol. De relevantie van maritieme kennis is begrijpelijk omdat circa 45 procent van de wereldbevolking op minder dan honderd kilometer van de kust woont; acht van de tien grootste steden met hun grote economische activiteit liggen aan de kust. Dit vereist een goede infrastructuur, zoals dijken, havens en offshore platforms, voor bescherming tegen het water, voor transport van grondstoffen en producten, en voor het delven van olie en gas en het winnen van windenergie.

Naast de praktische relevantie is ook de menselijke fascinatie voor waterbewegingen en golven in het verleden en heden een belangrijke drijfveer voor onderzoek van de meerwetenschappelijke kanten van watergolven. Wie is niet verbaasd over de zich nooit herhalende golven aan het strand, of de merk-

waardige golfpatronen in een rivier of van regenwater op een hellende straat. Ook in wetenschappelijk opzicht is Nederland zichtbaar: bekende vergelijkingen dragen de namen van Korteweg en de Vries en van Burgers; de ideeën van Huygens zijn net zo belangrijk in de theorie van watergolven als in die van licht. Heden ten dage worden numerieke codes die zijn ontwikkeld door 'Delft' dagelijks door duizenden over de hele wereld gebruikt. Software zoals 'SWAN' (Booij, Van Vledder e.a.) en 'Delft3D' (Guus Stelling e.a.) maakt dat Delft met de TU Delft en Deltares een tweede zwaartepunt is in het nationaal maritiem onderzoek.

Hieronder geven we in vogelvlucht een aantal ontwikkelingen in het begrijpen van vrije oppervlaktegolven. We beginnen met een korte samenvatting van de basisvergelijkingen die de beweging van het wateroppervlak beschrijven. Via Newton, Euler, Boussinesq en Zakharov arriveren we bij een variationale formulering die vooral voor modelleren en simuleren zeer bruikbaar blijkt. Daarna illustreren we verschillende verschijnselen en geven enkele toepassingen.

De behandelde voorbeelden zijn een keuze uit ons onderzoek van de laatste jaren op

de Universiteit Twente in samenwerking met het onderzoeksinstituut LabMath-Indonesia in Bandung.

Beschrijving van oppervlaktegolven

Klassieke mechanica

De klassieke mechanica voor wrijvingsloze puntmassa's begint bij Newton. We nemen een aantal puntmassa's, met gelijke constante massa m , die kunnen bewegen in een vlak, waar de posities op tijd t worden aangegeven met $q_k(t)$, samengevat in de vector $q(t)$. De bijbehorende impulsvector is gerelateerd aan de snelheid: $p = mq$. Als de deeltjes een kracht op elkaar uitoefenen beschreven met een krachtveld F wordt de gehele dynamica samengevat als een stelsel van twee differentiaalvergelijkingen van eerste orde in de tijd:

$$\begin{cases} \dot{q} = p/m, \\ \dot{p} = F. \end{cases} \quad (1)$$

In geval F een conservatief krachtveld is, met potentiaal V , zodat $F(q) = -\nabla V(q)$, is het mogelijk de energie van het systeem te definiëren als de som van kinetische en potentiële energie, in dit geval gegeven door

$$H(q, p) = \frac{1}{2m} |p|^2 + V(q). \quad (2)$$

Het is eenvoudig te verifiëren dat inderdaad $dH/dt = 0$ voor oplossingen van de dynami-

sche vergelijkingen. Maar, meer nog, de dynamische vergelijkingen kunnen nu in een canonicke vorm geschreven worden als

$$\begin{cases} \dot{q} = \partial H / \partial p, \\ \dot{p} = -\partial H / \partial q. \end{cases} \quad (3)$$

Dit is de zogenaamde Hamiltoniaanse vorm, en H wordt de Hamiltoniaan genoemd, met q en p als canonicke geconjugeerde variabelen. Deze beschrijving geeft ook een directe relatie met het principe van stationaire actie in de klassieke mechanica, omdat de vergelijkingen precies de nodige voorwaarden zijn waaraan stationaire punten van de actiefunctionaal moeten voldoen, welke functionaal gegeven wordt door

$$\int [p\dot{q} - H(q, p)] dt. \quad (4)$$

De gebruikelijke Newtonse beschrijving wordt gevonden, na eliminatie van de impulsen, als een stelsel niet-lineair gekoppelde oscillatoren: $\dot{q} = -\partial V(q)/\partial q$.

Hamilton-consistent modelleren

De Hamiltoniaanse formulering heeft niet alleen een intrinsieke schoonheid, maar blijkt ook voor modelleren een grote waarde te hebben. De structuur garandeert onder andere energiebehoud: $\partial H(q, p)/\partial t = 0$. In het algemeen zal, wanneer we het stelsel via discretisatie, truncatie of anderszins door een benadering vervangen, een conservatieve kracht $F(q, p)$ niet vanzelf conservatief blijven, dus energiebehoud gaat verloren. Passen we echter de benadering toe op de potentiaal, of meer algemeen op de Hamiltoniaan, en leiden we dan het nieuwe stelsel af via de Hamiltoniaanse formulering, dan is er ook in het nieuwe stelsel behoud van energie. Deze laatste aanpak wordt Hamilton-consistent modelleren genoemd en is de basis waarop wij benaderingen, inclusief numerieke discretisaties, van de golfdynamica uitvoeren.

Continuïumdynamica

Voor de beschrijving van water heeft Euler de generalisatie gegeven van de wet van Newton. In die continuümtheorie zijn er oneindig veel 'deeltjes'. De snelheid, bijvoorbeeld, wordt dan gegeven door $\mathbf{u}(\mathbf{x}, t)$ als de snelheid van het deeltje dat ten tijde t op de plaats $\mathbf{x} = (x, y, z)$ is. Onsamendrukbaarheid van water (bij constante temperatuur), en de praktisch gesproken afwezigheid van viscositeit leiden tot aanzienlijke vereenvoudigin-

gen, in het bijzonder tot energiebehoud. Voor de beweging van een laag water moet ook de beschrijving van het bewegende oppervlak meegenomen worden. Die is gekoppeld aan de interne beweging als consequentie van de continuümaanname dat het vrije oppervlak, dat we beschrijven als $z = \eta(x, y, t)$, steeds uit dezelfde deeltjes bestaat (we sluiten brekende golven uit in het vervolg). Dan volgt daaruit de vergelijking $\partial_t \eta = N$, waarin ∂_t de partiële tijdafgeleide is, en N de normaalcomponent van de snelheid op het oppervlak. Deze kinematische relatie is de continuïteitsvergelijking.

In het onderzoek naar 'golven' in de daarop volgende eeuwen werd, tot op de dag van vandaag, meestal de extra aanname gemaakt dat de stroming rotatievrij is. Dat vereenvoudigt de situatie verder, omdat dan de snelheid gerelateerd is aan een vloeistofpotentiaal Φ volgens $\mathbf{u} = \nabla \Phi$. De inwendige vergelijkingen reduceren dan ten gevolge van de incompressibiliteit ($\nabla \cdot \mathbf{u} = 0$) tot het Laplace-probleem $\Delta \Phi = 0$. Op het oppervlak komt een tweede, dynamische vergelijking, die krachten op het oppervlak, zoals bijvoorbeeld van de wind of een schip kan bevatten. Wij zullen alleen 'vrije' oppervlakken beschouwen zoals het geval is met een uniforme atmosfeer boven het water.

Maar zelfs met al deze aannamen zijn de vergelijkingen niet eenvoudig, omdat voor een niet-vlak wateroppervlak of bodemprofiel het Laplace-probleem niet expliciet opgelost kan worden, en de twee vergelijkingen op het oppervlak niet-lineair zijn. Vele vereenvoudigde modellen werden in de negentiende eeuw opgesteld, steeds door een benadering te zoeken voor de oplossing van het Laplace-probleem onder verschillende aannamen voor karakteristieke grootheden van de golven. Vergelijkingen die in alleen oppervlaktegrootheden geschreven zijn, worden Boussinesq-achtige vergelijkingen genoemd.

Een speciaal geval daarvan is de Korteweg-de Vries-vergelijking (1895) voor eenrichtingsgolven. Bijzonder was dat die vergelijking aantoonde dat er een golfprofiel bestaat, later soliton genoemd, dat zich onvervormd en met constante snelheid voortplant. Daarmee werd een verklaring gegeven voor de observatie van Scott Russell van een golf met die eigenschap (1834). Dezelfde solitonoplossing, maar dan in de Maxwell-vergelijkingen, heeft ons later dankzij Hasegawa (1973) de mogelijkheid gegeven om signalen door glaskabels te transporteren, en heeft daarmee de moderne communicatie mogelijk gemaakt.

Variationele structuur

Het duurde tot 1968 voordat Zakharov [16], en later Broer [2], expliciet maakten dat ook oppervlaktegolven een variationele structuur hebben, en als een continu Hamiltons systeem geschreven kunnen worden. Miles [13] toonde iets later aan dat een herschrijven van een 'druk-principe' geformuleerd door Luke in 1967 precies tot een actieprincipe leidt zoals in de klassieke mechanica [12]. De canonieke variabelen zijn in dit geval de waterhoogte η , zoals te verwachten is, en de potentiaal $\varphi(x, y, t) = \Phi(x, y, z, t)$ voor $z = \eta(x, y, t)$ op het oppervlak (niet zomaar te verwachten). Overeenkomstig de klassieke systemen is de Hamiltoniaan H weer de totale energie: de som van potentiële energie P en kinetische energie K . Dan luiden de vergelijkingen:

$$\begin{cases} \partial_t \eta = \delta_\varphi H, \\ \partial_t \varphi = -\delta_\eta H. \end{cases} \quad (5)$$

Hierin is $\delta_\varphi H$ de partiële variatieafgeleide met betrekking tot φ , dat is (voor een dichtheidsfunctionaal H) die functie die via het L_2 -inproduct de linearisatie van de functionaal oplevert. (Analoog aan de gradiënt van een scalaire functie, die via het $\mathbb{R}^n$ -inproduct de richtingsafgeleide van de functie oplevert.)

De totale potentiële energie is eenvoudig te vinden uit de potentiële energie $\rho g z$ van een eenheidsvolume op hoogte z : de integraal over het ruimtelijk gebied levert (met weglating van een constante term)

$$P(\eta) = \frac{1}{2} \rho g \int \eta^2 dx dy. \quad (6)$$

De kinetische energie is — als gebruikelijk in de klassieke mechanica — de helft van de massa maal de snelheid in het kwadraat, nu geïntegreerd over het watervolume. Uitgedrukt in de snelheidspotentiaal is dat de Dirichlet-integraal behorende bij het Laplace-probleem, dus

$$K(\varphi, \eta) = \frac{1}{2} \rho \int \int_{-D}^{\eta} |\nabla \Phi|^2 dz dx dy. \quad (7)$$

Merk op dat K nu ook van η afhangt omdat Φ moet voldoen aan de Laplace-vergelijking met randvoorwaarde $\Phi = \varphi$ op $z = \eta(x, y, t)$. Aan de bodem, $z = -D$, moet de normaalafgeleide nul zijn.

Omdat K een kwadratische functionaal is in φ , kan de Dirichlet-integraal geschreven worden als $K(\varphi, \eta) = \frac{1}{2} \rho \int \varphi N(\eta) \varphi dx dy$. Hierin is $N(\eta) \varphi$ de normaalafgeleide op het

oppervlak, lineair in φ met de zogenaamde Dirichlet-to-Neumann (D-to-N) operator $N(\eta)$ die van η afhangt. Daarmee vinden we meteen dat $\delta_\varphi H = \delta_\varphi K = N(\eta)\varphi$ en daarmee de kinematische voorwaarde op het oppervlak. Voor de dynamische vergelijking moet bij het nemen van de partiële variatieafgeleide naar η bedacht worden dat de potentiaal $\Phi = \varphi$ ook op het gevarieerde oppervlak moet gelden.

Voor het speciale geval van infinitesimale golfhoogte worden de niet-lineaire termen verwaarloosd. Boven een vlakke bodem is het Laplace-probleem dan oplosbaar met Fouriertheorie. De lineaire vergelijkingen worden dan

$$\begin{cases} \partial_t \eta = N_0 \varphi, \\ \partial_t \varphi = -g\eta. \end{cases} \quad (8)$$

Hierin is $N_0 = N(\eta = 0)$ de ruimtelijke (pseudo-differentiaal) operator die in de spectrale ruimte werkt als een vermenigvuldiging: $\widehat{N_0 \varphi}(k) = \Omega(k)^2 \hat{\varphi}(k)/g$ met $\Omega(k)^2 = gk \tanh(kD)$ waarin k het golfgetal, of in 2D de lengte van de golfvector, is. De vergelijkingen leiden dan (na eliminatie van φ) tot het resultaat dat tijdharmische oplossingen met frequentie ω aan de zogenaamde dispersierelatie moeten voldoen:

$$\omega^2 = \Omega(k)^2. \quad (9)$$

Variationele Boussinesq-modellen

De variationele beschrijving hebben we op vele manieren uitgebuit voor het vinden van Hamilton-consistente benaderingen. Dit garandeert energiebehoud, hetgeen bij klassieke Boussinesq-vergelijkingen vaak niet eenvoudig te controleren is. Ter onderscheiding spreken we van variationele Boussinesq-modellen. In de laatste jaren hebben we een tweetal consistente benaderingen ontwikkeld.

Een klasse van zogenaamde AB-vergelijkingen [4–6, 14] ontstaat door een ontwikkeling van de D-to-N operator $N(\eta)$ in willekeurige orde in η . Deze benadering is geldig voor alle golf lengten. Daardoor kan met een spectrale implementatie aan de niet-rationale dispersierelatie exact worden voldaan boven een vlakke bodem. Voor variërende bodem nemen we een quasi-homogene benadering. Daarmee vervalt een van de grootste beperkingen van klassieke Boussinesq-vergelijkingen, die namelijk slechts geldig zijn voor voldoende grote golf lengten. Als gevolg hiervan kunnen wij nu ook de lange golven

die uit de diepe oceaan naar de kust toe korter worden, goed berekenen. Ook bekende nadelen van spectrale implementaties (de periodiciteit van de rekengebieden, het toevoegen van harde wanden et cetera.) zijn inmiddels overwonnen door al deze effecten te bewerkstelligen via brontermen in de Hamiltoniaanse vergelijkingen. De simulaties zijn uiterst efficiënt, op een laptop normaliter minder dan 10 procent van de fysische tijd.

In een andere klasse van variationele Boussinesq-modellen [4], met een *Finite Element*-implementatie, gebruiken we een extra optimaliteitscriterium voor het modelleren van de kinetische energie [1, 5]. Dat is gebaseerd op het principe van Dirichlet, namelijk dat de potentiaal die de minimale waarde van de kinetische energie geeft precies de oplossing is van het Laplace-probleem. Dit leidt tot effectieve benaderingen voor de kinetische energie door voor de potentiaal in de Dirichlet-functionaal slechts een combinatie van enkele geoptimaliseerde verticale profielen te nemen. Voor elke geometrie en gegeven binnenkomende golven wordt dan a priori de beste dispersierelatie bepaald. Dit model wordt in het bijzonder gebruikt voor gecompliceerde havengeometrieën boven sterk veranderende bodemdiepten.

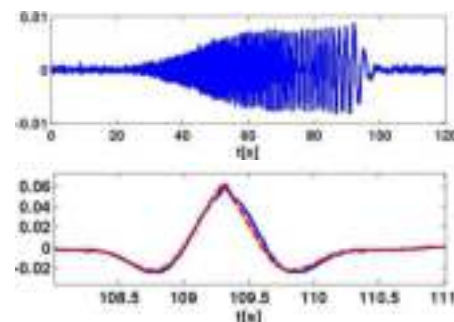
Onregelmatige golven en freak waves

Karakteristieke golfgrootheden

Bij het waarnemen van golven zien we voornamelijk een golfprofiel, een momentopname; of een zogenaamd tijdsignaal: de op en neer gaande beweging van het wateroppervlak op een vaste plaats. Grootheden als golf lengte L , periode T van het tijdsignaal en de golfvoortplantingssnelheid c kunnen alleen precies gedefinieerd worden voor harmonische golven in één horizontale dimensie. Met $k = 2\pi/L$ het golfgetal en $\omega = 2\pi/T$ de frequentie wordt een harmonische golf beschreven door $\eta(x, t) = A \cos(kx - \omega t + \psi)$, waarin A de amplitude van de golf is (de helft van de golfhoogte) en ψ een fasefactor. De snelheid c is dan $c = L/T$. Om in de limiet voor kleine amplitude te voldoen aan de lineaire vergelijkingen, moet er een relatie bestaan tussen k en ω , de dispersierelatie $\omega^2 = \Omega(k)^2$ zoals hierboven genoemd. Omdat $\Omega(k)$ voor positieve k een concave functie is, volgt hieruit dat de fasesnelheid $c = \omega/k$ op zodanige manier afhangt van k dat lange golven een grotere snelheid hebben dan kortere golven.

Focuserende golf

Een mooi voorbeeld om het effect van lineaire

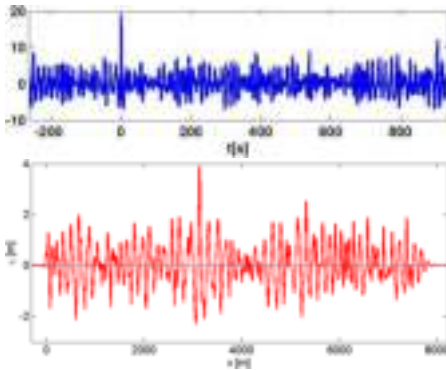


Figuur 1 Boven: Het gemeten tijdsignaal op positie 10 m in de golftank (MARIN-experiment 202002); horizontaal de tijd in seconden, verticaal de waterhoogte in meter. Langere en snellere golven volgen de langzamere korte golven. Onder: Superpositie van alle golven op een plaats doordat de snelle golven de langzame hebben ingehaald. Constructieve interferentie leidt tot een grote golfhoogte. Het gemeten tijdsignaal (blauw) vergeleken met het berekende (rood) tijdsignaal op positie 50,2 m. De simulatie is uitgevoerd met de tweede-orde AB-vergelijking met exacte dispersierelatie [11].

dispersie te illustreren is het geval van een zogenaamde focuserende golf, een van de methoden om hoge golven op te wekken in een golftank (tank is vaktiaal voor proefbassin), zie Figuur 1. Aan één zijde van de tank wordt een rij van golven opgewekt. Opeenvolgende golven krijgen langere golf lengte en iets hogere amplitude. De langere golven hebben grotere snelheid en halen de kortere golven in. Als alles mooi afgestemd wordt, komen op één plaats en tijd alle golven samen tot een golf met een grote amplitude die de som is van de amplitudes van de afzonderlijke golven als ter plaatse alle fases nul zijn. Vanwege symmetrie vervalt daarna die hoge golf weer in de afzonderlijke golfcomponenten, nu met de langste golven vooruitlopend op de langzamere korte golven. Het verspreiden van de golven die op zee worden opgewekt gedurende een storm laat eenzelfde patroon zien. Aan de kust worden de langste golven (die verstoringen tijdens de lange reis het best kunnen weerstaan) als deining waargenomen.

Onregelmatige golven

Superpositie van verschillende harmonische golven, zoals in het voorbeeld van de focuserende golf, kan leiden tot zeer verschillende golfprofielen, zoals bekend uit de Fouriertheorie. Golven opgewekt door de wind hebben ook verschillende frequenties, en iedere frequentie (band) kan met een verschillende energie vertegenwoordigd zijn, hetgeen weer te geven is in een energiespectrum. Maar windgolven hebben bovendien een stochastisch karakter doordat ook de fase per frequentie verschillend zal zijn, hetgeen gemiddeld wordt als random keuze met uniforme verdeling over het interval $[0, 2\pi)$. We spreken dan over onregelmatige golven.



Figuur 2 Boven: Met horizontaal de tijd en verticaal de waterhoogte in meter toont dit tijdsignaal de zogenaamde Draupner freak wave op $t=0$ gemeten vanaf het Draupner-platform op 1 januari 1995. De tophoogte is 18,5 m in de omringende golven met gemiddelde tophoogte van ongeveer 4 m. (Data van Dr. Sverre Haver, Statoil.) Onder: Profiel van een onregelmatige golf met horizontaal de afstand in meters. Deze berekening komt overeen met een experiment (# 103001) in een golf tank van MARIN, geschaald met een factor 50. Op $x=0$ wordt de golf opgewekt overeenkomstig de waterhoogte op een meetpositie in de tank. De naar rechts lopende golven worden in de simulatie aan het einde van het interval kunstmatig gedempt. De golf heeft een significante golfhoogte van $H_s=3,2$ m. Maar rond positie 3000 is een 'toevallige' freak wave te zien, met een golfhoogte op die plaats van 6,8 m. In een simulatie blijkt dat die hoge golf voor lange tijd een zeer grote hoogte behoudt door constructieve interferentie [6].

Figuur 2 toont een zeer speciaal golfveld op zee, en onregelmatige golven in een golf-tank. Deze laatste worden opgewekt om het gedrag van schepen te kunnen testen in zo realistisch mogelijke zee-situaties. Overigens: de schaling van laboratorium naar de werkelijke situatie is mogelijk omdat in beide gevallen de gravitatie (dimensie m/s^2) de onderliggende processen bepaalt: schaling van tijd gaat dan met de wortel uit de schaling van afstanden.

Met de randomfase-aanname kunnen uit de energieverdeling statistische grootheden gedefinieerd worden voor onregelmatige golven. Een belangrijke grootheid is de significante golfhoogte H_s . Oorspronkelijk bedoeld als de 'gemiddelde' golfhoogte zoals die geschat zou worden door een ervaren waarnemer, blijkt die waarde in feite dichter bij vier maal de standaarddeviatie van het spectrum te liggen. Daaruit volgt dan (omdat de golfhoogten aan een Gaussische verdeling voldoen) dat H_s ongeveer de gemiddelde hoogte is van de verzameling van de 33 procent hoogste golven. Blijkbaar baseert de 'waarnemer' zijn schatting vooral op de meest zichtbare golven.

Niet-lineariteit

De niet-lineaire termen in de vergelijkingen leiden in samengestelde golven tot een herverdeling van energie tussen de golven: de niet-lineariteit verandert het spectrum. Dat is temeer zo als het spectrum zeer nauw is, dat

wil zeggen als begonnen wordt met een energieverdeling in een smal interval. Het effect van niet-lineariteit kan geïllustreerd worden aan de hierboven genoemde (KdV) soliton. Het bestaan van deze golfvorm is een subtiële balans tussen dispersie, die de golfvorm breder zou maken, en niet-lineariteit die de vorm steiler zou maken. Vanuit een theoretisch gezichtspunt is het aardig op te merken dat een soliton ook oplossing is van een *constrained minimum*-principe voor de energie, onder de voorwaarde van gegeven horizontale impuls (een relatief evenwicht in de Hamiltoniaanse beschrijving omdat de impuls een van de Hamiltoniaan onafhankelijke bewegingsgrootte is).

Freak waves

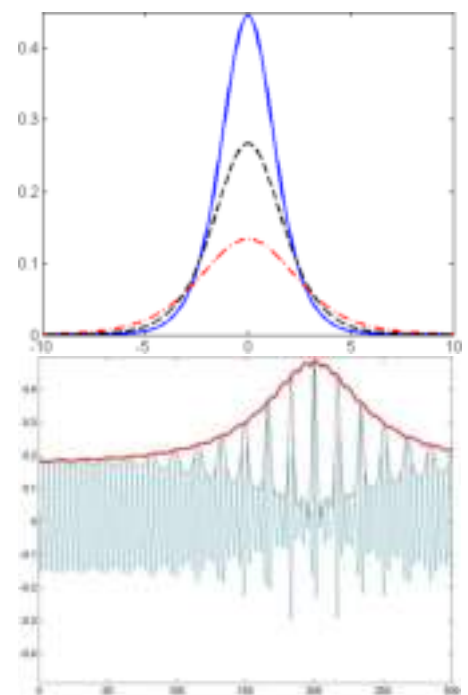
Zogenaamde 'freak' of 'rogue' waves of monstergolven zijn uitzonderlijk hoge golven (in vergelijking met omringende golven) die zonder aanwijsbare oorzaak plotseling optreden. Van oudsher waren verhalen bekend van zeevarenden over onverwachte monstergolven, maar die werden meestal afgedaan als visserslatijn. Naast aanwijsbare schade aan schepen, en veel schepen die op onverklaarbare wijze zijn vergaan, werd het meest overtuigende bewijs van het bestaan van monstergolven gegeven door videowaarnemingen op het Draupner-platform in de Noordzee gedurende een storm op nieuwjaarsdag 1995. De video registreerde het passeren van een golf met een tophoogte van 18,5 meter, in een omringende zee met significante golfhoogte van 12 meter. Schade aan een dek hoog boven het wateroppervlak gaf extra bevestiging. De hoge top van bijna drie maal de gemiddelde tophoogte is uitzonderlijk. In Figuur 2 is het tijdsignaal van de geregistreerde Draupner-golf te zien.

Sinds die tijd is er veel onderzoek gedaan naar freak waves. Maar veel vragen over het ontstaan, en het voorspellen van dit soort golven zijn nog steeds niet beantwoord, zelfs niet voor langkammige (1D-) golven. Twee gezichtspunten komen naar voren, overeenkomend met de twee karakteristieke effecten in golven.

Een onderzoeksrichting concentreert zich op niet-lineaire effecten, het uitwisselen van energie tussen golven. Dit is al zichtbaar voor een gemoduleerde harmonische golf. Als de modulatielengte voldoende groot is, vindt energie-uitwisseling plaats waardoor de hoogste golven groter worden ten koste van de anderen, bekend als de Benjamin-Feir-instabiliteit. Later werd gevonden dat deze instabiliteit een aanzet is tot een niet-lineaire

golftrein, waarvan de omhullende een niet-bewegend solitonprofiel heeft. De naam daarvoor, *soliton on finite background*, verwijst naar de KdV-solitonvorm, maar met het verschil dat de limietwaarden niet nul zijn maar naderen naar een achtergrond van harmonische golven, zie Figuur 3. Als dit mogelijk is voor een harmonische golf als drager en basis van de instabiliteit, is dat misschien ook mogelijk met een onregelmatige golf als drager. De instabiliteit treedt alleen op voor voldoende nauw spectrum, gekarakteriseerd door een voldoende grote waarde van de zogenaamde BF-index. Voor scheepvaart over de oceaan worden tegenwoordig veiligheidswaarschuwingen gegeven voor gebieden waar de BF-index hoog is.

Een andere mogelijke reden voor een hoge golf is constructieve interferentie, zoals we al zagen bij de focuserende golf. In feite is het statistisch gezien al zo dat op grond van de Gaussische verdeling van golfhoogten (als een gevolg van de centrale limietstelling) een golfhoogte van twee maal H_s eens in de duizend golven voorkomt. In een storm van bijvoorbeeld drie uur, met golven met periode van negen seconden, zal dat dus statistisch



Figuur 3 Boven: Enkele karakteristieke KdV-solitonprofielen, gegeven in dimensieloze met de diepte geschaalde grootheden door $\eta(x) = \alpha \text{sech}^2(\beta x)$ met $\alpha = 4\beta^2/3$, die onvervormd voortbewegen met constante snelheid $V = 1 + \alpha/2$. Onder: 'Soliton on finite background': de modulatie van de naar rechts lopende gemoduleerde harmonische golf wordt van links naar de extreme positie steeds groter ten gevolge van niet-lineaire effecten, waarna op symmetrische wijze een afname plaatsvindt. De 'soliton' is nu de stilstaande rode omhullende waaronder dit modulatieproces zich afspeelt: het is de rand van het natte deel van de wand van de golf tank waarin dit proces zich voltrekt [3].

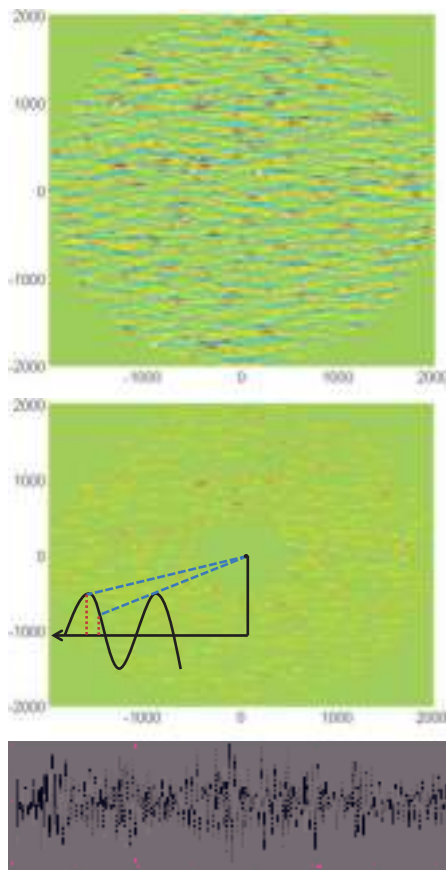
een keer voorkomen, op elke plaats in het stormgebied. In lange tijdreeksen van experimenten bij MARIN met onregelmatige golven hebben wij inderdaad verschillende freak waves aangetroffen, zie Figuur 2. Nadere analyse leerde dat de BF-index te klein was, net zoals bij de Draupner-golf, om die hoge golven te verklaren ten gevolge van niet-lineaire effecten. Numerieke simulatie met lineaire en niet-lineaire codes bevestigde dat. In deze onderzochte gevallen is de hoge golf een direct effect van de lineaire constructieve interferentie van enkele naburige hoge golven, daarmee

andere soortgelijke observaties bevestigend [15]. Dat geeft tenminste een aanwijzing dat ook in onregelmatige golven (met een breed spectrum) zoals die in windzeeën, lineaire interferentie tot hoge, freak-achtige golven kan leiden.

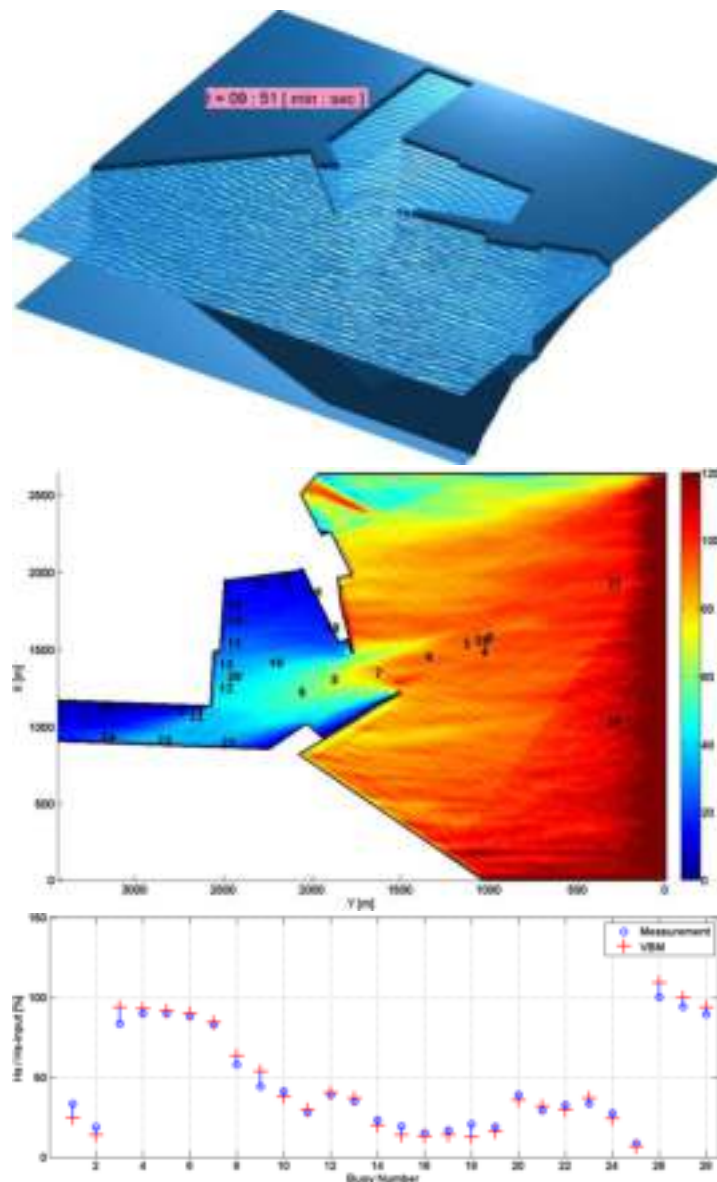
Golfvoorspelling uit radarbeelden

Voor meerdere offshore-activiteiten zou het zeer wenselijk zijn een voorspelling te hebben van de te verwachten aankomende golven. Denk bijvoorbeeld aan het plaatsen van windmolens, of het landen van een helikop-

ter op een scheepsdek. In beide gevallen is een relatief korte periode van kalme golven voldoende om zo'n actie uit te voeren, hetgeen met een goede voorspelling dan ook zou kunnen in overigens tamelijk ruwe zeeën. Een x-band radar die op veel schepen aanwezig is kan de golven waarnemen in een ringvormig gebied tussen 500 en 2000 meter van de radarpositie. Momenteel kan uit de (sterk vertekende) beelden slechts een statistische grootheid, zoals de piekperiode (of de daaraan gerelateerde piekgolflengte) bepaald worden, maar bijvoorbeeld niet de ab-



Figuur 4 Boven: Momentopname van een synthetisch gemaakte windzee in een schijf met straal van 2000 m met de radarpositie in de oorsprong. De golven komen uit het noorden naar de radarpositie, met een spreiding van ongeveer 600 naar beide kanten. De significante golfhoogte is $H_s = 3$ m, en de piekperiode $T_p = 9$ s. Met een bodemdpte van 50 m zoals hier genomen is de piekgolflengte ongeveer 125 m. Midden: Een geïdealiseerde versie van een radarbeeld van de windzee. Met de radar op een hoogte van 30 m boven het wateroppervlak in het midden van het typisch ringvormige gebied van radarbereik zijn de golven berekend die zichtbaar zijn voor de radar. Het schaduw effect wordt geïllustreerd in het plaatje als inzet. De kleurschakering van een echt radarbeeld is gerelateerd aan de signaalreflectie. De afname van de zichtbaarheid met de afstand maakt het mogelijk de werkelijke golfhoogten te bepalen. Onder: Het tijdsignaal van de waterhoogte van de windzee op de radarpositie (blauw) vergeleken met een voorlopig resultaat van de berekende simulatiewaarden (rood), over een interval van 12 min. Het verschil in het begin tussen de signalen geeft een indicatie van de tijd die nodig is, ongeveer 90 s, om op de radarpositie de volledige informatie te krijgen van de golven buiten de binnenring met een straal van 500 m.



Figuur 5 Simulaties met het geoptimaliseerde variationele Boussinesq-model met een *Finite Element*-implementatie voor de haven van Limassol, Cyprus. In de 3D-figuur is ook de bodem zichtbaar. De fijnheid van het onregelmatige numerieke grid hangt af van de diepte, zodat de nauwkeurigheid gelijk blijft bij veranderende golflengte. Van belang voor de kwaliteit van de haven is het 'verstoringprofiel'. Dat is op elke plaats het quotiënt van de significante golfhoogte en de golfhoogte van de binnenkomende golven. In het 2D-bovenaanzicht is dit quotiënt aangegeven door kleurschakering, van rood (100%) tot blauw (0%). Met numerieke simulaties kunnen verschillende golven, uit verschillende richtingen worden behandeld, meer dan in laboratoriumproeven. Enkele simulatieresultaten zijn vergeleken met die van laboratoriummetingen zoals getoond.

solute golfhoogten, noch de zich voortplantende golven en dus ook niet de waterhoogte op de radarpositie. Lopend onderzoek in onze groep heeft als doel een deterministische voorspelling te geven van de waterhoogte op de radarpositie en verloopt ruwweg langs de volgende lijnen.

Een zee zal in veel gevallen bestaan uit een groot aantal kortkammige golven die opgewekt zijn door een nabije sterke wind. Karakteristieke getallen zijn piekperioden tussen 5 en 12 seconden en golfhoogten tussen 1 en 4 meter, voornamelijk afhankelijk van de windsterkte en de duur daarvan. De kortkammigheid (ter onderscheiding van de langkammige (1D-) golven) is een gevolg van fluctuaties van de windsterkte en -richting tijdens het generatieproces. Een voorbeeld van een (synthetisch gemaakte) zee met $T_p = 9$ s en $H_s = 3$ m is te zien in Figuur 4 (boven). De kleurschakering correspondeert met de waterhoogte. In veel gevallen zal naast een windzee ook deining uit een andere richting aanwezig zijn: veel langere, en minder gespreide golven, afkomstig van een storm op een verre locatie. Tezamen geeft zo'n multimodaal golfveld een nog gecompliceerder beeld.

Het vertekende beeld dat een radar geeft van zo'n golfveld lijkt enigszins op dat in Figuur 4 (midden). In die figuur is het schaduwgolfveld te zien (in het typische ringvormige bereik van de radar). Dat is het gevolg van het geometrisch effect dat een dichter bij de radar zijnde golf het radarsignaal afdekt voor een deel van de daaropvolgende golf. Daarnaast is de kleurschakering nu een maat van de intensiteit van het teruggekaatste deel van

het uitgezonden signaal en niet van de waterhoogte. Omdat de terugkaatsing bepaald wordt door de interactie van het radarsignaal met de rimpels op de golven (korte capillaire golfjes), is er geen eenvoudige schaling van de intensiteit naar de werkelijke waterhoogte. Ons onderzoek heeft geleerd dat de (geometrische) zichtbaarheid als functie van de afstand tot de radar een karakteristieke functie is van de twee dimensieloze grootheden die dit geometrisch effect bewerkstelligen; en daaruit kan dan wel, bij benadering, een waarde voor de golfhoogte worden berekend uit alleen de radarbeelden. Om de waterhoogte op de radarpositie te bepalen uit de (synthetische) radarbeelden, verbeteren we eerst de beelden door het opvullen van de schaduwgebieden met zo realistisch mogelijke golfpatronen; voorlopig is dat meer een kunst dan kunde, vooral als de gebieden groot zijn, zoals het geval is achter een hoge, freak-achtige, golf. Met de verbeterde beelden construeren we een dynamisch middellingsproces. Door assimilatie krijgen we dan een evolutie die het radargat van 500 meter kan overbruggen en een benadering geeft van de waterhoogte op de radarpositie. Voor een zee van windgolven blijkt de reconstructie opmerkelijk goed te gaan (althans voor de lineaire golfvelden en propagatie die we tot nu toe uitgevoerd hebben), leidend tot een benadering met een correlatie van meer dan 90 procent, zie Figuur 4c [7]. Het blijkt dat we in feite de windgolven nauwkeurig genoeg bepalen vanaf een afstand van ongeveer 1000 meter vanaf de radar, waaruit de lengte van de voorspellingshorizon berekend kan worden.

Havengolven

Ook voor het ontwerpen van groot- en kleinschalige waterbouwkundige constructies is het begrijpen en nauwkeurig simuleren van golven van belang. De golven in een haven, bijvoorbeeld, moeten zo kalm mogelijk zijn, ook als de golven daarbuiten hoog zijn, terwijl het gewenst is dat de toegang tot de haven toch zo ruim mogelijk is. Terugkaatsing van golven aan de kadewanden kan leiden tot resonante staande golven (seiches) die het laden en lossen onmogelijk maken. Numerieke simulaties moeten daarom vooral in staat zijn (gedeeltelijk) reflecterende wanden, voor veel verschillende binnenkomende golfvelden, te modelleren. Een tweede eis is dat de kunstmatige numerieke rand op zee, waar de golfvelden worden opgewekt, doorlaatbaar is voor de teruggekaatste, weer naar buiten lopende golven. Ter illustratie is in Figuur 5 een resultaat te zien van golven in een haven waarvoor meetgegevens (op laboratoriumschaal) van Deltares beschikbaar zijn.

Dankwoord

De boven beschreven onderwerpen zijn de laatste jaren onderzocht door een aantal PhD-studenten werkzaam op UT en LabMath-Indonesia, Bandung, gefinancierd door projecten van NWO, STW, KNAW en IOP-Promised. We danken Andonowati, Gert Klopman en Onno Bokhove voor de mede begeleiding van enkele van die studenten: Arnida Latifah, Wenny Kristina, Lie She Liam, Ivan Lakhturov, Natanael Karjanto, Didit Adytia en Andreas Parama Wijaya. MARIN en Deltares worden bedankt voor het beschikbaar stellen van de experimentele data. ☞

Referenties

- 1 D. Adytia en E. van Groesen, Optimized Variational 1D Boussinesq modelling of coastal waves propagating over a slope, *Journal Coastal Engineering* 64 (2012), 139–150.
- 2 L.J.F. Broer, On the Hamiltonian theory of surface waves, *Appl. Sci. Res.* 30 (1974), 430–446.
- 3 E. van Groesen, Andonowati en N. Karjanto, Displaced Phase-Amplitude variables for Waves on Finite Background, *Phys. Lett. A* 354 (2006), 312–319.
- 4 E. van Groesen en Andonowati, Variational derivation of KdV-type of models for surface water waves, *Phys. Lett. A* 366 (2007), 195–201.
- 5 E. van Groesen, Andonowati, L. She Liam en I. Lakhturov, Accurate modelling of unidirectional surface waves, *Journal of Computational and Applied Mathematics* 234 (2010), 1747–1756.
- 6 E. van Groesen en I. van der Kroon, Fully dispersive dynamic models for surface water waves above varying bottom, Part 2: Hybrid spatial-spectral implementations, *Wave Motion* 49 (2012), 198–211.
- 7 E. van Groesen, A.P. Wijaya en Andonowati, Spatial evolution scenario from synthetic radar images to sea states near the radar, 2013
- 8 N. Karjanto en E. van Groesen, Qualitative comparisons of experimental results on deterministic freak wave generation based on modulational instability, *Journal of Hydro-environment Research* 3 (2010), 186–192.
- 9 G. Klopman, E. van Groesen en M.W. Dingemans, A variational approach to Boussinesq modelling of fully non-linear water waves, *J. Fluid Mech.* 657 (2010), 36–63.
- 10 I. Lakhturov, D. Adytia en E. van Groesen, Optimized Variational 1D Boussinesq modelling for broad-band waves over flat bottom, *Wave Motion* 49 (2012), 309–322.
- 11 A.L. Latifah en E. van Groesen, Coherence and Predictability of Extreme Events in Irregular Waves, *Nonlin. Processes Geophys.* 19 (2012), 199–213.
- 12 J.C. Luke, A variational principle for a fluid with a free surface, *J. Fluid Mech.* 27 (1967), 395–397.
- 13 J.W. Miles, On Hamilton's principle for surface waves, *J. Fluid Mech.* 83 (1977), 153–158.
- 14 L. She Liam en E. van Groesen, Variational derivation of KP-type equations, *Phys. Lett. A* 374 (2010), 411–415.
- 15 A. Slunyaev, E. Pelinovsky en C.G. Soares, Modeling freak waves from the North Sea, *Applied Ocean Research* 27(1) (2005), 12–22.
- 16 V.E. Zakharov, Stability of periodic waves of finite amplitude on the surface of a deep fluid, *J. Appl. Mech. Tech. Phys.* 2 (1968), 190–194.

Arjen Doelman

Mathematisch Instituut
Universiteit Leiden
doelman@math.leidenuniv.nl

Vivi Rottschäfer

Mathematisch Instituut
Universiteit Leiden
vivi@math.leidenuniv.nl

Lotte Sewalt

Mathematisch Instituut
Universiteit Leiden
lotte@math.leidenuniv.nl

Huib de Swart

IMAU
Universiteit Utrecht
h.e.deswart@uu.nl

Antonios Zagaris

Afdeling Toegepaste Wiskunde
Universiteit Twente
a.zagaris@utwente.nl

Julia Zijlstra

Research Group Computational Dynamics
CWI, Amsterdam
julia.zijlstra@cwi.nl

De complexe dynamica van fytoplankton

Fytoplankton speelt niet alleen een belangrijke rol in de voedselketen in oceanen en kustwaters, ook voor het klimaat is een goed begrip van de fytoplanktodynamica van groot belang. Met behulp van dynamische modellen analyseren Arjen Doelman, Vivi Rottschäfer, Lotte Sewalt, Huib de Swart, Antonios Zagaris en Julia Zijlstra dit gedrag. Een overzicht van verschillende typen modellen wordt gegeven, en het dynamische gedrag wordt geanalyseerd met verschillende technieken, zoals centre manifold reduction en meerschalenanalyse. Het gevonden gedrag kan erg complex zijn, zo kan er zelfs spatio-temporele chaos voorkomen.

Anderhalf miljard jaar geleden heeft de ontwikkeling van leven op aarde een enorme stap voorwaarts gemaakt toen een eukaryotische cel een fotosynthetische bacterie opnam en integreerde. Deze oerfusie heeft geresulteerd in de huidige situatie waarbij organismen met zogeheten *plastiden* zonne-energie om kunnen zetten naar biochemische energie. Deze omzetting is essentieel voor alle organismen op aarde die deze plastiden niet bevatten, zoals de mens, omdat op deze manier de zon een voedselketen kan voorzien van energie.

Naast alle groen dat we op land kennen, vindt een groot deel van de fotosynthese plaats binnen minuscule oceanorganismen. Deze organismen, met verzamelnaam *fytoplankton*, zijn de grootste voedselvoorziening in de oceaan. Doordat ze aan de basis van

veel voedselketens staan, produceren ze naar schatting de helft van alle voedsel op aarde. Ze worden geconsumeerd door kleine schaaldieren, *zoöplankton*, die vervolgens gegeten worden door een breed scala aan waterdieren, van kleine vissen tot walvissen. Bij fotosynthese worden kooldioxide en water omgezet in koolhydraten en zuurstof. Op deze manier verzorgt fytoplankton de helft van alle zuurstofproductie op de planeet. Het voert jaarlijks 45 gigaton koolstof uit de atmosfeer af, diep de oceaan in, wat een grote invloed heeft op ons klimaat en dus ook op de mens. Deze relatie is tweezijdig, want klimaatveranderingen hebben ook een directe uitwerking op de fytoplanktonsoorten. Ook zijn menselijke factoren van invloed op de planktongroei, bijvoorbeeld door de dump van gifti-

ge stoffen of veranderend watermanagement. De enorme opeenhoping van fytoplankton, *bloei*, kan zich uitstrekken over honderden vierkante kilometers en zelfs tot onder de ijskappen. De levensduur wordt voornamelijk bepaald door geofysische en milieuprocessen: de aanwezigheid van licht aan het wateroppervlak, de hoeveelheid voedsel — nutriënt — op de bodem, transportfenomenen, en de mate waarin fytoplankton wordt geconsumeerd door zoöplankton. Aangezien licht vanaf het zeeoppervlak naar beneden diffundeert en voedingsstoffen — nitraat, fosfaat, silicium en ijzer — zich van de bodem naar boven verspreiden, bevindt de grootste concentratie fytoplankton zich op een optimale diepte waar beide voldoende beschikbaar zijn. Een dergelijke opeenhoping noemen we *Deep Chlorophyll Maximum* (DCM) en deze kan zowel permanent (in de tropen) als seizoensafhankelijk (op hogere breedtegraden) aanwezig zijn.

Om de groei van fytoplankton te beschrijven en te begrijpen, moeten theoretische ecologie en toegepaste wiskunde de krachten bundelen. Echter, diepgewortelde onzekerhe-

den in dergelijke modellen maken dat het formuleren van een model op planetaire schaal onmogelijk, zo niet onzinnig is. Er zijn verschillende modellen die de nadruk leggen op verschillende aspecten. In dit artikel zullen we een aantal daarvan kort beschouwen en een overzicht geven van het dynamische gedrag dat ze vertonen.

De eerste modellen

Kort na de introductie van roofdier–prooi-modellen in de jaren twintig begon Richard Fleming, oceanograaf bij de *Scripps Institution of Oceanography*, de invloed van zoöplankton op de groei van fytoplankton te bestuderen. Zijn model bestond uit een eerste-orde, lineaire gewone differentiaalvergelijking (ODE) voor fytoplanktonconcentratie (P) waarbij zoöplanktonniveaus handmatig werden ingevoerd. In dit model ontbraken geofysische processen en natuurlijke limiterende factoren volledig, maar desondanks kon Fleming redelijke overeenstemming met data van zoöplanktonniveaus uit Het Kanaal krijgen.

Dit werk werd uitgebreid met een ruimtelijke dimensie door Gordon Riley, oceanograaf bij de *Woods Hole Oceanographic Institute* (WHOI). Om milieuprocessen te kunnen modelleren voegde hij aan het oorspronkelijke model diffusie, advection en lichtintensiteit toe:

$$P_t = DP_{zz} - VP_z + (f(L) - \ell)P. \quad (1)$$

Hier beschrijven constanten D en V de mate van diffusie en advection, respectievelijk. De fytoplanktongroei op diepte z en tijd t wordt gekwantificeerd door f , een stijgende functie voor de lichtintensiteit $L(z, t)$. Deze wordt gemodelleerd door het gegeven dat licht geabsorbeerd wordt door water en fytoplankton in de waterkolom erboven, zodat

$$L(z, t) = L_I(t)e^{-k_L z + k_P \int_0^z P(s, t) ds}. \quad (2)$$

Hier meten k_L en k_P de troebelheid van water en doorzichtigheid van fytoplankton; L_I is de intensiteit van het invallende licht. Voor f wordt vaak de stijgende, begrensde functie $f(L) = f_{\max}L/(L_H + L)$ gekozen; hierbij zijn $f_{\max}$ en L_H constanten. Stationaire profielen ($P_t = 0$) worden door een niet-autonome ODE beschreven. Het vinden van expliciete oplossingen van zulke vergelijkingen is eerder uitzondering dan regel, en er is dus een benadering nodig. Deze volgt uit het feit dat in

de bovenste laag van de oceaan de groeiomstandigheden aanzienlijk gunstiger zijn. Riley zelf versimpelde zijn model nog iets verder en voerde voor $f(L)$ een expliciete stapfunctie op diepte z_* in die stuksgewijs constant is en alleen van de diepte z afhangt. In dat geval is (1) dus lineair. Hij leidde een scherpe conditie voor het bestaan van een plankton populatie af,

$$\ell < f(z_*) - \frac{V^2}{4D}. \quad (3)$$

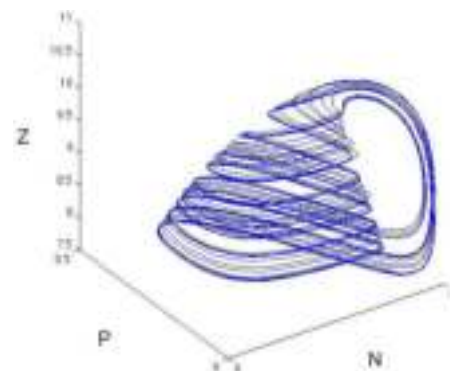
Uiteraard kwam Riley niet verder dan een summier begrip van fytoplanktodynamica. De totale biomassa, bijvoorbeeld, kon op deze manier niet voorspeld worden omdat (1) lineair is. Desondanks diende (1) als springplank voor meer omvangrijke modellen, waarbij voorwaarde (3) gebruikt werd als leidende-orde-benadering.

In de jaren zestig en zeventig werden Rileys modellen populairder door het werk van de eminente wiskundig-bioloog John Steele (voorzitter van WHOI). Steele vulde het model aan met biologische componenten — voornamelijk zoöplankton en nutriënt — en werkte met niet-lineaire interactietermen. Hij was ook de eerste die niet alleen stationaire maar ook tijdsafhankelijke profielen onderzocht. Zijn werk was een samenspel van theorie en experimenten en wordt beschouwd als zeer invloedrijk.

De scalaire modellen van Riley en Steele werden door vele onderzoekers geanalyseerd, zie [16] voor een gedetailleerde beschrijving van de literatuur. In wiskundig opzicht bereikte dit onderzoek zijn hoogtepunt in de publicatie van [10], waar $f: \mathbb{R}_+ \rightarrow \mathbb{R}$ een strikt stijgende functie is. Hier bewezen de auteurs dat de globale attractor voor oplossingen van (1)–(2) gevormd wordt door een stationair profiel. Wanneer een dergelijk, uniform positief profiel bestaat, is dit globaal stabiel; zo niet, dan verdwijnt de fytoplanktonconcentratie. In [8] wordt een gedetailleerde voorwaarde voor het bestaan van een dergelijk profiel gegeven, die uiteraard met een algemenere vorm van (3) overeenkomt.

NPZ-modellen

Een van de vele soorten gebruikte modellen is het NPZ-model. NPZ-modellen zijn systemen van drie ODEs die de interactie van nutriënt (N), fytoplankton (P) en zoöplankton (Z) beschrijven. De achterliggende structuur is dat fytoplankton de nutriënt opneemt, terwijl zoöplankton zich voedt met fytoplankton. Beide planktonsoorten dragen bij aan



Figuur 1 Chaotische 'teacup' attractor, gesimuleerd aan de hand van het NPZ-model in [7].

de hoeveelheid nutriënt [6], waardoor de drie componenten een voedselketen vormen. Een algemeen voorbeeld van een dergelijk NPZ-model is [3, 5]

$$\begin{aligned} N_t &= - \text{opname door } P + \text{sterfte } P \\ &\quad + \text{afval } Z + \text{afval van roofdier van } Z \\ &\quad + \text{balanstern}, \\ P_t &= \text{groei} - \text{sterfte} - \text{zinken} \\ &\quad - \text{opname door } Z, \\ Z_t &= \text{groei} - \text{sterfte}. \end{aligned} \quad (4)$$

Hierbij wordt het model beschouwd in een specifieke laag in de oceaan. Het zinken van fytoplankton betekent daarom dat er fytoplankton verdwijnt uit het systeem. De balanstern bepaalt dat de hoeveelheid nutriënt door een externe factor op een zekere balans gehouden wordt. De opname van nutriënt door fytoplankton wordt vaak gemodelleerd door de bovengenoemde functie f [3, 6–7].

Al in 1991 is in een overeenkomstig NPZ-model [7], numeriek, een chaotische attractor gevonden, zie Figuur 1. Het chaotische gedrag werd verklaard doordat het NPZ-model kon worden opgesplitst in twee subsystemen (NP en PZ) die beide oscillerend gedrag vertonen met een verschillende frequentie.

Acht jaar later werd een soortgelijke analyse gedaan in een NPZ-model dat binnen de klasse (4) valt [3], deze keer met een uitgebreide numerieke bifurcatieanalyse waarbij wederom een chaotische attractor en een opeenvolging van periodeverdubbelingen gevonden werd. Dit chaotische gedrag wordt nog eens bevestigd door een grootschalig praktijkexperiment in [1] waarin verschillende plankton populaties ruim zes jaar bestudeerd zijn.

NPZ-modellen blijken goed gereedschap te zijn om onderzoek te doen naar plankton en

in de afgelopen decennia zijn de modellen ook veelvuldig gekoppeld aan grootschaliger dynamica, zoals oceaancirculatie. Ook wordt gesteld dat NPZ-modellen als voorbeeld kunnen dienen voor de dynamica in andere ecosystemen [3]. De voornaamste reden voor het gebruik van deze systemen is de relatieve eenvoud waarmee ODEs bestudeerd kunnen worden. Desalniettemin zet Franks [6] vraagtekens bij het gemak waarmee deze modellen voor waar aangenomen worden. Hij roept op om nieuwe hypothesen over planktodynamica op te stellen en zorgvuldig om te gaan met conclusies die worden getrokken uit de huidige NPZ-modellen.

Beperkte hoeveelheid licht én nutriënt

In de vorige paragrafen zijn modellen beschouwd waarin fytoplanktongroei louter afhangt van de aanwezigheid van licht. Deze aanname resulteerde in dynamica waarbij op lange termijn alle oplossingen een stationair profiel benaderen. Echter, dit is zelden realistisch. In grote voedselarme gebieden van de oceaan is de situatie drastisch anders, en daarom is in [9] een model geïntroduceerd dat voedselschaarste meeneemt. Bovendien is de dynamica van dit model vele malen interessanter omdat oscillaties en zelfs spatiotemporele chaos voorkomen, zie Figuur 2.

Dit intrigerende dynamische gedrag is gedeeltelijk geanalyseerd in [14–15], waar we de ontwikkeling van een klein DCM patroon volgen dat ontstaat als reactie op veranderingen in de omgeving. In de buurt van de transkritische bifurcatie waar dit gebeurt, gebruiken

we standaard bifurcatie methodes, zoals centre manifold reduction (CMR) en methodes waarin meerdere schalen worden meegenomen. Door dit te koppelen aan klassieker methoden (WKB-analyse, integraalasymptotiek, speciale functies) breiden we ons begrip van het gedrag uit en relateren dit aan expliciete kenmerken van de omgeving. Zo is bewezen dat de DCM stabiel is daar waar CMR geldig is: alle kleine fytoplanktonpopulaties bloeien uit tot een DCM met een kleine, kwantificeerbare massa.

Wanneer we de analyse uitbreiden tot verder weg van het bifurcatiepunt, waar CMR niet meer toepasbaar is, is de analogie met modellen voor voedselrijke gebieden, zoals eerder beschreven, weg. Het spectrum bevat één enkele instabiele eigenwaarde, behorend bij de DCM, en daarnaast een aftelbare verzameling van stabiele eigenwaarden, afkomstig van de verspreiding van nutriënt. Alle oplossingen die bij de stabiele eigenwaarden horen, hebben alleen niet-lineaire invloed op de ontwikkeling van de DCM. Toch blijkt dit het mechanisme te zijn dat achter een uiteindelijke destabilisatie van de DCM, via een Hopf-bifurcatie, zit. Intrinsiek aan deze analyse is dat deze bifurcatie het begin van een cascade van periodeverdubbelingen aangeeft. Door onze berekeningen lijkt het ook waarschijnlijk dat een punt waar twee transkritische bifurcaties samenvallen als bron van chaotisch gedrag fungeert.

Dit werk is ook van belang voor het ontwikkelen van nieuwe wiskundige methoden, omdat het klassieke toegepaste analyse met dynamische-systeemtheorie combineert om een destabilisatie te beschrijven die buiten het bereik van de lokale analyse valt. Dat geeft ons voor het eerst een niet-triviale attractor in een planktonsysteem dat uitgebreid is met een ruimtelijke variabele. Ook onderstreept de analyse de enorme complexiteit van planktodynamica in voedselarm water.

Algemene reactie-diffusiesystemen

Naar aanleiding van [14–15] is een project gestart om te onderzoeken wat de bron van chaotisch gedrag in dergelijke reactie-diffusiesystemen is. Hiertoe beschouwen we een meer algemeen maar versimpeld modelprobleem:

$$\begin{aligned} U_t &= \mathcal{L}U + \alpha U + F(U, V), \\ V_t &= \varepsilon [\mathcal{L}V + \beta U + \gamma V + G(U, V)]. \end{aligned} \quad (5)$$

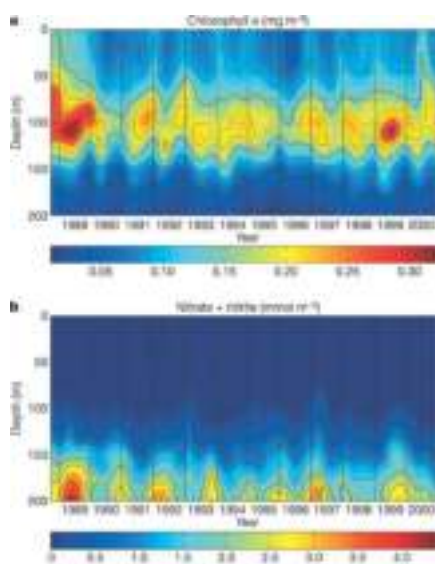
Hierbij zijn U, V functies van $(z, t) \in \Omega \times \mathbb{R}_+$

met $\Omega \subset \mathbb{R}^n$ en is $\mathcal{L}$ een lineaire differentiaaloperator. Functies F en G zijn ten minste kwadratisch en α, β en γ constanten. In de context van planktononderzoek beschrijft U de planktonconcentratie en V hoeveel de nutriënt afwijkt van de als randvoorwaarde gegeven maximale concentratie. Operator $\mathcal{L}$ modelleert de transportprocessen, parameter α meet de sterfte en maximale groei van plankton en $\gamma = 0$. De nutriëntopname is gemodelleerd door βU en $F = -G$. Systeem (5) is zodanig opgezet dat er voor het gelineariseerde probleem een verzameling $\mathcal{O}(\varepsilon)$ eigenwaarden $\{\mu_m\}_{m \geq 0}$ en een verzameling $\mathcal{O}(1)$ eigenwaarden $\{\lambda_m\}_{m \geq 0}$ bestaan. Zo benadert (5) de situatie in [15].

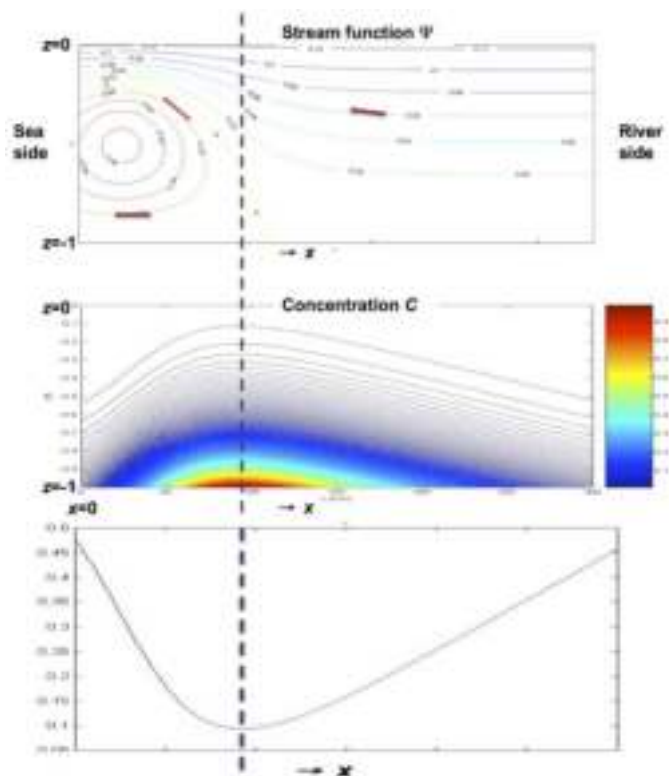
Afhankelijk van Ω en de randvoorwaarden, kunnen de eigenfuncties en de eigenwaarden expliciet bepaald worden. Daarmee kan het lineaire deel van (5) herschreven worden in termen van de eigenfuncties en eigenwaarden, waardoor het systeem verandert in een oneindig-dimensionaal stelsel van ODEs. Nu beschouwen we de situatie waarin de grootste eigenwaarde klein en positief is, $\lambda_1 = \varepsilon^\sigma r$, terwijl alle andere eigenwaarden stabiliteit impliceren. Als $\sigma > 1$, dus wanneer λ_1 veel kleiner is dan de (eerste) μ_k eigenwaarden, blijkt uit een formele verstoringsanalyse dat de leidende orde van het systeem gegeven wordt door een eendimensionale ODE die gekoppeld is aan algebraïsche relaties. Tot zover komt de analyse overeen met de eerdergenoemde CMR. Echter, anders dan bij CMR, geldt onze methode ook voor $\sigma = 1$.

Voor het geval $\sigma = 1$, dus als de enige positieve eigenwaarde van dezelfde orde grootte is als oneindig veel negatieve, kleine eigenwaarden, is de analyse veel interessanter. We krijgen een oneindig-dimensionaal stelsel van ODEs, afkomstig van de V -vergelijking, samen met een enkele vergelijking voor de grootste eigenfunctie behorend bij λ_1 . Deze zijn wederom gekoppeld aan algebraïsche vergelijkingen, vergelijkbaar met [15].

Een formele analyse van dit oneindig-dimensionale stelsel geeft een eerste-orde benadering van een niet-triviaal vast punt. Wanneer de lineaire stabiliteit van dit vaste punt bestudeerd wordt blijken slechts twee eigenmodi niet exponentieel uit te doven. Doordat onze methode leidt tot concrete uitdrukkingen, is daarmee een expliciete waarde voor de bifurcatieparameter r te bepalen, waarvoor het vaste punt een Hopf-bifurcatie ondergaat. Rijkere gedrag dan een Hopf-bifurcatie, samen met geassocieerde periodieke oplossingen, lijkt uitgesloten. Dit verandert wanneer we de methode toe-



Figuur 2 DCM-data van het diep-waterstation ALOHA (Hawaii) die oscillaties vertonen. (a): Chlorofyl concentratie. (b): Nitraat and nitriet concentraties (gecombineerd). Figuur afkomstig uit [9].



Figuur 3 Boven: stroomlijnen, die de circulatie tonen. Midden: ruimtelijk patroon van zwevend stof; waarden geschaald met maximum concentratie. Onder: geschaalde grootste eigenwaarde versus afstand. De verticale lijn geeft de locatie van het troebelheidsmaximum aan.

passen op (5), waarbij de niet-lineaire term in de V -vergelijking, $\varepsilon G(U, V)$, geen voorfactor ε heeft. Uit numerieke simulaties blijkt dat zulke systemen chaos kunnen bevatten.

Plankton-groei in kustzeeën en estuaria

Behalve in de open oceaan worden ook in kustzeeën en estuaria, overgangsgedebieden van zout water naar zoet rivierwater, geregeld hoge concentraties fytoplankton waargenomen. Het bloeiseizoen is de vroege lente, wanneer de lichtintensiteit en nutriëntconcentratie sterk toenemen. Maar ook in andere situaties, zoals tijdens en na een grote rivierafvoer, kan zich bloei voordoen. Het voorspellen van bloei is een belangrijke taak voor beheerders van kustwateren, vanwege de ecologische gevolgen, sommige planktonsoorten zijn bijvoorbeeld giftig voor de mens.

De dynamica verschilt wezenlijk van die in de oceaan: dieptes zijn kleiner en variëren sterk (zand- en modderbanken, geulen, eb en vloed, stormvloed). Ook is turbulente menging meer aanwezig (getijden, golfbreking) en zoetwateraanvoer zorgt voor horizontale gradiënten in nutriëntconcentraties. Bovendien wordt licht sneller geabsorbeerd door hoge concentraties zwevend stof. Kenmerkend in estuaria is dat vanaf zee de troebelheid eerst

toeneemt en vervolgens weer afneemt. De eb- en vloedcyclus en stormen zorgen voor fluctuaties van troebelheid in de tijd.

De structuur van de eerste, eendimensionale wiskundige modellen is identiek aan die van de hiervoor besproken oceaanmodellen, maar parameterwaarden zijn anders. Zo beschouwt Cloern [2] een model voor voedselrijke estuaria met fytoplanktonconcentratie als enige variabele. Variaties in de verticale menging lijken de oorzaak te zijn van waargenomen periodieke fluctuaties in chlorofylconcentraties: tijdens doortijd is er weinig menging, waardoor er minder fytoplankton naar de bodem wordt getransporteerd, waar zich zoöplankton bevindt. Het omgekeerde doet zich voor tijdens springtij. In [11] worden ook ruimtelijke variaties beschouwd, door een numeriek model met twee door laterale diffusie gekoppelde kolommen. PDEs bepalen de planktonconcentratie en zwevend-stofconcentratie. De omwoeling van sediment en verticale diffusiecoëfficiënten worden berekend met een door getij en wind aangedreven hydrodynamisch model. Belangrijke conclusies zijn dat de timing van wind en getij sterk de dynamiek beïnvloedt. Zo zorgt een langdurige harde wind voor grotere lichtabsorptie en uitsterfing van plankton door omwoeling van sediment.

Onze studie [13] richt zich daarentegen op ruimtelijke variaties van fytoplankton in de stroomrichting van een estuarium. Laatstgenoemde wordt gemodelleerd als een kanaal, met constante diepte en breedte, dat wordt aangedreven door een zoetwaterafvoer. Het model beschrijft netto stroming en zwevend-stofconcentratie (aangedreven door omwoeling en advectie), alsook planktonconcentratie en nutriëntconcentratie. Middels een tweedimensionale eigenwaardeanalyse wordt de groei van fytoplankton lineair geanalyseerd. Figuur 3 toont de netto stroming, zwevend-stofconcentratie en grootste eigenwaarde tegen de afstand tot de zeewaartse rand. In deze berekening is nutriënt horizontaal uniform. In overeenstemming met de waarnemingen heeft de groei van plankton een minimum op de locatie van het troebelheidsmaximum. In het gebied meer zeewaarts laten observaties bovendien een chlorofylmaximum zien. Dit gedrag wordt verklaard door de afname van nutriënt richting zee die groei sterk beperkt.

Grootschalige modellen

Vanzelfsprekend zijn er ook realistische modellen ontwikkeld die als doel hebben een grootschalig ecosysteem nauwkeurig te beschrijven — iets dat onmogelijk is met simpele NPZ-modellen. Een van de belangrijkste vragen die men uiteindelijk beoogt te beantwoorden is natuurlijk die van de globale invloed van fytoplankton op het klimaat (en tegelijkertijd andersom). Iets minder ambitieus is de vraag hoe groot het totaal door fytoplankton veroorzaakte kooldioxide-transport is. In lokale situaties wil men bijvoorbeeld juist expliciet inzicht krijgen in de invloed van menselijke activiteiten, zoals de toevoer van kunstmatige nutriënten via industriële afvalwateren aan ecosystemen. Dit afval kan giftige algenbloei of hypoxie — een tekort aan zuurstof in het water — veroorzaken. Beide effecten kunnen dramatische gevolgen hebben voor het lokale onderwaterleven.

Grootschalige modellen zijn complexe rekenmodellen die ongeschikt zijn voor een directe wiskundige analyse. Belangrijke factoren als de biochemische en hydrodynamische processen worden in groot detail beschreven. Tegelijkertijd worden in deze modellen tijdsafhankelijke variaties (lichtinval, watertemperatuur) meegenomen, om de getijden, dag-en-nachtcycli en seizoenswisselingen na te bootsen. Ook worden vaak planktonpopulaties in meerdere componenten met verschillende eigenschappen opgedeeld om daarvan het effect te onderzoeken.



Foto: USGS/NOAA/Landsat 7

Figuur 4 Fytoplanktonbloei bij Gotland, Zweden (2005)

Het is niet de bedoeling van dit verhaal om in detail in te gaan op de ‘state of the art’ op dit gebied. Succesvolle modellen zijn vaak nog duidelijk toegespitst op een lokale situatie, bijvoorbeeld om de grootschalige en schadelijke algenbloei (*Harmful Algal Blooms*, HAB) voor de kust van Oregon te begrijpen. Andere modellen richten zich juist op het beschrijven van het gedrag van de fytoplanktonsoort *Phaeocystis*, een belangrijke soort omdat deze een cruciale rol speelt bij kooldioxidetransport. In zo’n op één fytoplanktonsoort toegespitst model wordt de cyclus van zuurstof en de invloed van meerdere nutriënten op de evolutie en dynamica van fytoplankton op een zo realistisch mogelijke manier beschreven. Andere modellen concentreren zich juist weer op het effect dat het ‘trace element’ ijzer op fytoplanktonpopulaties heeft, en dan vaak ook nog eens in een specifiek gebied in de zuidelijke oceanen. Dit is met name interessant omdat gebrek aan ijzer een directe negatieve invloed heeft op fytoplanktongroei.

Hoewel dit soort rekenmodellen steeds beter worden in het reproduceren van waarnemingen en het doen van voorspellingen, is ook in de context van deze modellen de belangrijkste conclusie dat de complexe dynamica van fytoplankton op veel vlakken nog onvoldoende begrepen is.

Referenties

- 1 E. Benincà et al., Chaos in a long-term experiment with a plankton community, *Nature* 451 (2008), 822–826.
- 2 J.E. Cloern, Tidal stirring and phytoplankton bloom dynamics in an estuary, *J. Mar. Res.* 49 (1991), 203–221.
- 3 A.M. Edwards en J. Brindley, Zooplankton mortality and the dynamical behaviour of plankton population models, *Bull. Math. Biol.* 61 (1999), 303–339.
- 4 R.H. Fleming, The control of diatom populations by grazing, *J. Conseil Intl. Explor. Mer* 14 (1939), 210–227.
- 5 P.J.S. Franks, NPZ models of plankton dynamics: their construction, coupling to physics, and application, *J. Oceanogr.* 58 (2002), 379–387.
- 6 P.J.S. Franks, Planktonic ecosystem models: perplexing parameterizations and a failure to fail, *J. Plankton Res.* 31 (2009), 1299–1306.
- 7 A. Hastings en T. Powell, Chaos in a three-species food chain, *Ecology* 72 (1991), 896–903.
- 8 S.-B. Hsu en Y. Lou, Single phytoplankton species growth with light and advection in a water column, *SIAM J. Appl. Math.* 70 (2010), 2942–2974.
- 9 J. Huisman, N.N. Pham Thi, D.M. Karl en B.P. Sommeijer, Reduced mixing generates oscillations and chaos in the oceanic deep chlorophyll maximum, *Nature* 439 (2006), 322–325.
- 10 H. Ishii en I. Takagi, Global stability of stationary solutions to a nonlinear diffusion equation in phytoplankton dynamics, *J. Math. Biol.* 16 (1982), 1–24.
- 11 C.L. May et al., Effects of spatial and temporal variability of turbidity on phytoplankton blooms, *Mar. Ecol. Prog. Ser.* 254 (2003), 111–128.
- 12 G.A. Riley, H. Stommel en D.F. Bumpus, Quantitative ecology of the plankton of the western North Atlantic, *Bull. Bingham Oceanogr. Coll.* XII(3) (1949), 1–169.
- 13 H.E. de Swart, H.M. Schuttelaars en S.A. Talke, Phytoplankton growth in turbid estuaries, a simple model, *Cont. Shelf Res.* 29 (2009), 136–147.
- 14 A. Zagaris, A. Doelman, N.N. Pham Thi en B.P. Sommeijer, Blooming in a nonlocal, coupled phytoplankton nutrient model, *SIAM J. Appl. Math.* 69 (2009), 1174–1204.
- 15 A. Zagaris en A. Doelman, Emergence of steady and oscillatory localized structures in a phytoplankton nutrient model, *Nonlinearity* 24 (2011), 3437–3486.
- 16 A. Doelman en A. Zagaris, Complex phytoplankton dynamics: the mathematical perspective (2012), www.mathclimate.org/education/annotated-reading-lists.

Anna von der Heydt

IMAU

Utrecht University

a.s.vonderheydt@uu.nl

Henk Dijkstra

IMAU

Utrecht University

h.a.dijkstra@uu.nl

El Niño: processes, dynamics and predictability

The El Niño variability in the equatorial Tropical Pacific is characterized by sea-surface temperature anomalies and associated changes in the atmospheric circulation. Through an enormous monitoring effort over the last decades, the relevant time scales and spatial patterns are fairly well-documented. Anna von der Heydt and Henk Dijkstra describe the hierarchy of models that have been developed to understand the physics of this phenomenon and to make predictions of future variability. It turns out that the predictability of El Niño events is still limited to about 6–9 months due to inherent nonlinear processes.

About once every four years, the sea surface temperature in the equatorial eastern Pacific is a few degrees higher than normal [13]. Near the South American coast, this warming of the ocean water is usually at its maximum around Christmas. Long ago, Peruvian fishermen called it El Niño, the Spanish phrase for the 'Christ Child'. During the last decades, the equatorial Pacific has been observed in unprecedented detail [10]. The El Niño relevant quantities in the equatorial ocean and atmosphere system are sea-level pressure, sea-surface temperature, sea-level height, surface wind and ocean sub-surface temperature.

What makes El Niño unique among other interesting phenomena of natural climate variability is that it has both a well-defined pattern in space and a relatively well-defined time scale. The sea-surface temperature anomaly during the maximum of the El Niño 1997–1998, in December 1997, obtained by subtracting a long term mean (see Figure 1a) from the actual pattern, is shown in Figure 1b. The anomaly is characterized by higher than usual equatorial Pacific temperatures east of the date line [14]. A scalar index of this sea-surface temperature anomaly pattern is the NINO3 index, defined as the sea-surface tem-

perature anomaly averaged over the region 5°S – 5°N , 150°W – 90°W (see the black box in Figure 1a). In the time series (Figure 1c), the high NINO3 periods are known as El Niño's

and the low NINO3 periods as La Niña's. Most El Niño's and La Niña's peak around December and are hence phase-locked to the seasonal cycle.

The mean surface winds over the equatorial Pacific, the trade winds, blow from east to west and are driven by an area of average high pressure in the eastern part of the Pacific and a low pressure area over Indonesia. The Southern Oscillation consists of an irregular strengthening and weakening of the trade winds, related to the changes in surface pres-

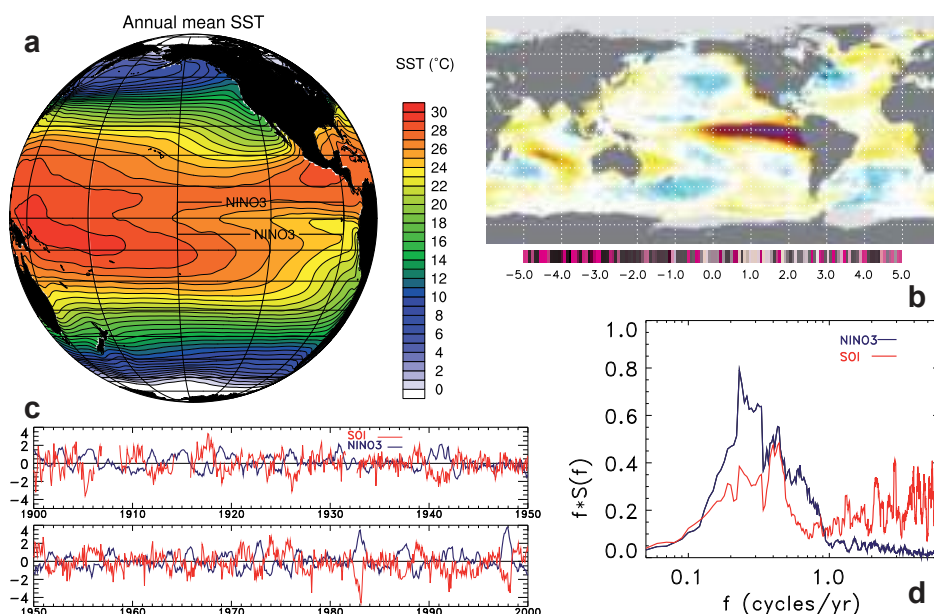


Figure 1 (a) The time mean sea-surface temperature of the Pacific Ocean. The black box indicates the NINO3 region. (b) Sea-surface temperature anomaly field of December 1997, at the height of the 1997/1998 El Niño. (c) Monthly mean SOI and NINO3 index over the period 1900–1999. (d) Spectra of monthly mean SOI and NINO3. Shown are normalized periodograms smoothed over 11 bins, that is over 0.11 cycles per year. Note that the x-axis is logarithmic, and $f \times S(f)$ rather than $S(f)$ is shown, in order that equal areas make equal contributions to the variance.

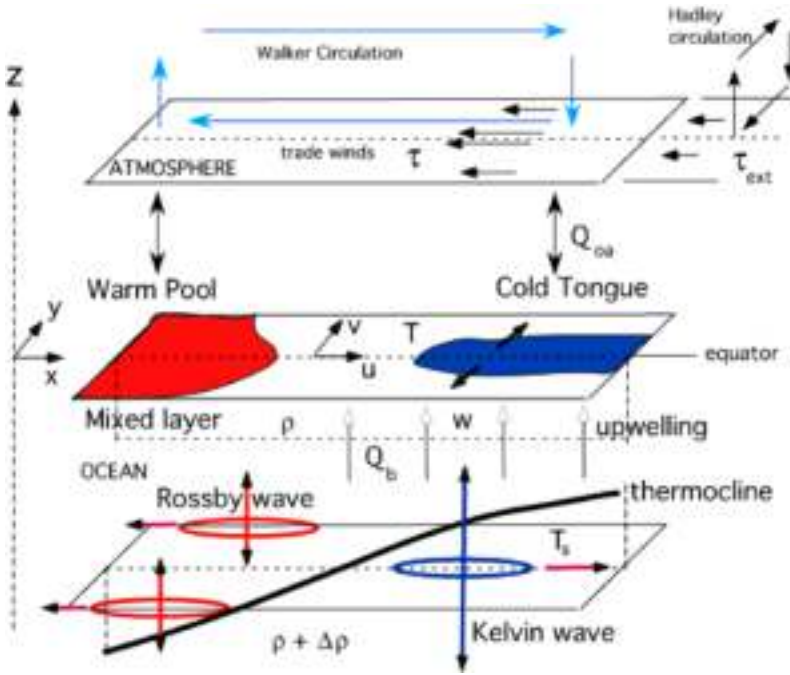


Figure 2 Overview of the oceanic and atmospheric processes of the equatorial coupled ocean-atmosphere system.

sure. An index that captures the amplitude of the sea-level pressure anomaly pattern is the Southern Oscillation Index (SOI), which is the normalized difference of the pressure anomalies between Tahiti (18°S, 150°W) and Darwin (12°S, 131°E).

It was Jacob Bjerknes [1] who discovered that El Niño and the Southern Oscillation are part of the same phenomenon now referred to as ENSO. Generally, periods with a high NINO3 index have a low SOI and vice versa (Figure 1c) which indicates that the trade winds are weaker (stronger) during El Niño's (La Niña's). The average time scale of ENSO is about four years, as can be seen from the spectra of the NINO3 index and the SOI in Figure 1d. The SOI contains more high-frequency variability than the NINO3 index.

Basic processes

The vertical structure of the upper ocean can be considered as layered with a well-mixed upper layer down to a depth of 50 m and a sharp transition between warm and cold water (the thermocline) at about 150 m depth (Figure 2). The temperature in the mixed layer changes due to air-sea interaction, exchange processes with deeper layers and advection. Let the net heat flux from the atmosphere into the ocean and that at the bottom of the mixed layer be denoted by Q_{oa} and Q_b (positive when heat is going into the mixed layer), respectively. Under the approximation that the temperature is vertically homogeneous over

the layer, one can integrate the heat balance equation over the layer which results in an equation for the mixed-layer temperature T , i.e.

$$\frac{\partial T}{\partial t} = -(u \frac{\partial T}{\partial x} + v \frac{\partial T}{\partial y}) + F_T^H + \frac{Q_{oa} + Q_b}{\rho C_p H_m}, \quad (1)$$

where C_p is the heat capacity of water, H_m the mean depth of the mixed layer, ρ its mean density and where (u, v) are the horizontal velocities in the mixed layer. The quantity F_T^H represents the horizontal mixing of heat through lateral turbulent exchange processes. Equation (1) expresses that the mixed layer temperature can change due to advection (first two terms in the right hand side), horizontal exchange (third term) and vertical exchange (last term) of heat.

The effects of the wind stress on the upper ocean are threefold. The wind stress due to the trade winds causes water to pile up near the western part of the basin. This induces a higher pressure in the upper layer western Pacific than that in the eastern Pacific and consequently a shallowing of the thermocline towards the east (Figure 2). Second, these winds cause a divergence of mass at the equator that is compensated by vertical mass transport, the so-called upwelling. Finally, the wind stress is responsible for the presence of the upper ocean currents.

When the amplitude and/or the direction of the wind stress changes, the upper ocean adjusts through wave dynamics. The most

important waves involved in this adjustment process are:

1. Equatorial Kelvin waves. For such a wave, the meridional structure of the thermocline is maximal at the equator. Its amplitude decays exponentially in meridional direction with a decay scale λ_0 of about 300 km. The zonal velocity has the same spatial structure as the thermocline and the meridional velocity is zero. The group velocity of these non-dispersive waves, say c_o , is about 2 ms^{-1} . It takes about $\tau_K = 3$ months for the wave to cross the Pacific basin of width $L = 15,000 \text{ km}$.
2. Long equatorial Rossby waves. For such a wave, the meridional thermocline structure has an off-equatorial maximum. For each of these waves, the meridional velocity is much smaller than the zonal velocity. The group velocity of these non-dispersive waves is $c_j = -c_o/(2j + 1)$. The $j = 1$ long Rossby wave travels westward with a velocity which is $1/3$ of that of the Kelvin wave and hence takes about $\tau_R = 9$ months to cross the Pacific basin.

When a Kelvin wave meets the east coast, the dominant contribution of the reflected signal are long Rossby waves. Similarly, when a Rossby wave meets the west coast, it is reflected dominantly as a Kelvin wave.

The coupling of the ocean-atmosphere system can be understood by looking at the effect of sea-surface temperature anomalies. A sea-surface temperature anomaly will give through local heating (as measured by a parameter α_T) a lower-level wind stress anomaly (as measured by a parameter y). The resulting wind stress anomaly on the ocean-atmosphere surface will: (i) change the thermocline slope through horizontal pressure differences in the upper ocean; (ii) change the strength of the upper ocean upwelling; and (iii) affect the upper ocean currents (u, v) in the mixed layer. The changes in velocity field and thermocline field will affect the sea surface temperature, according to (1).

The strength of the coupling in the system, i.e. between the atmosphere and the ocean, is determined by the combined effects of how much zonal wind anomaly is generated by sea-surface temperature anomalies and how much of the momentum of this wind is transferred as stress to the upper ocean layer. The strength of the coupling is measured by the parameter μ which is a (dimensionless) product [3] of α_T and y , i.e.

$$\mu = \frac{y \alpha_T \Delta T L^2}{c_o^2 c_a^2}, \quad (2)$$

where the quantity ΔT is the mean zonal temperature difference over the basin, L the zonal length of the basin and $c_a = 60 \text{ m s}^{-1}$ the free Kelvin wave speed in the atmosphere.

Physics of El Niño variability

Much of the theory on El Niño variability was developed with the help of the Zebiak–Cane (ZC) model [19], which precisely represents the physical processes in Figure 2. In this model the zonal wind stress τ^x over the ocean is represented by

$$\tau^x = \tau_{ext}^x + \mu A(T), \quad (3)$$

where $A(T)$ represents the atmospheric response of the model to the sea-surface temperature T . The part τ_{ext}^x is determined externally, i.e. by processes outside the Pacific basin and it is usually taken to be constant.

At zero coupling strength ($\mu = 0$), the ocean circulation and consequently the sea-surface temperature is determined by the external zonal wind stress τ_{ext}^x . A small easterly wind causes a small amount of upwelling and a small slope in the thermocline. At small μ , additional easterly wind stress due to coupling occurs because of the zonal temperature gradient. This leads to larger upwelling and a larger thermocline slope, strengthening the cold tongue in the eastern part of the basin.

With increasing μ , the Pacific climate state becomes unstable. Necessary conditions for instability can be obtained by determining the linear stability boundary through normal mode analysis. In such an analysis, an arbitrary perturbation is decomposed in modes (e.g. Fourier modes) and the growth or decay of each of these modes is investigated. When the background state is stationary, the time-

dependence of each mode is of the form $e^{\sigma t}$, where $\sigma = \sigma_r + i\sigma_i$ is the complex growth rate. If μ is the control parameter, then the linear stability boundary μ_c is the first value of μ where $\sigma_r = 0$ for one particular normal mode. A mode with $\sigma_i \neq 0$ is oscillatory (with a period $2\pi/\sigma_i$) while a mode with $\sigma_i = 0$ is called stationary. If the growth factor of a mode is positive, sustained oscillations will occur and if it is negative the mode will be damped. In the presence of atmospheric noise, however, the spatial pattern of a damped mode can still be excited. [4] provides an overview of the dependence of the growth factor (Figure 3a) and period (Figure 3b) of the least stable mode (the ENSO mode) on the background conditions. The dashed curves in Figure 3 represent zero growth (so-called neutral conditions) of the ENSO mode. Present-day estimates for the Pacific background state correspond to the area near the points A and B, with a period of 2–6 years and near neutral conditions.

The main positive feedbacks responsible for the instability of the coupled ocean-atmosphere system are referred to the thermocline, upwelling and zonal advection feedback [12]. The thermocline feedback can be described as follows. A positive sea-surface temperature anomaly in the eastern Pacific will lead to a weakening of the trade winds to the west of this anomaly. This will lead to a smaller slope of the thermocline and hence the cold water will be further away from the surface in the east. This causes a positive heat flux Q_b in (1) which will amplify the initial disturbance. The upwelling feedback can be understood in a similar way, considering again a positive sea-surface temperature anomaly in the eastern part of the basin and associated changes in the wind. In this

case, weaker easterly winds induce less upwelling and hence less cold water enters the mixed layer, again amplifying the sea surface temperature perturbation. The zonal advection feedback arises through zonal advection of heat, particularly in regions of strong sea surface temperature gradients. A warm sea-surface temperature anomaly then can lead to wind anomalies, and, if positive, intensify zonal ocean currents, which bring more warm water to the location of the original perturbation.

In the ZC model, oscillatory El Niño behaviour is found slightly above critical conditions. Attempts have been made to derive conceptual models to explain the physics of the oscillation period. Most of these share the ideas that one of the positive feedbacks acts to amplify sea-surface temperature anomalies and that wave adjustment processes in the ocean eventually cause a delayed negative feedback. These common elements are grouped together in the delayed oscillator mechanism. The differences between the views are subtle [6] and are related to the role of the boundary wave reflections, the dominant feedback which is responsible for amplification of anomalies and the view of the dynamical adjustment processes in the ocean. There are four types of ENSO ‘oscillators’: (i) the ‘classical’ delayed oscillator; (ii) the recharge/discharge oscillator; (iii) the western Pacific oscillator; and (iv) the advective/reflective oscillator. Below, we only describe (i), the other oscillators are described extensively in [6] and [18].

In the ‘classical’ delayed oscillator view, the eastern boundary reflection is unimportant, the thermocline feedback is dominant and individual Kelvin and Rossby waves control ocean adjustment. A minimal model [15]

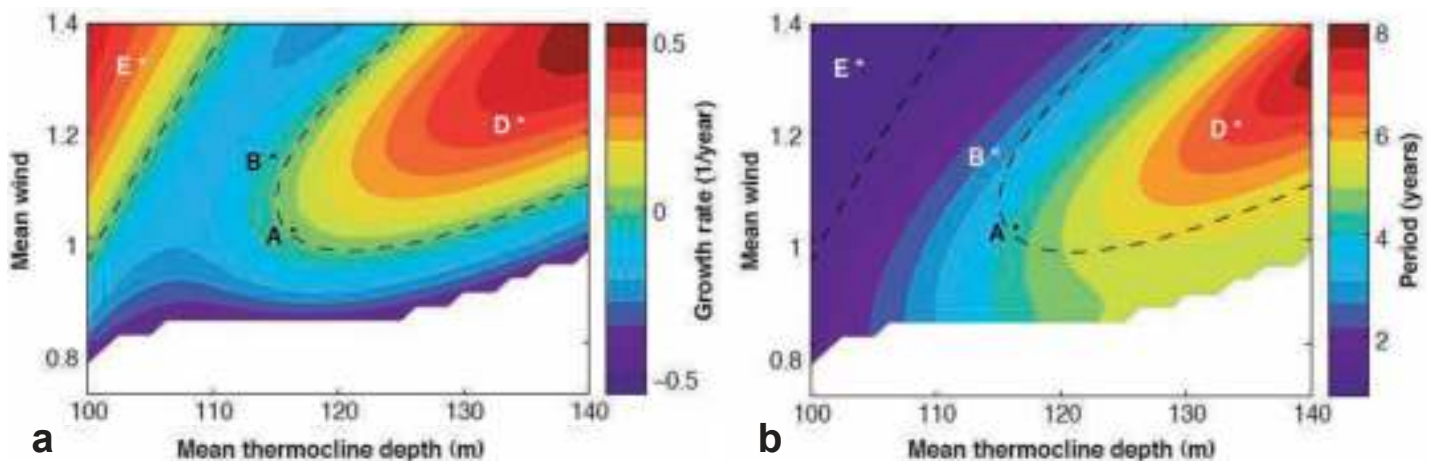


Figure 3 Dependence of (a) growth rate (σ_r in yr^{-1}) and (b) period ($2\pi/\sigma_i$ in yr) of the ENSO mode on the mean wind intensity (in units of $0.5 \text{ cm}^2 \text{ s}^{-2}$) and the mean thermocline depth (in m) [4].

representing this behaviour is a differential delay equation of the form

$$\frac{dT(t)}{dt} = ah_e(x_c, t - \tau_1) - bh_o(x_c, t - \tau_2) - cT^3(t). \quad (4)$$

In this equation, a , b and c are constants, T is the eastern Pacific temperature anomaly, which is influenced by midbasin (at $x = x_c$) equatorial thermocline depth anomalies h_e and off-equatorial anomalies h_o . Furthermore, $\tau_1 = \frac{\tau_K}{2}$ and $\tau_2 = \frac{\tau_R}{2} + \tau_K$ where τ_K and τ_R are the basin crossing times of the Kelvin wave and the gravest ($j = 1$) Rossby wave.

When a Kelvin wave, which deepens the thermocline, arrives in the eastern Pacific, local amplification of temperature perturbations occurs through the thermocline feedback, represented by the first term in the right hand side of (4). The wind response excites (off-equatorial) Rossby waves which travel westwards, reflect at the western boundary and return as a Kelvin wave which rises the thermocline and provides a delayed negative feedback (second term in (4)). The delay τ_2 is the time taken for the Rossby wave to travel from the centre of wind response to the western boundary plus the time it takes the reflected Kelvin wave to cross the basin. The nonlinear in (4) term is needed for equilibration to finite amplitude.

A linear mechanism for ENSO's phase locking to the seasonal cycle was proposed in [5]. The coupling strength μ varies on a seasonal time scale and is maximal during June–July and minimal during December–January. The variation of the eastern Pacific thermocline depth h in the 'classical' delayed oscillator picture can be described by

$$\frac{dh(t)}{dt} = dF(\mu(t - \tau_1)h(x_c, t - \tau_1)) - eF(\mu(t - \tau_2)h(x_c, t - \tau_2)) - fh(t). \quad (5)$$

Here F is a certain nonlinear function, the damping of thermocline anomalies is assumed to be linear (the third term in the right hand side of (5)) and d, e and f are again constants. At peak time of ENSO, the eastern thermocline anomalies are maximal and hence $dh/dt \approx 0$. For small dissipation, the warming by the Kelvin wave (first term in (5)) has to balance the cooling due to the Rossby wave (second term in (5)).

Suppose that the peak of ENSO would occur during June, i.e. with maximum μ . In this case, the warm Kelvin wave was excited during April and because coupling is strong over April–June, it has a relatively large amplitude.

The Rossby wave was, however, excited about 6 months earlier, i.e. during November–December when coupling was small. It has therefore a small amplitude and cannot balance that of the Kelvin wave ($\mu(t - \tau_1)h(x_c, t - \tau_1)$ is large while $\mu(t - \tau_2)h(x_c, t - \tau_2)$ is small). On the other hand, when peak conditions are in December, the Kelvin wave is excited in October during relatively small coupling and its amplitude is only weakly amplified. The Rossby wave is now excited in June when coupling is large and is amplified to significant amplitude to balance the Kelvin wave thermocline signal.

ENSO's predictability barrier

Two different mechanisms have been suggested to explain the irregularity of ENSO, as for example seen in the NINO3 time series (Figure 1c). These are: (i) deterministic chaotic behaviour; and (ii) impact of small-scale processes usually referred to as 'noise', possibly through non-normal growth behaviour.

Deterministic chaotic behaviour can result from nonlinear resonances due to interaction of the seasonal cycle and the ENSO mode [7, 16], leading to frequency locking. A typical example of a one-dimensional map where frequency locking occurs is the circle map given by

$$x_{n+1} = x_n + \Omega - \frac{K}{2\pi} \sin 2\pi x_n \pmod{1} \quad (6)$$

on the interval $[0, 1)$. Here, mod 1 indicates

that if x_n becomes larger than 1 or smaller than 0, its value is adjusted by adding or subtracting 1 to get a value within the unit interval. $\Omega \in [0, 1)$ can be considered to be the forcing frequency and the control parameter K is a measure of the nonlinearity of the system. To look for frequency locked regions, one varies Ω for fixed K and searches for rational rotation numbers W defined by

$$W = \lim_{n \rightarrow \infty} \frac{x_n - x_0}{n}, \quad (7)$$

where in this definition the value of x_n is considered, without the mod 1 function.

When the coupling strength μ is increased above criticality in the seasonally forced ZC model the NINO3 index displays regimes of frequency locking to the annual cycle alternated by chaotic regimes [8]. In these frequency-locked regimes, the ENSO period is a rational multiple of the seasonal cycle. When δ_s indicates the parameter which measures the strength of the upwelling feedback, the behaviour of the ZC model in the (μ, δ_s) parameter plane is captured [8] in the so-called Devil's terrace (Figure 4). The colours indicate the frequency ratio of frequency locked solutions and the grey areas indicate the chaotic solutions. Although chaotic solutions are nearly everywhere in parameter space over a fractal surface, the locked regime is broad, with the frequency ratio becoming smaller for smaller δ_s . For many solutions found, the phase lock-

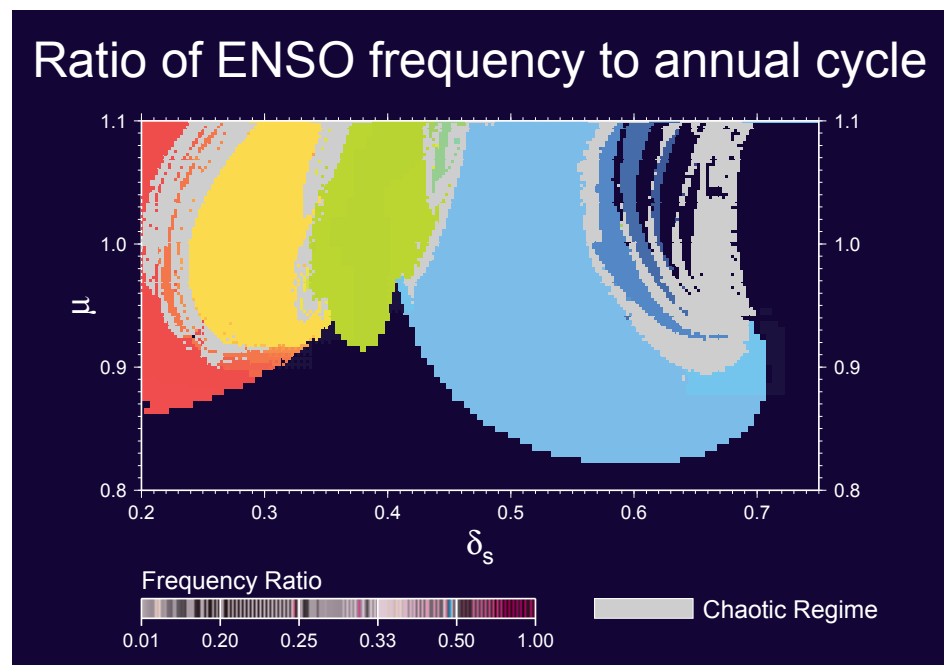


Figure 4 The Devil's terrace as computed in [8]. Colours indicate the frequency locked regimes, with chaotic regimes in between. Along a section for fixed μ , a so-called Devil's staircase appears.

ing to the annual cycle is near to that observed with January being the preferred month of the peak of the warm event.

From the perspective of El Niño, processes that evolve independently with smaller time and space scales can be considered as noise. In the atmosphere, for example, the 30–60 day or Madden–Julian oscillation [9] gives rise to westerly wind bursts. These are events of anomalous westerlies lasting typically a week, which are strongest somewhat west of the region of the El Niño wind response. Kelvin waves are excited very effectively by the Madden–Julian oscillation because they travel eastward with a similar speed. The spectra of the NINO3 index and the SOI in Figure 1 are moderately peaked compared to those of many deterministic versions of the ZC model. Hence, it is generally agreed that noise contributes substantially to the irregularity of El Niño. If the Tropical Pacific climate system is close to neutral conditions, non-normal growth of perturbations, which occurs because the operator in the ZC model is not self-adjoint, can lead to additional irregular behaviour [11].

During boreal spring the coupled ocean-atmosphere system is thought to be at its frailest state. Then the system is most susceptible to perturbations, which can be in-

ternal small-scale stochastic events (random wind bursts) or random external influences (like the monsoon). This leads to a spring predictability barrier in April/May independent of when the forecast is started. The existence of this barrier (still) limits the predictability of El Niño (with a broad range of models) to about 6–9 months.

Another issue is to predict how the characteristics of ENSO variability, such as amplitude or frequency change under global warming. In this case, the pattern of the long-term mean climate in the tropical Pacific becomes a crucial factor, because this pattern partly determines the strength of the different feedbacks involved. It is still not clear whether the ENSO system is above criticality or just below criticality (and hence excited by noise). Future changes in the global climate as a whole can therefore have a significant impact on the characteristics of ENSO variability. When it comes to predictions of long-term changes in ENSO characteristics under global warming, state-of-the-art climate models project a wide range of changes to future ENSO variability [2], indicating that the interaction of the long-term mean state of climate and modes of inter-annual variability remains poorly understood. A study with a ZC type model has indicated that changes in the back-

ground temperature and external wind stress mainly shift the position of the cold tongue in the eastern equatorial Pacific [17]. Such a shift affects local temperature gradients and can impact the relative strength of the different feedbacks, e.g. strengthen the zonal advection feedback. From the perspective of Figure 3, a change in the background climate could move the ENSO system across the critical stability threshold (dashed curve in Figure 3).

Summary

Although there are still many issues which are not sufficiently well explored, we can conclude that the ENSO phenomenon can be understood within a (weakly) nonlinear framework of equatorial ocean-atmosphere interaction. The oscillatory nature of ENSO derives from an internal mode of variability of the coupled system involving coupled feedbacks and (ocean) wave dynamics. Possible mechanisms for ENSO irregularity have been suggested, but it is not clear at the moment which one is dominant. Noise and the inherent instability of the system are important limiting factors for the predictability of El Niño. The growth of perturbations depends on the season and leads to a spring predictability barrier.

References

- 1 J.P. Bjerknes, Atmospheric teleconnections from the equatorial Pacific, *Monthly Weather Review* 97 (1969), 163–172.
- 2 M. Collins, The El Niño - Southern Oscillation in the second Hadley Centre Coupled model and its response to greenhouse warming, *Clim. Dyn.* 24 (2005), 89–104.
- 3 H.A. Dijkstra, *Nonlinear Physical Oceanography: A Dynamical Systems Approach to the Large Scale Ocean Circulation and El Niño*, Springer, New York, 2005.
- 4 A. Fedorov and S. Philander, Is El Niño Changing?, *Science* 288 (2000), 1997–2002.
- 5 E. Galanti and E. Tziperman, ENSO's phase locking to the seasonal cycle in the fast-SST, fast-wave and mixed-mode regimes, *J. Atmos. Sci.* 57 (2000), 2936–2950.
- 6 F.-F. Jin, An equatorial recharge paradigm for ENSO. II: A stripped-down coupled model, *J. Atmos. Sci.* 54 (1997), 830–8847.
- 7 F.-F. Jin, J.D. Neelin and M. Ghil, El Niño on the devil's staircase: Annual subharmonic steps to chaos, *Science* 264 (1994), 70–72.
- 8 F.-F. Jin, J.D. Neelin and M. Ghil, El Niño/Southern Oscillation and the annual cycle: Subharmonic frequency-locking and aperiodicity, *Physica D* 98 (1996), 442–465.
- 9 R. Madden and P. Julian, Observations of the 40–50-day tropical oscillation — A review, *Monthly Weather Review* 122 (1994), 814–835.
- 10 M. McPhaden and co-authors, The Tropical Ocean-Global Atmosphere observing system: a decade of progress, *J. Geophys. Res.* 103 (1998), 14,169–14,240.
- 11 A.M. Moore and R. Kleeman, Stochastic Forcing of ENSO by the Intraseasonal Oscillation, *J. Climate* 12 (1999), 1199–1220.
- 12 J.D. Neelin, The slow sea surface temperature mode and the fast-wave limit: Analytic theory for tropical interannual oscillations and experiments in a hybrid coupled model, *J. Atmos. Sci.* 48 (1991), 584–606.
- 13 S.G.H. Philander, *El Niño and the Southern Oscillation* Academic Press, New York, 1990.
- 14 E. Rasmusson and T.H. Carpenter, Variations in tropical sea surface temperature and surface wind fields associated with the Southern Oscillation/El Niño, *Monthly Weather Review* 110 (1982), 354–384.
- 15 M. Suarez and P.S. Schopf, A delayed action oscillator for ENSO, *J. Atmos. Sci.* 45 (1988), 3283–3287.
- 16 E. Tziperman, L. Stone, M.A. Cane and H. Jarosh, El Niño chaos: overlapping of resonances between the seasonal cycle and the Pacific ocean-atmosphere oscillator, *Science* 264 (1994), 72–74.
- 17 A.S. von der Heydt, A. Nnafie and H.A. Dijkstra, Cold tongue/Warm pool and ENSO dynamics in the Pliocene, *Clim. Past* 7 (2011), 903–915.
- 18 C. Wang and J. Picaut, Understanding ENSO physics: A review, In C. Wang, S.-P. Xie, and J.A. Carton, editors, *Earth's Climate: The Ocean-Atmosphere Interaction*, pages 21–48. American Geophysical Union.
- 19 S.E. Zebiak and M.A. Cane, A model El Niño-Southern Oscillation, *Monthly Weather Review* 115 (1987), 2262–2278.

Arjen Doelman

Mathematisch Instituut
Universiteit Leiden
doelman@math.leidenuniv.nl

Maarten Eppinga

Research Group Environmental Sciences
Universiteit Utrecht
m.b.eppinga@uu.nl

Geertje Hek

Institut International de Lancy
Genève
g.m.hek@uva.nl

Jens Rademacher

Fachbereich 3 – Mathematik
Universität Bremen
jdmr@uni-bremen.de

Max Rietkerk

Research Group Environmental Sciences
Universiteit Utrecht
m.g.rietkerk@uu.nl

Eric Siero

Mathematisch Instituut
Universiteit Leiden
esiero@math.leidenuniv.nl

Koen Siteur

Research Group Environmental Sciences
Universiteit Utrecht
k.siteur@uu.nl

Klimaatverandering, verwoestijning en biljart

Door geleidelijk veranderende omgevingsfactoren kunnen gezonde ecosystemen abrupt veranderen in woestijnen. Om dergelijke catastrofale veranderingen te begrijpen zijn conceptuele modellen geformuleerd voor het gedrag van vegetatiepatronen. Arjen Doelman, Maarten Eppinga, Geertje Hek, Jens Rademacher, Max Rietkerk, Eric Siero en Koen Siteur beschrijven de reactie-diffusievergelijkingen waaruit zulke modellen bestaan en analyseren de patronen die er in voorkomen. Onder toenemende ecologische stress, zoals verminderde regenval, is het mogelijk dat vegetatiepatronen destabiliseren tot alleen een woestijn als stabiele toestand overblijft.

Wereldwijd staan ecosystemen bloot aan veranderingen in bijvoorbeeld klimaat of landgebruik. Droge, maar ‘gezonde’, begroeide ecosystemen kunnen zich lange tijd aanpassen aan geleidelijke omgevingsveranderingen, maar vervolgens plotseling en op een meestal onomkeerbare manier veranderen in woestijn. Vaak wordt zo’n catastrofale verandering voorafgegaan door het ontstaan van vegetatiepatronen, die zowel een gestreept als een labyrintachtig karakter kunnen hebben (zie Figuur 1).

Ecologische modellen en wiskundige vragen

In de ecologie worden conceptuele modellen geformuleerd voor de mechanismen die ten grondslag liggen aan het langzaam ontstaan en abrupt verdwijnen van vegetatiepatronen. Zie Figuur 2 voor een conceptuele interpretatie van het verwoestijningsproces.

Een centrale vraag is of aan een vegetatiepatroon te zien is hoe dicht het systeem in de buurt van de verwoestijningscatastrofe is [16].

Het meest eenvoudige model is dat van Christopher Klausmeier [7]. Het is een reactie-diffusiemodel in twee componenten: water U en biomassa V . Het model beschrijft een systeem op een hellend terrein, met heuvelafwaarts stromend water gemodelleerd door de advectieterm CU_x in (1). Om ook vegetatiepatronen in vlakke gebieden te kunnen beschrijven, beschouwen we hier een generaliseerd Klausmeier–Gray–Scott-model (gKGS) [18]. Wij beperken ons tot patronen die maar van één ruimtelijke variabele x afhangen. In een tweedimensionale setting geeft voortzetting van periodieke patronen zonder verdere variatie in de y -richting streepvormige structuren [7]. Verder kiezen we x onbe-

grensd, $x \in \mathbb{R}$. Dit is een natuurlijke aanname binnen het vakgebied ‘pattern formation’, die aangeeft dat de schaal van de geobserveerde patronen veel kleiner is dan de domeingrootte. In herschaalde vorm ziet het gKGS-model er als volgt uit:

$$\begin{aligned} U_t &= (U^\gamma)_{xx} + CU_x + A(1 - U) - UV^2, \\ V_t &= \delta^2 V_{xx} - BV + UV^2. \end{aligned} \quad (1)$$

De subscripten x en t zijn de partiële afgeleiden naar x en t ; de parameter A representeert de (gemiddelde, jaarlijkse) regenval, B is een sterfteparameter voor de vegetatie en C is een maat voor de helling van het terrein. De niet-lineaire diffusieterm $(U^\gamma)_{xx}$ modelleert de verspreiding van water in vlakke terreinen; zie [4, 18]. De waarde van γ is typisch 2, hoewel de lineaire keuze $\gamma = 1$ ook voorkomt in de vegetatiepatronenliteratuur [6, 18]. De verhouding δ tussen de (typische) diffusiesnelheden van de vegetatie en het water is vanzelfsprekend zeer klein, zodat het model op een natuurlijke manier een *singulier verstoord* karakter heeft. Er zijn meerdere veel realistischere modellen in de literatuur [15, 19], maar begrip van de dynamica van patronen in conceptuele modellen als (1) kan



Figuur 1 Streep- en labyrintvegetatiepatronen in de Sahel in Niger. De typische afstand tussen begroeide stukken terrein is in de orde van 10–100 m.

helpen inzicht te krijgen in wat er in het echt gebeurt. Vooral de grof versimpelde plant–water-interactietermen — de niet-lineariteiten UV^2 — en het ontbreken van onderscheid tussen oppervlakte- en grondwater geven aan dat (1) inderdaad vooral gezien moet worden als fenomenologisch model. Ook andere belangrijke factoren in het verwoestijningsproces, zoals bijvoorbeeld de intensiviteit van de begrazing van de vegetatie, worden niet in dit model meegenomen.

Het gKGS-model reduceert tot het Klausmeier-model [7] in de limiet $C \rightarrow \infty$ [18]. De gekozen schaling geeft aan hoe sterk het Klausmeier-model voor vegetatiepatronen gerelateerd is aan het al veel langer bekende Gray–Scott-model voor autocatalytische chemische reacties, dat ook in de wiskundeliteratuur uitgebreid bestudeerd is: (1) neemt de vorm van het Gray–Scott-model aan voor $\gamma = 1$ en $C = 0$ [2, 5, 11]. Merk op dat vanwege de niet-lineaire diffusieterm, de basisvraag naar existentie en uniciteit van oplossingen van (1) niet triviaal is [8].

In de context van het model correspondeert klimaatverandering met een langzame

verandering van de parameters. Vegetatiepatronen komen overeen met *stabiele*, ruimtelijk periodieke oplossingen in (1). In wiskundige termen komt het proces van verwoestijning dus overeen met het verdwijnen van periodieke patronen ten gevolge van langzaam variërende parameters. Er is een uitgebreide wiskundige literatuur over patronen — en met name het ontstaan ervan — in reactie-diffusievergelijkingen met *vaste* parameters. Hieronder volgt een schets van een aantal basisideeën, uitmondend in de introductie van de zogenaamde *Busse-ballon*. Deze Busse-ballon werd oorspronkelijk geïntroduceerd in de context van (hydrodynamische) convectie [1] en is een in het algemeen tamelijk complex gebied in (parameter, golfgetal)-ruimte waarin stabiele ruimtelijk periodieke oplossingen bestaan. Intuïtief betekent klimaatverandering, ofwel het variëren van parameter A , dat de dynamica van systeem (1) een baan zal beschrijven binnen deze Busse-ballon: verwoestijning komt overeen met het moment dat deze baan de Busse-ballon verlaat en ‘neerslaat’ in de triviale homogene oplossing $(U(x, t), V(x, t)) \equiv (U^0, V^0) = (1, 0)$ van (1), die de kale woestijn voorstelt en een ‘sterke aantrekker’ is voor de dynamica van (1). Zowel het karakter van deze baan als het precieze moment waarop deze baan de Busse-ballon verlaat, geven aanleiding tot diepe, nog onbeantwoorde, wiskundige vragen.

De existentie van patronen

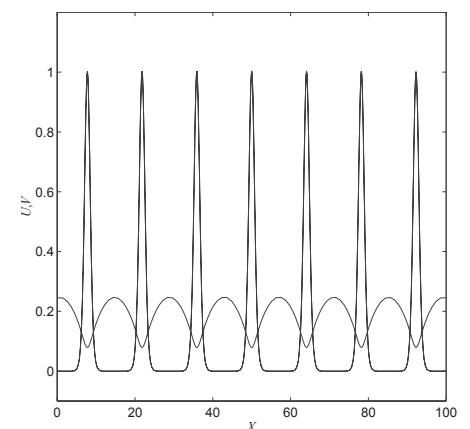
Het begrijpen van de dynamica van (1) begint met het bestuderen van eenvoudige patronen. De allereenvoudigste zijn de homogene evenwichten, tijds- en ruimteonafhankelijke oplossingen van (1) van de vorm

$(U(x, t), V(x, t)) \equiv (U^*, V^*)$. Buiten de genoemde ‘desert state’ (U^0, V^0) zijn er nog twee oplossingen,

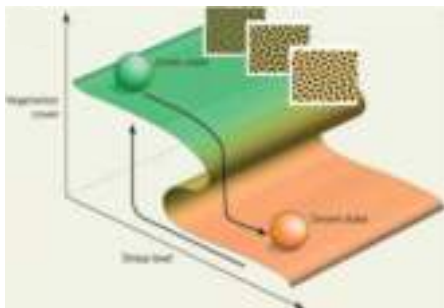
$$(U^\pm, V^\pm) = \left(\frac{A \mp \sqrt{A^2 - 4AB^2}}{2A}, \frac{A \pm \sqrt{A^2 - 4AB^2}}{2B} \right),$$

die twee verschillende homogeen begroeide situaties voorstellen. Voor $A = A_{sn} = 4B^2$ vallen deze samen en verdwijnen in een zogenaamde zadel-knoop-bifurcatie. Voor $A < A_{sn}$, een kleine jaarlijkse regenval, is de woestijn dus de enige homogene oplossing; voor $A > A_{sn}$ is er coëxistentie van drie verschillende homogene evenwichten.

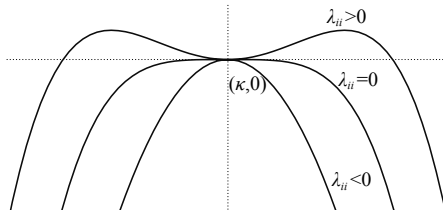
Vegetatiepatronen komen overeen met (lopende) ruimtelijk periodieke oplossingen $(U(x, t), V(x, t)) = (u^p(\xi), v^p(\xi))$, waarbij $\xi = x - ct$ een meebewegend coördinatenstelsel beschrijft (met snelheid c), en (u^p, v^p) een periodieke oplossing is van de *gewone* differentiaalvergelijking (ODE) in ξ die je uit (1) krijgt door daarin $(U, V) = (u(\xi), v(\xi))$ te substitueren. Zie Figuur 3 voor een representatie van $(u^p(\xi), v^p(\xi))$ als functie van x (met $c = 0$). Omdat deze ODE bestaat uit twee gekoppelde tweede-orde-vergelijkingen in ξ — één voor u en één voor v — definieert het een 4-dimensionaal dynamisch systeem; het periodieke patroon $(u^p(\xi), v^p(\xi))$ correspondeert daarom met een gesloten kromme in een 4-dimensionale ruimte. Het aantonen van de existentie van dit soort banen is in het algemeen zeer moeilijk, maar het singuliere karakter van (1) geeft een meetkundige structuur die gebruikt kan worden om het bestaan van ‘vegetatiepatronen’ $(u^p(\xi), v^p(\xi))$ in (1) analytisch aan te tonen [2].



Figuur 3 Een singulier, ruimtelijk periodiek vegetatiepatroon als oplossing van (1) met $\gamma = 1$, $A = 0,02$, $B = 0,14$, $C = 0$, $\delta = 0,1$. Het singuliere karakter van (1) is duidelijk uit de lokaal piekende V -component: tussen twee opeenvolgende vegetatiebanden is er geen of nauwelijks begroeiing.



Figuur 2 Een globale schets van het verwoestijningsproces: de plotselinge catastrofe wordt voorafgegaan door zich langzaam vormende vegetatiepatronen. Zie ook [16].



Figuur 4 De sideband-instabiliteit

Stabiliteit

Dat een zeker patroon bestaat, betekent niet dat het ook waargenomen zal worden. Hiervoor is het noodzakelijk dat het *stabiël* is als oplossing van (1). Stabiliteit ‘meet’ of willekeurig kleine verstoringen van een patroon al dan niet de neiging hebben om te groeien. Is dat laatste het geval, dan noemt men het patroon *instabiël*, en zal het niet waargenomen kunnen worden, omdat er zowel in de natuur als in (numerieke) simulaties altijd verstoringen zijn. Wiskundig bepaalt men stabiliteit als volgt: men neemt een stationaire oplossing $(U^*(\xi), V^*(\xi))$ van (1), met ξ gedefinieerd als hierboven, en voegt hier een kleine verstoring aan toe:

$$\begin{aligned} (U(x, t), V(x, t)) = \\ (U^*(\xi), V^*(\xi)) + \varepsilon(\tilde{U}(\xi, t), \tilde{V}(\xi, t)), \end{aligned} \quad (2)$$

waarbij $0 < \varepsilon \ll 1$ ‘asymptotisch klein’ is. Als $(\tilde{U}(\xi, t), \tilde{V}(\xi, t))$ begrensd blijft, of zelfs kleiner wordt, noemen we $(U^*(\xi), V^*(\xi))$ *stabiël* en is het patroon dus waarneembaar: kleine verstoringen blijven klein, en de oplossing $(U(x, t), V(x, t))$ van (1) blijft dicht bij $(U^*(\xi), V^*(\xi))$ als de tijd t evolueert.

Om inzicht te krijgen in het (leidende-orde-) gedrag van de verstoring $(\tilde{U}(\xi, t), \tilde{V}(\xi, t))$ substitueert men (2) in (1) en ontwikkelt naar ε . Omdat de basistoestand $(U^*(\xi), V^*(\xi))$ oplossing is van de bijbehorende ODE-reductie, verdwijnen hierbij alle termen zonder voorfactor ε . Men kan dus een factor ε uitdelen en vervolgens de limiet $\varepsilon \rightarrow 0$ nemen. Dit geeft het gelineariseerde systeem

$$\partial_t \begin{pmatrix} \tilde{U} \\ \tilde{V} \end{pmatrix} = \begin{pmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} \\ \mathcal{L}_{21} & \mathcal{L}_{22} \end{pmatrix} \begin{pmatrix} \tilde{U} \\ \tilde{V} \end{pmatrix}, \quad (3)$$

waarin $\mathcal{L}_{ij} = \mathcal{L}_{ij}(\partial_\xi; U^*(\xi), V^*(\xi))$ ξ -afhankelijke, tweede-orde-operatoren zijn (voor $i = j$) die expliciet afhangen van $(U^*(\xi), V^*(\xi))$. Omdat dit systeem lineair is, kan $(\tilde{U}(\xi, t), \tilde{V}(\xi, t))$ met behulp van Fourier-theorie ontbonden worden in $(\tilde{u}(\xi), \tilde{v}(\xi))e^{\lambda t}$, zodat de vraag naar de stabiliteit van $(U^*(\xi), V^*(\xi))$ reduceert tot het vinden van ‘eigenwaarden’

$\lambda \in \mathbb{C}$ van het spectrale probleem

$$\begin{pmatrix} \mathcal{L}_{11}(\frac{d}{d\xi}; \xi) & \mathcal{L}_{12}(\frac{d}{d\xi}; \xi) \\ \mathcal{L}_{21}(\frac{d}{d\xi}; \xi) & \mathcal{L}_{22}(\frac{d}{d\xi}; \xi) \end{pmatrix} \begin{pmatrix} \tilde{u} \\ \tilde{v} \end{pmatrix} = \lambda \begin{pmatrix} \tilde{u} \\ \tilde{v} \end{pmatrix}. \quad (4)$$

Als er λ bestaan met $\Re(\lambda) > 0$ dan zijn er exponentieel groeiende verstoringen en is $(U^*(\xi), V^*(\xi))$ dus instabiël.

In het algemeen zal het spectrum niet uit geïsoleerde eigenwaarden bestaan. Als $(U^*(\xi), V^*(\xi))$ een homogeen evenwicht is, kan met behulp van een Fourier-transformatie (door middel van $\mathcal{L}_{ij}(\frac{d}{d\xi}) = \mathcal{L}_{ij}(ik)$ met $k \in \mathbb{R}$) de operatormatrix in (4) vervangen worden door een gewone (constante coëfficiënten) 2×2 matrix. Het spectrum bestaat dan uit twee expliciet uit te rekenen krommes $\lambda = \lambda_{\pm}(k)$ in het complexe vlak.

Als $(U^*(\xi), V^*(\xi))$ een periodiek vegetatiepatroon is, is het lineaire systeem periodiek in ξ . Met behulp van Floquet–Bloch-theorie kan worden aangetoond dat het spectrum van (4) bestaat uit een collectie van oneindig veel niet noodzakelijk gladde beelden van de eenheidsdissc S^1 . Ook hier is in het algemeen moeilijk een vinger achter te krijgen, maar kan de singuliere structuur van het systeem worden uitgebuit om tot een expliciete analytische karakterisering van het spectrum te komen [12]. Het belangrijkste verschil tussen spectra van periodieke en homogene patronen is dat er vanwege translatiesymmetrie altijd een spectrale kromme is die door de oorsprong $(\lambda, k) = (0, 0)$ gaat. Dit betekent dat periodieke patronen (op $\mathbb{R}$) niet exponentieel stabiel zijn, maar op zijn best ‘diffusief stabiel’ (verstoringen nemen af als oplossingen van de warmtevergelijking) [17]. Omdat de spectrale krommen symmetrisch zijn onder $k \rightarrow -k$ is er dus ook een speciale, robuuste manier waarop een periodiek patroon instabiël kan worden: door verandering van de kromming van de spectrale kromme door $\lambda = 0$. Dit is een zogenaamde *sideband-instabiliteit*, die een belangrijke rol speelt in de dynamica van vegetatiepatronen (zie Figuur 4).

Ontstaan van patronen en Busse-ballon

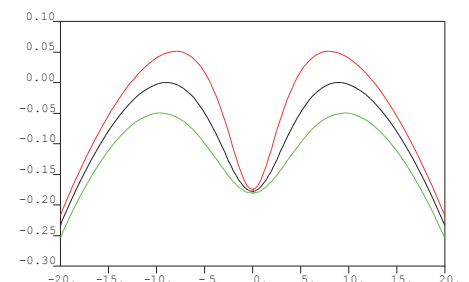
Uit de hierboven geschetste berekeningen volgt dat de ‘desert state’ $(U^0, V^0) = (1, 0)$ altijd stabiel is. Het homogene-begroeiings-evenwicht (U^-, V^-) is altijd instabiël, (U^+, V^+) is stabiel bij voldoende regenval. Als de regenval-parameter A afneemt dan wordt bij de bifurcatiewaarde $A = A_{TH}$ ook (U^+, V^+) instabiël. Voor $A = A_{TH}$ raakt de spectrale tak $\Re(\lambda_{\pm}(k))$ in de ‘meest kritische golfgetallen’

$k = \pm k_{TH}$ aan de $\{\Re(\lambda) = 0\}$ -as; voor A iets onder $A = A_{TH}$ zijn er twee continue intervallen van k -waarden waarvoor $\frac{2\pi}{k}$ -periodieke verstoringen exponentieel groeien (zie Figuur 5). Omdat $\lambda(k_{TH})$ in het algemeen (puur) imaginair is, zullen deze verstoringen zowel tijds- als ruimtelijk periodiek zijn: het zijn lopende golven. Als $\lambda(k_{TH}) = 0$ staan de golven stil en wordt de bifurcatie een Turing-bifurcatie genoemd — dit is het geval bij een vlak terrein, ofwel als $C = 0$ in (1). Op hellende terreinen is er sprake van een Turing–Hopf-bifurcatie.

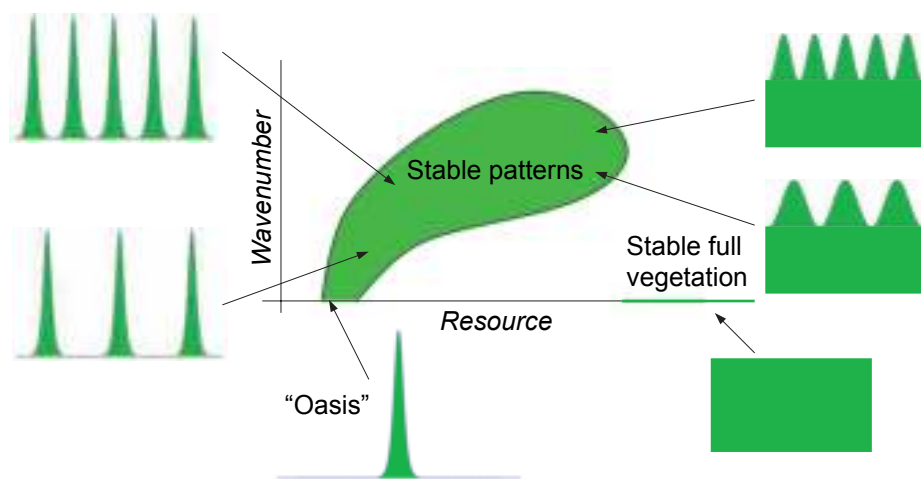
De lineaire aanpak suggereert dus dat er ruimtelijk periodieke patronen met golfgetallen dichtbij $k = k_{TH}$ zullen ontstaan als A de waarde A_{TH} passeert. Of dit daadwerkelijk het geval is wordt bepaald door niet-lineaire effecten: alleen als de bifurcatie superkritisch is, is de lineaire ‘intuïtie’ correct, bij een subkritische bifurcatie ontstaan dergelijke patronen in het algemeen niet. Met *Ginzburg–Landau-theorie* kan het karakter van de bifurcatie worden bepaald, en tevens welke van de bifurkerende ‘golftreinen’ stabiel is en welke niet [9]. In [18] is aangetoond dat de Turing–Hopf-bifurcatie in (1) altijd superkritisch is en dat er een continue familie van stabiele golftreinen ontstaat als A de waarde A_{TH} passeert.

In de woestijncontext betekent de Turing–Hopf-bifurcatie dat er periodieke vegetatiepatronen ontstaan vanuit een volledig begroeid terrein. Dichtbij A_{TH} hebben deze nog niet de typische structuur van kale grond afgewisseld met vegetatie, maar moet men eerder denken aan een begroeiing met periodiek variërende dichtheid. Met name door de singuliere structuur van het model neemt de familie van stabiele vegetatiepatronen bij afnemende A snel de vorm aan die men verwacht.

De *Busse-ballon* is gedefinieerd als het gebied in (parameter A , golfgetal κ)-ruimte waarvoor (1) stabiele periodieke oplossingen heeft, de *waarneembare* vegetatiepatronen. Zie Figuur 6 voor een schets. Merk hierbij op dat het ‘niet-lineaire golfgetal’ κ conceptueel



Figuur 5 De kromme $\lambda_+(k)$ uit het spectrum van (U^+, V^+) in het $(k, \Re(\lambda(k)))$ -vlak voor $A > A_{TH}$ (stabiël, groen), $A \approx A_{TH}$ (zwart) en $A < A_{TH}$ (instabiël, rood); $C = B = 0,2$, $\delta^2 = 0,001$, $\gamma = 2$.



Figuur 6 Schets van een typische reactie-diffusie-Busse-ballon

verschilt van het lineaire golfgetal k in bovenstaande stabiliteitsanalyse.

Veruit het grootste deel van de Busse-ballon kan in het algemeen niet analytisch worden bepaald. Er is een classificatie mogelijk van het karakter van de rand van de Busse-ballon [14], maar er is nog bijzonder weinig bekend over de structuur ervan. Zelfs het betrouwbaar bepalen van deze rand met numerieke continuatieprogramma's is een subtiële aangelegenheid, die op zich het onderwerp van wiskundig onderzoek is [13]. Opvallend aan Busse-ballonnen voor reactie-diffusievergelijkingen van type (1) is dat deze zich uitstrekken tot de rand $\kappa = 0$. Met andere woorden, patronen met een geïsoleerde homocliene puls — oases in de ecologische context — spelen een belangrijke rol bij het begrijpen van de Busse-ballon. Juist deze patronen kunnen in het singuliere geval wel analytisch worden bestudeerd. Onderzoek

hieraan heeft tot een intrigerende wiskundige *spin-off* geleid: de rand van de Busse-ballon heeft in de buurt van de homocliene oase een fijnstructuur van twee om elkaar heen draaiende en elkaar aftelbaar vaak doorsnijdende Hopf-bifurcatiekrommen [3, 18]. En passant is ook een vermoeden uit de literatuur van Ni [10] gevalideerd, dat, in ecologische bewoording, claimt dat oases de stabielste patronen zijn die dus als laatste zullen destabiliseren. Dus hoewel de Busse-ballon in het algemeen zeer moeilijk analytisch te begrijpen is, is er verrassend genoeg zowel bij het ontstaan van patronen — de Turing–Hopf-neus —, als bij het definitief verdwijnen ervan — de homocliene oases — wel degelijk analyse mogelijk.

Biljarten in de Busse-ballon

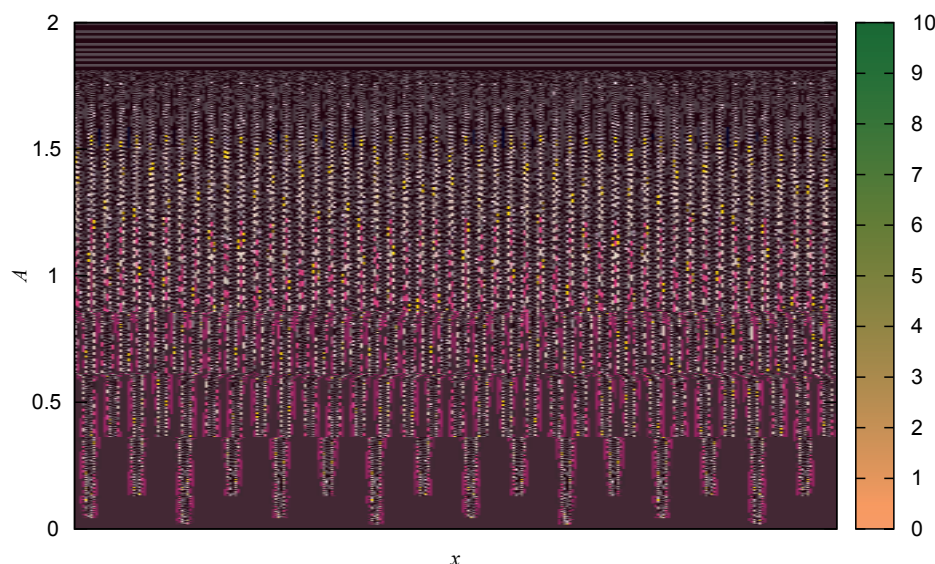
Nu we het gebied van alle mogelijk waarneembare vegetatiepatronen, de Busse-ballon, kennen, kunnen we proberen te begrijpen

pen hoe (1) reageert op een langzaam veranderend klimaat, ofwel, wat er gebeurt als A langzaam gaat afnemen in de tijd. Zie Figuur 7 voor een numerieke simulatie.

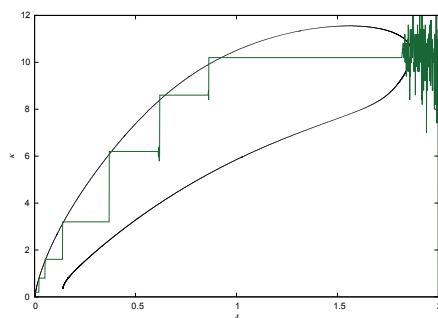
Wat opvalt is dat de verwoestijnscatastrofe voorafgegaan wordt door een aantal 'mini-catastrofes' waarbij het vegetatiepatroon 'opeens' een ander golfgetal aanneemt. Tevens blijkt het systeem bij simulaties niet de meest stabiele oase te bereiken: verwoestijning vindt net als in de realiteit al eerder plaats bij een periodiek patroon. De cruciale rol van de Busse-ballon wordt duidelijk als we de baan van een simulatie door de Busse-ballon plotten (zie Figuur 8).

De mini-catastrofes vinden plaats als deze baan de rand van de Busse-ballon raakt of doorsnijdt: het patroon wordt door de rand naar beneden gekaatsd. Dit kaatsen is uitstekend te begrijpen, buiten de Busse-ballon zijn tenslotte geen stabiele vegetatiepatronen. Een in essentie onbeantwoorde vraag is welk patroon (golfgetal) wordt aangenomen ná dit kaatsen. Omdat veruit het grootste deel van de rand van de Busse-ballon bestaat uit patronen die door het sideband-mechanisme instabiel worden, komt begrijpen hiervan neer op inzicht in de dynamica van patronen in de buurt van een sideband-instabiliteit onder langzaam veranderende omstandigheden.

Uit computersimulaties blijkt dat het veranderingsproces gevoelig afhangt van de snelheid waarmee de regenval-parameter A varieert. Als deze snelheid (relatief) hoog is, dan neemt de voorspellende waarde van de Busse-ballon af: in systemen als (1) wordt de sideband-instabiliteit typisch gevolgd door een periode-verdubbelingsdestabilisatie die lineair sterker groeit en dus zal domineren. Voor nog hogere snelheden is aanpassing helemaal niet meer mogelijk en zal verwoestijning meteen plaatsvinden bij het golfgetal $\kappa = k_{TH}$ van de Turing–Hopf-bifurcatie. Dit wordt in de realiteit echter niet waargenomen. Ook de hoeveelheid aan het systeem toegevoegde ruis heeft een significante invloed op de dynamica: grofweg kan gezegd worden dat het systeem dichter in de buurt van de rand van de Busse-ballon blijft als de hoeveelheid ruis toeneemt. Ook dit is wiskundig nog niet begrepen. Tot slot is het nog geheel onduidelijk wat de precieze mechanismen zijn die aanleiding geven tot de laatste verwoestijningsstap: de baan van het periodieke patroon raakt de rand van de Busse-ballon maar 'kaatsd' niet meer. In plaats daarvan verlaat hij de Busse-ballon definitief om 'neer te slaan' in het enige overgebleven stabiele patroon (U^0, V^0), de woestijn. Omdat deze stap plaatsvindt in het



Figuur 7 De dynamica van vegetatiepatronen bij langzaam afnemende A (begroeide gebieden zijn donker/groen). In de simulatie is een kleine (ruimtelijke) stochastische verstoring toegevoegd aan het deterministische systeem (1).



Figuur 8 Biljarten in de Busse-ballon. De baan geeft het dominante golfgetal van het bovenstaande evoluerende vegetatiepatroon aan.

gebied waar asymptotische analyse mogelijk lijkt, is er hoop dat dit proces binnenkort beter begrepen zal worden.

Terug naar de woestijn

De belangrijkste vraag is natuurlijk: “Wat kan

het opgebouwde wiskundig inzicht in de dynamica van eenvoudige patronen in het een-dimensionale model (1) nu betekenen voor een beter begrip van verwoestijning?”

Het is duidelijk dat (1) niet direct gebruikt kan worden om kwantitatieve voorspellingen te doen over de overlevingskans van bestaande begroeiing aan de rand van een woestijn. Daarvoor zijn veel realistischere modellen nodig. Echter, inzicht in de dynamica van patronen in vereenvoudigingen als (1) geeft wel fundamenteel inzicht in de verwoestijnings-catastrofe. Wiskundig gezien blijkt de catastrofe direct gerelateerd te zijn aan het onder afnemende regenval, of toenemende stress, ‘verlaten’ van de Busse-ballon horende bij het model. De ‘afstand’ tussen een patroon en de rand van de Busse-ballon zou wellicht gezien kunnen worden als een soort maat voor de kans op verwoestijning, maar dat is bij een

langzaam veranderend klimaat waarschijnlijk een veel te grove maat. In de realiteit is het door de vanzelfsprekende aanwezigheid van stochastische effecten zelfs aannemelijk dat waargenomen vegetatiepatronen zich meestal dichtbij van de rand van de Busse-ballon zullen bevinden. De echte vraag blijft dus wat een model of een echt systeem er uiteindelijk toe brengt om te biljarten en vervolgens al dan niet de Busse-ballon te verlaten.

Dankwoord

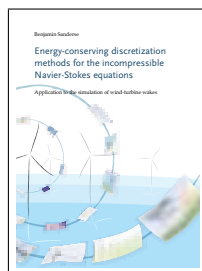
Dit onderzoek aan vegetatiepatronen en verwoestijning is mogelijk gemaakt met steun vanuit het NWO-programma Complexity.

Referenties

- 1 F.H. Busse, Nonlinear properties of thermal convection, *Rep. Prog. Phys.* 41 (1978), 1929–1967.
- 2 A. Doelman, T. J. Kaper en P. Zegeling, Pattern formation in the one-dimensional Gray-Scott model, *Nonlinearity* 10 (1997), 523–563.
- 3 A. Doelman, J.D.M. Rademacher en S. van der Stelt, Hopf dances near the tips of Busse balloons, *Discr. Cont. Dyn. Syst. (S)* 5(1) (2012), 61–92.
- 4 E. Gilad, J. von Hardenberg, A. Provenzale, M. Shachak en E. Meron, Ecosystem engineers: from pattern formation to habitat creation, *PRL* 93(9) (2004), 098105.
- 5 P. Gray en S.K. Scott, Autocatalytic reactions in the isothermal, continuous stirred tank reactor - oscillations and instabilities in the system $A + 2B \rightarrow 3B$; $B \rightarrow C$, *Chem. Eng. Sc.* 39 (1984), 1087–1097.
- 6 B.J. Kealy en D.J. Wollkind, A nonlinear stability analysis of vegetative Turing pattern formation for an interaction-diffusion plant-surface water model system in an arid flat environment, *Bull. Math. Biol.* 74 (2012), 803–833.
- 7 C.A. Klausmeier, Regular and irregular patterns in semi-arid vegetation, *Science* 284 (1999), 1826–1828.
- 8 M. Meyries, J.D.M. Rademacher en E. Siero, preprint (2013).
- 9 A. Mielke, The Ginzburg-Landau equation in its role as modulation equation, in *Handbook of Dynamical Systems, II*, (ed. B. Fiedler) (2002), 759–835.
- 10 W.-M. Ni, Diffusion, cross-diffusion, and their spike-layer steady states, *Not. AMS* 45(1) (1998), 9–18.
- 11 J. E. Pearson, Complex patterns in a simple system, *Science* 261 (1993), 189–192.
- 12 H. van der Ploeg en A. Doelman, Stability of spatially periodic pulse patterns in a class of singularly perturbed reaction-diffusion equations, *Ind. Univ. Math. J.* 54(5) (2005), 1219–1301.
- 13 J.D.M. Rademacher, B. Sandstede en A. Scheel, Computing absolute and essential spectra using continuation, *Phys. D* 229(2) (2007), 166–183.
- 14 J.D.M. Rademacher en A. Scheel, Instabilities of wave trains and Turing patterns in large domains, *Int. J. Bif. Chaos* 17(8) (2007), 2679–2691.
- 15 M. Rietkerk, M.C. Boerlijst, F. van Langevelde, R. Hillerislambers, J. van de Koppel, L. Kumar, H.H.T. Prins en A.M. de Roos, Self-Organization of Vegetation in Arid Ecosystems, *The American Naturalist* 160(4) (2002), 524–530.
- 16 M. Rietkerk, S.C. Dekker, P.C. de Ruiter en J. van de Koppel, Self-organized patchiness and catastrophic shifts in ecosystems, *Science* 305, (2004) 5692.
- 17 B. Sandstede, A. Scheel, G. Schneider en H. Uecker, *J. Diff. Eqns.* 252 (2012), 3541–3574.
- 18 S. van der Stelt, A. Doelman, G. Hek en J.D.M. Rademacher, Rise and fall of periodic patterns for a Generalized Klausmeier-Gray-Scott model, *J. Nonl. Sc.* 23(1) (2012), 39–95.
- 19 J. Von Hardenberg, E. Meron, M. Shachak en Y. Zarmi, Diversity of Vegetation Patterns and Desertification *PRL* 87(19) (2001), 198101.

In de verdediging

| In defence



Energy-conserving discretization methods for the incompressible Navier–Stokes equations

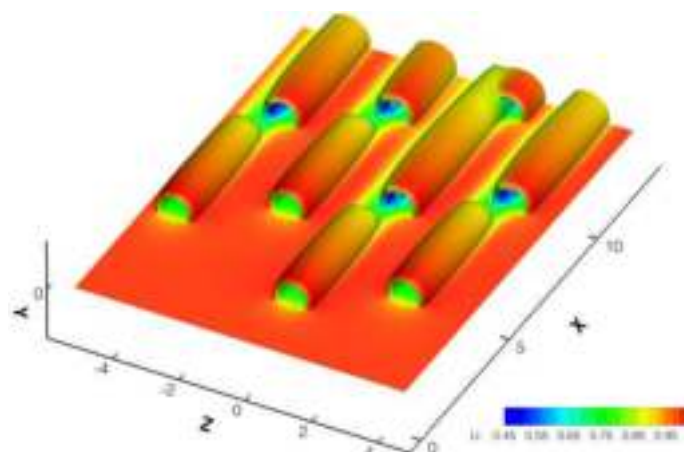
Benjamin Sanderse

“Onderzoek doen aan de Navier–Stokes-vergelijkingen is extra spannend zolang de existentie en uniciteit van hun oplossingen niet bewezen is.” Zo luidt de volgens hemzelf mooiste stelling bij het proefschrift van Benjamin Sanderse, die op 19 maart cum laude promoveerde aan de Technische Universiteit Eindhoven. Op het eerste gezicht een verrassend ‘zuiver wiskundige’ stelling voor iemand die werkte aan windmolenparksimulaties voor het Energieonderzoek Centrum Nederland (ECN) in Petten. Een groot deel van zijn onderzoek en zijn belangrijkste resultaten waren echter heel fundamenteel numeriek wiskundig van aard. “Het werk aan de Navier–Stokes-vergelijkingen, ook al is dat heel bescheiden en met behulp van numerieke methoden, is juist zo interessant omdat je een kleine bijdrage kan leveren aan het begrip van het totale probleem. Misschien vinden we met nauwkeurige numerieke simulaties wel hints die het wiskundige bewijs kunnen helpen”, aldus Sanderse.

Sanderse schreef zijn proefschrift *Energy-conserving discretization methods for the incompressible Navier–Stokes equations; Application to the simulation of wind-turbine wakes* onder begeleiding van prof.dr.ir. Barry Koren en deed een groot deel van het onderzoek aan het Centrum Wiskunde & Informatica. Hij ontwikkelde numerieke methoden voor de simulatie van turbulente stromingen, beschreven door de incompressibele Navier–Stokes-vergelijkingen, en gebruikte deze methoden om de stroming van lucht in windmolenparken uit te rekenen.

Windmolenparken en numerieke simulaties

In windmolenparken op zee staan grote aantallen windmolens dichtbij elkaar opgesteld. Nadeel daarvan is dat de windmolens de luchtstro-



Simulatie van een park met negen turbines. Achterliggende turbines hebben duidelijk een langer zog en de stromingssnelheid neemt er af tot onder de 50% van de vrije-veldwindsnelheid.

Pas gepromoveerden brengen hun werk onder de aandacht.
Heeft u tips voor deze rubriek of bent u zelf pas gepromoveerd?
Laat het weten aan onze redacteur.

Redacteur: Geertje Hek
la Voie-du-Coin 7
1218 Grand-Saconnex
Zwitserland
verdediging@nieuwarchief.nl

mingen deels voor elkaar verstoren. De afname in stromingssnelheid leidt tot een vermogensverlies van gemiddeld zo'n tien procent, vergeleken met individueel opgestelde windmolens, en de toename in turbulente wervelingen zorgt voor een verhoogde belasting op de turbinebladen.

Met kennis over de luchtstromingen in een windmolenpark zouden het ontwerp van de turbines en de opstelling van windmolens in parken verbeterd kunnen worden. Simulaties van luchtstromingen aan de hand van rekenmodellen komen echter niet altijd goed overeen met de werkelijkheid. Turbulente stromingen worden gemodelleerd door de Navier–Stokes-vergelijkingen, partiële differentiaalvergelijkingen die de stroming van lucht en vloeistoffen beschrijven. Om numerieke simulaties te doen, worden deze continue vergelijkingen gediscretiseerd. Dit brengt echter vaak fouten met zich mee, omdat de oplossing van een gediscretiseerd probleem in het algemeen niet precies overeenkomt met de oplossing van het continue probleem. Als een eenmaal ontstane fout niet te hard groeit gedurende de berekening, wordt een numeriek algoritme stabiel genoemd.

In het geval van de Navier–Stokes-vergelijkingen introduceren veel standaard discretisatiemethoden kunstmatige diffusie om de oplossing stabiel te houden. Zulke 'numerieke' diffusie kan turbulente stromingsfenomenen echter uitdempen. Omdat juist die fenomenen belangrijk zijn voor het begrip van de aërodynamica, bestudeerde Sanderse in zijn proefschrift methoden die bepaalde invarianten van de continue vergelijkingen behouden wanneer oplossingen worden benaderd. Dit zorgt voor stabiele simulaties en voor fysisch realistische oplossingen. Specifieker, 'zijn' methodes introduceren geen kunstmatige diffusie, maar zijn toch stabiel doordat ze de totale kinetische energie van de stroming in het niet-viskeuze geval behouden. De kinetische energie van de stroming is in dit geval een (kwadratische) invariant die een bovengrens aan de oplossing stelt. Als de discretisatie die energie behoudt, is de oplossing dus ook begrensd, waardoor stabiliteit verzekerd is.

Een lastig en groot probleem

In een windmolenpark is er een wisselwerking tussen de lokale stroming rondom individuele turbinebladen en het grote zog dat achter iedere turbine ontstaat en de stroming rond achterliggende molens beïnvloedt. Het gedetailleerd uitrekenen van de stroming rondom de turbinebladen vraagt om rekencellen met een typische afmeting in de orde van millimeters, terwijl de belangrijke zogeffecten over afstanden van honderden meters tot kilometers plaatsvinden. Het is daarom zelfs met enorm krachtige (parallele) computers onmogelijk om beide effecten nauwkeurig uit te rekenen. Voor turbinebladsimulaties wordt daarom het zog versimpeld of maar beperkt uitgerekend. Een standaardaanpak voor zogsimulaties is om niet het turbineblad zelf te modelleren maar wel hun effect: het turbineblad wordt versimpeld door gebruik te maken van zogenaamde *actuator disks* (trekkende schijven) met voorgeschreven krachten op de plek van de turbine. Voor het implementeren van deze krachten, met hun resulterende discontinuïteiten in de stromingsvariabelen, heeft Sanderse een nieuwe, scherpe en niet-diffusieve methode voorgesteld.

Energiebehoud en nauwkeurighedsanalyses

Voor het zover was construeerde hij eerst energiebehoudende discretisatiemethoden van het continue probleem, in het bijzonder tweede en vierde-orde nauwkeurige eindige volumemethoden op zogenaamde

staggered cartesische roosters. De variabelen worden hierbij niet alle op dezelfde roosterpunten gekozen, maar om en om, met als resultaat een discretisatie die massa, impuls, energie en vorticeit behoudt. Sanderse heeft nieuwe randvoorwaarden voor deze methoden afgeleid, die zodanig zijn dat de bijdragen van de rand aan de discrete energievergelijking de bijdragen van de continue energievergelijking nabootsen. Hij analyseerde bovendien de bijdrage van de rand op de nauwkeurigheid van de methoden.

Als de (incompressibele) Navier–Stokes-vergelijkingen eenmaal ruimtelijk gediscretiseerd zijn, komt de tijdsintegratie aan bod. Hiervoor worden standaard Runge–Kutta-methoden (RK-methoden) gebruikt, maar waarom die precies werken voor de Navier–Stokes-vergelijkingen was nog onbekend. RK-methoden zijn ontworpen voor gewone differentiaalvergelijkingen $\frac{du}{dt} = f(u, t)$, maar eenmaal gediscretiseerd in de ruimte zijn de Navier–Stokes-vergelijkingen geen systeem van gewone differentiaalvergelijkingen, maar een systeem van differentiaal-algebraïsche vergelijkingen. Het belangrijkste resultaat in het proefschrift van Sanderse is zijn nauwkeurighedsanalyse van de RK-methoden, toegepast op de Navier–Stokes-vergelijkingen. Er bestond al wel theorie over hoe RK-methoden op differentiaal-algebraïsche vergelijkingen kunnen worden toegepast, maar dat was nog niet eerder voor Navier–Stokes gedaan. Door die theorie te gebruiken kon Sanderse heel veel bestaande methoden in een bredere context plaatsen en werd duidelijk waarom deze bestaande methoden werken.

Fundamentele wiskunde voor praktische doeleinden

De vraag rijst hoe Sanderses fundamentele werk te rijmen is met het feit dat zijn onderzoek volledig door ECN werd gefinancierd. "Het mooie is, dat mijn resultaat wel direct toegepast kan worden voor het uitrekenen van praktische problemen. Ik heb het ook in mijn code gestopt voor het simuleren van stromingsproblemen en het was in die zin dus ook meteen nuttig voor ECN", aldus Sanderse. Hij roemt de combinatie van het werken bij het CWI en ECN. Op het CWI heerst een fantastische onderzoekssfeer, een ideale plek voor concentratie en rust, maar er was ook altijd wat te doen naast het werk. De absolute onderzoeksvrijheid (met dank aan Barry Koren) vond hij een enorm voorrecht. Op het ECN werd hij altijd bewust gemaakt van de noodzaak om wiskundige modellen naar de praktijk te brengen en zijn onderzoek toe te passen op windenergie-problemen.

Sanderse bracht ook twee maanden door bij het National Renewable Energy Laboratory (NREL) in het Amerikaanse Boulder, Colorado. Een buitenkans voor Sanderse, die heel erg van wielrennen houdt. Niet alleen kon hij volledig ongestoord aan zijn onderzoek werken, hij kon het werk ook afwisselen met wielrennen en wandelen in de Rocky Mountains.

Ten tijde van het interview combineerde hij wielrennen met zwemmen en hardlopen om te trainen voor een halve triatlon, die hij inmiddels heeft voltooid. Na zijn promotie is hij bij Shell in Amsterdam aan de slag gegaan. Hij werkt daar aan numerieke methoden voor de simulatie van meergefase-stromingen (gas, olie en water). Ook hiervoor worden de Navier–Stokes-vergelijkingen gebruikt. Of hij de diepgang en de langetermijnprojecten van het academische onderzoek gaat missen of niet zal moeten blijken. Voorlopig vindt Sanderse het vooral interessant om te zien hoe er in het bedrijfsleven wordt gewerkt, en dat wiskundige methoden niet alleen in artikelen voortleven, maar echt worden gebruikt voor het oplossen van praktische problemen.



Cornelis Johannes van Duijn

Executive Board

Eindhoven University of Technology

c.j.v.duijn@tue.nl

Iuliu Sorin Pop

Mathematics and Computer Science

Eindhoven University of Technology

i.pop@tue.nl

Shock waves and two phase porous media flows

Models for porous media flow with multiple fluids (water and oil, for example) are important for studying techniques such as geological storage of CO₂ and water driven oil recovery. These models consist of conservation laws and often contain nonlinearities that preclude the existence of classical solutions. Hans van Duijn and Sorin Pop discuss important aspects of the theory of hyperbolic conservation laws, in particular existence and uniqueness of weak solutions, admissibility conditions and shock solutions. This theory explains saturation overshoots occurring in porous media flows under non-equilibrium conditions.

Mathematical models and simulation tools are essential for an optimal exploitation of geological resources. In many cases, for instance groundwater flow, water driven oil recovery, or geological CO₂ storage [1, 7, 12], the underlying mathematical models consist of conservation laws. These are nonlinear evolution equations having either a hyperbolic or a parabolic character. The nonlinearities and data appearing in the models allow for the existence of so-called weak solutions, i.e. discontinuous solutions. The weak solutions are defined, based on test functions and partial integration. Natural questions arising in this context are existence, uniqueness, and regularity of these solutions. For parabolic problems, uniqueness generally holds. This is not the case of hyperbolic problems, where uniqueness requires additional admissibility conditions, leading to so-called *entropy solutions*.

In this paper we discuss two admissibility concepts. They are based on the observation that hyperbolic equations arise from evolution equations involving higher order derivatives, by letting such terms vanish. In porous media models, this is similar to vanishing capillarity.

Flows in porous media

In this paper we analyze a model that describes the flow of two incompressible and immiscible fluids through a porous medium. Such models arise in geological CO₂ storage. We follow the approach in which the porous medium is considered as a continuum, characterized by averaged properties such as porosity ϕ and permeability K . The presence of each fluid is described by a saturation, S_w for water and S_n for CO₂. The model equations arise from a combination of mass balance and Darcy laws for both fluids [1, 7]. When restricting ourselves to horizontal flows through homogeneous reservoirs, the following one-dimensional description results (with $\alpha \in \{w, n\}$):

$$\phi \frac{\partial S_\alpha}{\partial t} + \frac{\partial v_\alpha}{\partial x} = 0 \quad (\text{mass balance}), \quad (1)$$

$$v_\alpha = -K \frac{k_{r\alpha}}{\mu_\alpha} \frac{\partial p_\alpha}{\partial x} \quad (\text{Darcy}). \quad (2)$$

For each fluid, v_α is the averaged velocity, μ_α the dynamic viscosity and p_α the pressure in fluid α . Furthermore, $k_{r\alpha}$ denotes the relative permeability of fluid α . Experiments

show that it is an increasing function of the corresponding fluid saturation, i.e.

$$k_{r\alpha} = k_{r\alpha}(S_\alpha). \quad (3)$$

One can insert (2) into (1) to reduce the system to two equations, involving, however, four unknowns, two saturations and two pressures. To close the system one assumes the pores are occupied completely by the two fluids (thus no other fluid or void is present), implying

$$S_w + S_n = 1. \quad (4)$$

The difference of the fluid pressures, called *capillary pressure*, depends on the CO₂ saturation S_n . In standard models it is given by

$$p_n - p_w = P_c(S_n), \quad (5)$$

where P_c is strictly increasing.

From (1) and (4) one concludes that the sum of the fluid velocities, $v = v_n + v_w$, is constant in space. It remains constant in time too if the injection process is at constant flow rate. Introducing reference quantities the model reduces to one scalar equation

$$\partial_t u + \partial_x f(u) = \varepsilon (H(u) \partial_x P) \quad (6)$$

for all $t > 0$ and $x \in \Omega$,

which involves only one dimensionless num-

ber ε , the *capillary number*. This is related to the ratio of viscous forces and surface tension acting across the interface between the two fluids.

In (6) u denotes the CO_2 saturation, $u = S_n$, P is the dimensionless capillary pressure and Ω is the dimensionless interval of interest. Below we consider $\Omega = \mathbb{R}$. The nonlinear functions are:

$$\begin{aligned} f(u) &= \frac{k_o(u)}{k_o(u) + Mk_w(u)}, \\ H(u) &= k_w(u)f(u). \end{aligned} \quad (7)$$

Typically, the function f has a convex-concave profile.

In view of (5) one has

$$P = P(u). \quad (8)$$

Common choices in the porous media literature are

$$\begin{aligned} k_o(u) &= u^{1+p}, \quad k_w(u) = (1-u)^{1+q} \\ \text{and } P(u) &= (1-u)^{-\frac{1}{\lambda}}, \end{aligned}$$

where $p, q > 0$, $\lambda > 1$ and $M > 0$ are model specific, [1].

Non-equilibrium models

The above-mentioned model is based on permeability — saturation and capillary pressure — saturation functions under so-called ‘equilibrium conditions’. In this case, measurements are performed after achieving a static distribution of the fluids inside the pores. However, capillary pressure functions measured under non-equilibrium conditions, when the fluids do not reach a steady state, are different from those under equilibrium ones. Further, unexpected saturation profiles have been measured in [2] for infiltration in a thin and long column filled by homogeneous sand. As shown in Figure 1, the saturation profiles for a low flux at the inflow are monotone, as predicted by equilibrium models (see e.g. [10]). For higher fluxes, so-called ‘saturation overshoots’ are observed: at the infiltration front, the saturation values are higher than the ones at the influx. As the flux increases, this profile becomes even more intriguing: a saturation plateau that is higher than the inflow saturation appears between an infiltration front and a drainage front.

Such non-monotonic saturation profiles are ruled out by equilibrium models, there-

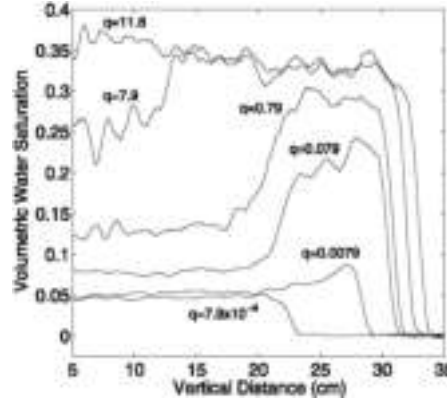


Figure 1 Saturation profiles measured for different fluxes (courtesy of D.A. DiCarlo); note that these profiles become non-monotonic with the increase of the flux.

fore non-equilibrium ones are required. A model incorporating dynamic effects in the capillary pressure is proposed in [6], where (5) becomes

$$P_c = P_c^e(S_n) + \hat{\tau} \partial_t S_n. \quad (9)$$

P_c^e is the capillary pressure-saturation function determined under equilibrium. The second term accounts for the non-equilibrium effects. The parameter $\hat{\tau}$ is assumed constant for simplicity; generally it may depend on S_n . In the simplified, dimensionless setting considered here, the capillary pressure becomes

$$P = P(u) + \varepsilon \tau \partial_t u, \quad (10)$$

with the dimensionless *dynamic number* τ . Then (6) reads

$$\begin{aligned} \partial_t u + \partial_x f(u) \\ = \varepsilon (H(u) \partial_x (P(u) + \tau \varepsilon \partial_t u)), \end{aligned} \quad (11)$$

for all $t > 0$ and $x \in \mathbb{R}$.

Remark. Both dimensionless numbers ε and τ depend on the reference quantities, see [11]. In particular, the dynamic number is proportional to the square of the reference velocity. In other words, a higher flux at the inflow results in a higher value for τ . This is a key observation for explaining the occurrence of saturation overshoots.

Admissible shocks

Formally, when letting ε tend to 0, both (6) or (11) become:

$$\partial_t u + \partial_x f(u) = 0 \quad \text{for } t > 0, x \in \mathbb{R}. \quad (12)$$

This equation is hyperbolic. We consider *Rie-*

mann initial data,

$$u(0, x) = u_0(x) = \begin{cases} u_\ell & \text{if } x < 0, \\ u_r & \text{if } x > 0, \end{cases} \quad (13)$$

with given states $u_\ell, u_r \in \mathbb{R}$. We only consider bounded solutions, satisfying

$$\begin{aligned} \lim_{x \rightarrow -\infty} u(t, x) &= u_\ell \\ \text{and } \lim_{x \rightarrow \infty} u(t, x) &= u_r \end{aligned} \quad (14)$$

for all $t > 0$. Moreover, since u is the CO_2 saturation, we only consider states satisfying $0 \leq u_\ell, u_r \leq 1$.

Having formulated the initial value problem (12)–(13), a natural question arising is the existence and uniqueness of solutions. Without entering into details, we mention that *classical solutions* fail to exist. Alternatively, one can multiply the original equation by a *test function* φ and integrate the resulting equation over the time-space domain. After integration by parts, a *weak solution* is defined as a bounded, measurable function satisfying

$$\begin{aligned} \int_0^\infty \int_{\mathbb{R}} u \partial_t \varphi + f(u) \partial_x \varphi \, dx \, dt \\ + \int_{\mathbb{R}} u_0(x) \varphi(0, x) \, dx = 0, \end{aligned} \quad (15)$$

for all $\varphi \in C_0^1([0, \infty) \times \mathbb{R})$, the space of continuously differentiable functions that vanish uniformly for large t and x . Clearly, any classical solution is also a weak solution, but the latter class is much wider. In particular, a weak solution may even be discontinuous.

The existence of a solution can be proved in the context above. However, uniqueness remains an open question, as this does not hold generally. For example, consider the *Burgers* equation,

$$\partial_t u + \partial_x \left(\frac{1}{2} u^2 \right) = 0 \quad \text{for } t > 0, x \in \mathbb{R}. \quad (16)$$

Let the initial condition in (13) with $u_\ell = 0$ and $u_r = 1$. One can verify that both functions given below are weak solutions:

$$u_a(t, x) = \begin{cases} 0 & \text{if } x \leq 0, \\ \frac{x}{t} & \text{if } 0 < x < t, \\ 1 & \text{if } x \geq t, \end{cases}$$

$$u_b(t, x) = \begin{cases} 0 & \text{if } x < \frac{1}{2}t, \\ 1 & \text{if } x > \frac{1}{2}t. \end{cases}$$

Note that u_a is a continuous transition from the left state $u_\ell = 0$ to the right state $u_r = 1$. These states are connected by a fan that becomes wider in time; such solutions are called *rarefaction waves*. At the same time, u_b is a *shock*, a jump from u_ℓ to u_r . This shock is located at $x(t) = \frac{1}{2}t$ and travels with a fixed velocity $s = x'(t) = \frac{1}{2}$. The shock satisfies the general *Rankine–Hugoniot* condition relating the shock speed to the states at its left and right sides,

$$s = \frac{f(u_\ell) - f(u_r)}{u_\ell - u_r}, \quad (17)$$

where f is the *flux function* in the conservation law (12). In general, one can prove that functions having discontinuities travelling with a velocity that respects (17), and satisfying the equation in classical sense away from the shock, are weak solutions (see e.g. [5]). This rises another question: *which solution is relevant?* In other words, additional criteria are required for selecting the weak solution of interest.

A natural approach is to use underlying physics and to investigate which structure has been lost by letting ε tend to zero in the original, *regularized* model. In particular, since the shock profiles are constant in time, one can think of similar solutions to the regularized model, *travelling waves* (TW). Such waves connect the same left and right states as the shock. In other words, one investigates solutions u^ε to the regularized equation, having the structure

$$u^\varepsilon(t, x) = v(\eta), \quad \text{with } \eta = (x - st)/\varepsilon, \quad (18)$$

and satisfying

$$\lim_{\eta \rightarrow -\infty} v^\varepsilon(\eta) = u_\ell \quad \text{and} \quad \lim_{\eta \rightarrow \infty} v^\varepsilon(\eta) = u_r. \quad (19)$$

A shock u is called *admissible* if it is the limit of a TW solution:

$$u(t, x) = \lim_{\varepsilon \rightarrow 0} v^\varepsilon(x - st). \quad (20)$$

Note that neither (6) nor (11) depend explicitly on t or x , so any translation of a TW remains a TW. Therefore we fix the wave by assuming that $v^\varepsilon(0) = (u_\ell + u_r)/2$.

In what follows we concentrate on shock solutions to (12), satisfying (in the sense of left and right limits)

$$\lim_{x \nearrow st} u(t, x) = u_\ell \quad \text{and} \quad \lim_{x \searrow st} u(t, x) = u_r. \quad (21)$$

Such shocks will be denoted by $\{u_\ell, u_r\}$. Following the above, we call the shock $\{u_\ell, u_r\}$ *admissible* if it is the limit ($\varepsilon \rightarrow 0$) of a TW v^ε connecting u_ℓ to u_r .

Oleinik's criterion

In the classical theory for hyperbolic conservation laws, a generally accepted admissibility criterion is due to O.A. Oleinik (see [9], Chapter II) and is based on the parabolic regularization of (12). In the context of two-phase flow in porous media, we consider (6) with the capillary pressure in (8). This deviates slightly for the standard approach, where the terms on the right are replaced by $\varepsilon \partial_{xx} u$. We seek TW solutions u^ε to (6), satisfying (21), and having the structure stated in (18). We assume that the waves v are smooth, i.e., twice continuously differentiable. Applying the chain rule, and integrating the resulting once with respect to η leads to

$$\begin{cases} A - sv + f(v) = H(v)(P(v))' & \text{for } \eta \in \mathbb{R}, \\ \lim_{\eta \rightarrow -\infty} v(\eta) = u_\ell \quad \text{and} \quad \lim_{\eta \rightarrow \infty} v(\eta) = u_r, \end{cases} \quad (22)$$

with $A, s \in \mathbb{R}$ to be determined. Note that the scaling in (18) allows eliminating ε from (22). However, if a TW solution exists, the profile of the corresponding u^ε becomes steeper as ε tends to 0. In the limit, this leads to a shock solution to (12).

The behaviour of v as $\eta \rightarrow \pm\infty$ implies that the term on the right in (22) tends to 0 as well, yielding

$$\begin{aligned} s &= \frac{f(u_\ell) - f(u_r)}{u_\ell - u_r} \\ \text{and } A &= \frac{u_r f(u_\ell) - u_\ell f(u_r)}{u_\ell - u_r}. \end{aligned} \quad (23)$$

Note that the TW speed is the shock speed introduced in (17). With $\beta(v) = \int H(v)P'(v)dv$, a solution of (22) solves

$$\begin{aligned} (\beta(v))' &= g(v) := A - sv + f(v), \\ \text{for } \eta &\in \mathbb{R}, \end{aligned} \quad (24)$$

subject to the initial condition $v(0) = (u_\ell + u_r)/2$.

Clearly, (23) is necessary, but not sufficient for the existence of TW solutions, as the latter implies that the solution of (24) also has the asymptotic behaviour (22). This behaviour can be reinterpreted in terms of phase plane

analysis, where u_ℓ acts as a source-type equilibrium, and u_r as a sink. Moreover, (24) should have no equilibria between the two states u_ℓ and u_r . This means that the term on the right in (24) does not vanish between the two states, hence g has a constant sign there. This shows that the waves are monotone. In mathematical terms, this becomes the celebrated *Oleinik entropy condition*, saying that shocks $\{u_\ell, u_r\}$ are admissible if and only if for all v between u_ℓ and u_r ,

$$\frac{f(v) - f(u_\ell)}{v - u_\ell} \geq s \geq \frac{f(v) - f(u_r)}{v - u_r}. \quad (25)$$

In particular, for (16) this implies that shocks $\{u_\ell, u_r\}$ are admissible if and only if $u_\ell > u_r$. Hence u_b is not admissible in the sense of Oleinik.

For model (6), where f has a convex-concave profile, this means that, if $u_\ell > u_r$, shocks $\{u_\ell, u_r\}$ are admissible only if the segment connecting the points $(u_\ell, f(u_\ell))$ and $(u_r, f(u_r))$ is above the graph of f and has no interior intersection points with the graph. For example, a shock $\{1, 0\}$ is not admissible in this case. A general feature of admissible shocks (according to this criterion) is that characteristics from the left and right sides of a shock converge into the shock (as time is increasing).

Non-classical shocks

Oleinik's approach uses a parabolic regularization of the hyperbolic equation (12), which corresponds to the equilibrium capillary pressure model (5). The TW solutions discussed before remain monotone. As mentioned, there are experimental results contradicting this monotonicity, which justifies non-equilibrium models as (11).

In this respect, one question is whether such models can justify *non-classical* shock solutions to (12), which violate Oleinik's admissibility condition (25). We refer to [9], where 'connectable' left and right states, or admissible shocks $\{u_\ell, u_r\}$, are defined based on a predefined *kinetic function*. Similarly, in [4] admissible shocks $\{u_\ell, u_r\}$ are defined as the limit $\varepsilon \rightarrow 0$ of TW solutions to (11), and satisfying (22).

From a physical point of view, such shocks are obtained in the vanishing capillary pressure limit of two-phase porous media flow models, when dynamic effects are included in the capillary pressure. Clearly, such shocks depend on τ ; for $\tau = 0$ one obtains the classical shocks in the previous subsection. Although formally the limit equation (12) rem-

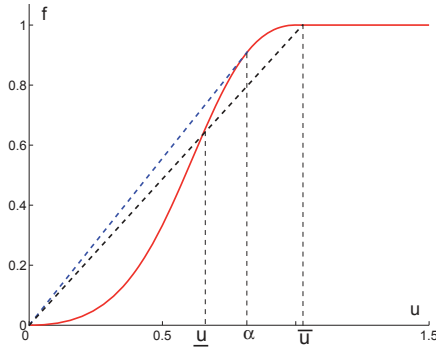


Figure 2 The function f for $p = 1$, $q = 1$ and $M = 2$; here α is the tangent point.

ains the same for any τ , the case $\tau > 0$ leads to important differences in the structure of the TW solutions and consequently in the admissibility of shocks. With (22), we seek v solving

$$\begin{cases} H(v)(s\tau v' - P(v))' = A - sv + f(v), \\ \quad \quad \quad \text{for } \eta \in \mathbb{R}, \\ \lim_{\eta \rightarrow -\infty} v(\eta) = u_\ell \text{ and } \lim_{\eta \rightarrow \infty} v(\eta) = u_r, \end{cases} \quad (26)$$

with s given by (23). Moreover, A has the same value, at least if H does not vanish; the *degenerate* case, when $H(u) = 0$ for e.g. $u = 0$ and $u = 1$, is analyzed in [3].

The existence of TW depends on τ . Here we consider $u_r < u_\ell$, and let α be the point where the tangent line through $(u_r, f(u_r))$ touches the graph of f (see Figure 2). In this context, the following results are proved in [4]:

Theorem. *If $u_r < \alpha \leq u_\ell$, there exists $\tau_* > 0$ such that:*

- If $0 \leq \tau \leq \tau_*$, TW connecting $u_\ell \leq \alpha$ to u_r exist and are monotone.*
- If $\tau > \tau_*$, there exists a unique $\bar{u}(\tau) > \alpha$ that can be connected to u_r through a monotone TW.*
- The $\bar{u} - \tau$ dependency is continuous and increasing for $\tau \geq \tau_*$.*

If $\tau > \tau_$, with $u_\ell = \bar{u}(\tau)$, let $\underline{u}(\tau) \in (u_r, \alpha)$ be the u -coordinate of the middle intersection point of the graph of f with the chord through $(u_\ell, f(u_\ell))$ and $(u_r, f(u_r))$. Then*

- For each $u_\ell \in (u_r, \underline{u}(\tau))$, there exists a TW connecting u_ℓ to u_r .*
- For each $u_\ell \in (\underline{u}(\tau), \bar{u}(\tau))$, no TW connecting u_ℓ to u_r exist.*
- For each $u_\ell \in (\underline{u}(\tau), \bar{u}(\tau))$, TW connecting u_ℓ to $\bar{u}(\tau)$ are possible.*

Given u_r , Figure 3 displays the values $\bar{u}(\tau)$ and $\underline{u}(\tau)$ for different values of τ . These points are computed by a shooting method (see [3]).

Case b of the theorem provides TW solutions connecting $u_\ell = \bar{u}(\tau) > \alpha$ to u_r . Letting $\varepsilon \rightarrow 0$, this TW justifies a shock solution $\{u_\ell, u_r\}$ to (12) that violates (25), which states that shocks $\{u_\ell, u_r\}$ are admissible only if $u_\ell \in (u_r, \alpha]$. If the choice of u_r , u_ℓ and τ places us the case f, the solution of the Riemann problem (12)–(13) combines two shocks: one upwards $\{u_\ell, \bar{u}(\tau)\}$ and one downwards $\{\bar{u}(\tau), u_r\}$. This is again a non-classical construction, where the solution is a shock down to u_r , possibly preceded by a rarefaction wave from u_ℓ to α if $u_\ell > \alpha$. This behaviour is displayed in Figure 4, presenting two numerical solutions computed for the regularized models with $\tau > \tau_*$, respectively $\tau = 0$. The former presents two fronts corresponding to the shocks mentioned before, the latter only one front connecting α to $u_r = 0$.

Saturation overshoot

The travelling wave analysis in the previous section provides the mathematical framework for explaining the occurrence of non-monotonic saturation profiles during infiltration. As observed in Figure 1, at low fluxes the saturation has a monotone profile. In this case, dynamic effects are negligible. The situation is changing as the flux is increasing; as mentioned before in a remark, a higher flux translates into a higher value of the parameter τ in the dimensionless setting. To relate the TW construction with the experiments, we mention that the left state u_ℓ represents the saturation at the inflow, whereas the right state u_r is nothing but the initial saturation (clearly, satisfying $u_r < u_\ell$). Then, as follows from the theorem, whenever

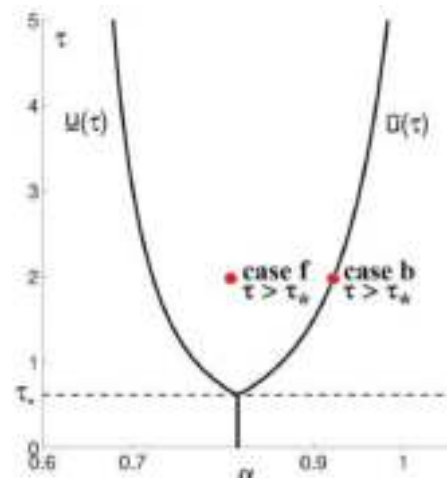


Figure 3 The branches $\bar{u}(\tau)$ (right) and $\underline{u}(\tau)$ (left) computed for $u_r = 0$. With $\tau > \tau_*$, in case b of the theorem $u_\ell = \bar{u}(\tau)$, while in case f $u_\ell \in (\underline{u}(\tau), \bar{u}(\tau))$.

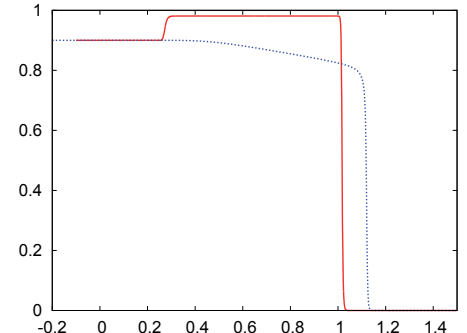


Figure 4 Numerical solutions for the full model (12), $\tau > \tau_*$ (solid) and $\tau = 0$ (dotted).

$\tau < \tau_*$, if $u_\ell < \alpha$, TW connecting u_ℓ to u_r are possible, also providing a monotonic saturation profile.

Increasing the flux at inflow has two implications in the dimensionless model: an increase of τ , as well as of u_ℓ . In this case, the theorem rules out the possibility of TW connecting u_ℓ directly to u_r . Instead, two fronts are encountered: an infiltration front from $\bar{u}(\tau)$ down to u_r and a drainage front from $\bar{u}(\tau)$ to u_ℓ . Between the two fronts, the saturation has a constant, *plateau* value $\bar{u}(\tau)$. These are also the saturation profiles in Figure 1, determined experimentally at high fluxes.

The present discussion is restricted to homogeneous media in one spatial dimension. Of course, realistic situations require two- or three-dimensional models, and heterogeneous media. One interesting phenomenon in the multi-dimensional case appears during infiltration in homogeneous media, when preferential flow is encountered in some parts of the medium. This leads to the development of *fingers*, which are thin regions of high saturation embedded into regions of low saturation. Such profiles are again ruled out by standard, equilibrium models, but permitted in the non-equilibrium case. Figure 5 presents such profiles, computed for the model (12) by a heterogeneous multiscale scheme (see [8]).

From practical point of view, the occurrence of preferential flow paths impacts the performance of the system. For example, in water driven oil recovery, when injecting water to displace oil, a water saturation overshoot implies that less oil is left in the reservoir. However, as seen above such overshoots are associated with the formation of fingers: inside fingers a high water saturation is encountered, whereas outside the saturation remains low. Water will prefer flowing through the fingers, leaving much oil unmobilized outside fingers. This reduces significantly the oil production. Similarly, when injecting supercritical CO_2 in the sub-surface, having fingers

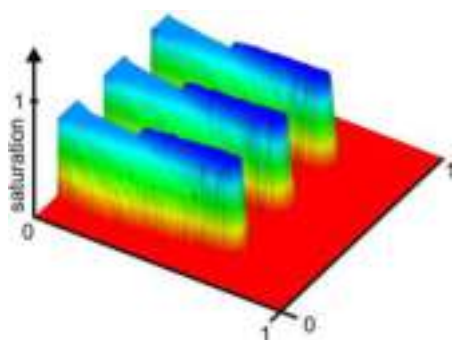


Figure 5 Fingers developed during infiltration in a heterogeneous medium (courtesy of F. Kissling). Note that even along a finger the saturation profile is non-monotone.

and preferential flow paths means that much of the storage reservoir is not reached by the CO_2 and therefore left unused. A proper management of such systems is based on the thorough understanding of the processes at various scales. This includes taking into account dynamic effects in controlling the flow regimes, so that the global efficiency is not depleted.

Conclusion

We discussed two mathematical models for two-phase flows in porous media. The difference between the two models is in the capil-

lary pressure-saturation dependency. In standard (equilibrium) models this is a nonlinear, monotone function, whereas non-equilibrium models also include the time derivative of the saturation. In both approaches, we let the capillary effects tend to zero to obtain admissibility conditions for shock solutions of the limit hyperbolic equation. These conditions depend on the parameter τ appearing in the dynamic capillarity term. We also discuss how non-equilibrium models can explain the saturation overshoot in the infiltration profiles, which are measured experimentally, but ruled out by the equilibrium theory. $\leftarrow$

References

- 1 J. Bear, *Dynamics of Fluids in Porous Media*, Dover, New York, 1972.
- 2 D.A. DiCarlo, Experimental measurements of saturation overshoot on infiltration, *Water Resour. Res.* 40 (2004), Wo4215.
- 3 C.J. van Duijn, Y. Fan, L.A. Peletier and I.S. Pop, Travelling wave solutions for degenerate pseudo-parabolic equation modelling two-phase flow in porous media, *Nonlinear Anal. Real World Appl.* 14 (2013), 1361–1383.
- 4 C.J. van Duijn, L.A. Peletier and I.S. Pop, A new class of entropy solutions of the Buckley-Leverett equation, *SIAM J. Math. Anal.* 39 (2007), 507–536.
- 5 A. Fasano, *Parabolic Free Boundary Problems in Industrial and Biological Applications*, *SIMA/e-Lecture Notes* 9 (2011).
- 6 S.M. Hassanizadeh and W.G. Gray, Thermodynamic basis of capillary pressure on porous media, *Water Resour. Res.* 29 (1993), 3389–3405.
- 7 R. Helmig, *Multiphase Flow and Transport Processes in the Subsurface*, Springer, 1997.
- 8 F. Kissling, R. Helmig and C. Rohde, Simulation of Infiltration Processes in the Unsaturated Zone Using a Multi-Scale Approach, *Vadose Z. Journal* 11 (2012), doi:10.2136/vzj2011.0193.
- 9 P.G. LeFloch, *Hyperbolic Systems of Conservation Laws. The Theory of Classical and Nonclassical Shock Waves*, *Lectures in Mathematics ETH Zürich*, Birkhäuser Verlag, Basel, 2002.
- 10 M. Lenzinger and B. Schweizer, Two-phase flow equations with outflow boundary conditions in the hydrophobic-hydrophilic case, *Nonlinear Anal. Real World Appl.* 73 (2010), 840–853.
- 11 S. Manthey, S.M. Hassanizadeh, R. Helmig and R. Hilfer, Dimensional analysis of two-phase flow including a rate-dependent capillary pressure-saturation relationship, *Adv. Water Res.* 31 (2008), 1137–1150.
- 12 J.M. Nordbotten, and M.A. Celia, *Geological Storage of CO_2* , Wiley, New Jersey, 2012.

Christoph Köhn

CWI, Amsterdam

c.koehn@cwi.nl

Margreet Nool

CWI, Amsterdam

margreet.nool@cwi.nl

Ute Ebert

CWI, Amsterdam

Faculteit Technische Natuurkunde, TU Eindhoven

ute.ebert@cwi.nl

Onweer: een bron van broeikasgassen en antimaterie

Bliksemschichten zijn indrukwekkende en wonderbaarlijke natuurverschijnselen. Ze zijn niet alleen gevaarlijk, maar ook een belangrijke bron van stikstofoxiden in de atmosfeer, een sterk broeikasgas. Sinds 1990 is bekend dat er naast bliksemschichten tussen en binnen wolken en naar de grond toe ook nog allerlei ontladingsvormen boven de wolken zijn, zoals sprites, elves en jets. Verder is waargenomen dat onweer flitsen van gammastraling kan uitzenden, en daarmee zelfs antimaterie kan produceren. Door hun toevallige optreden zijn rechtstreekse metingen aan bliksemschichten moeilijk uit te voeren. Ook kunnen ze maar zeer gedeeltelijk in het laboratorium nagebootst worden. Vandaar dat de wiskundige modellering van onweersverschijnselen uiterst belangrijk is. Christoph Köhn, Margreet Nool en Ute Ebert leggen uit hoe zij ontladingen simuleren met behulp van verschillende modellen waarin een dichtheidsbenadering en een deeltjesbeschrijving voor elektronen gecombineerd kunnen worden.

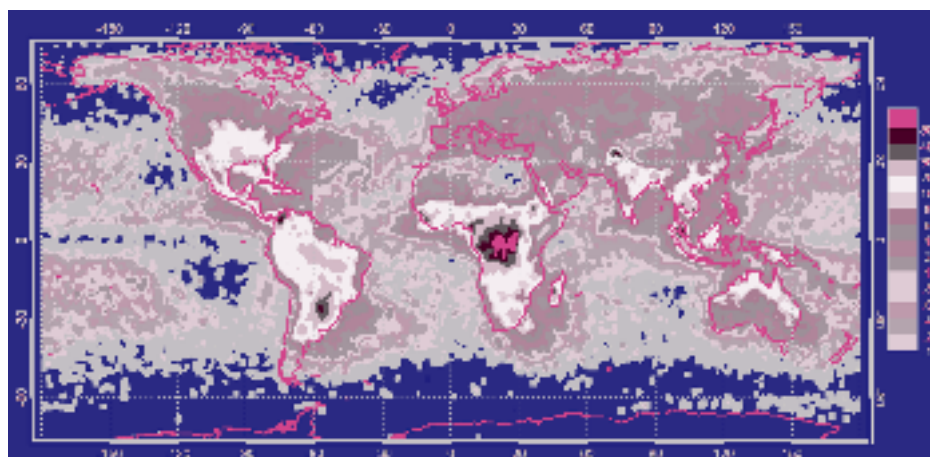
In 2010 werden door de NASA-satelliet Fermi [2] bundels van elektronen en positronen (hun antideeltjes) waargenomen boven Egypte. Hun bron was in tijd en ruimte gecorreleerd met een bliksemschicht boven Zambia, want geladen deeltjes volgen buiten de atmosfeer de lijnen van het aardmagnetisch veld. Het duidelijke positronensignaal op de satelliet was een opzienbarende ontdekking, omdat de rustenergie van een elektron of een positron 511 keV bedraagt en diensgevolge de productie van zulke paren een energie van meer dan 1 MeV vereist. Ter vergelijking: een elektron dat tussen twee plaalectroden wordt versneld met een spanning van 1 Volt, krijgt ten hoogste een energie van een elektronvolt (1 eV). Typische energiewaarden van elektronen in ontladingen zijn enkele eV, omdat de energie die nodig is om een molecuul in lucht te ioniseren — dat is de energie om een elektron uit de elektronenschil te stoten — tenminste 12 eV bedraagt voor zuurstof en tenminste 15 eV voor stikstof.

Om paren van elektronen en positronen te genereren, moeten fotonen (dat zijn licht-

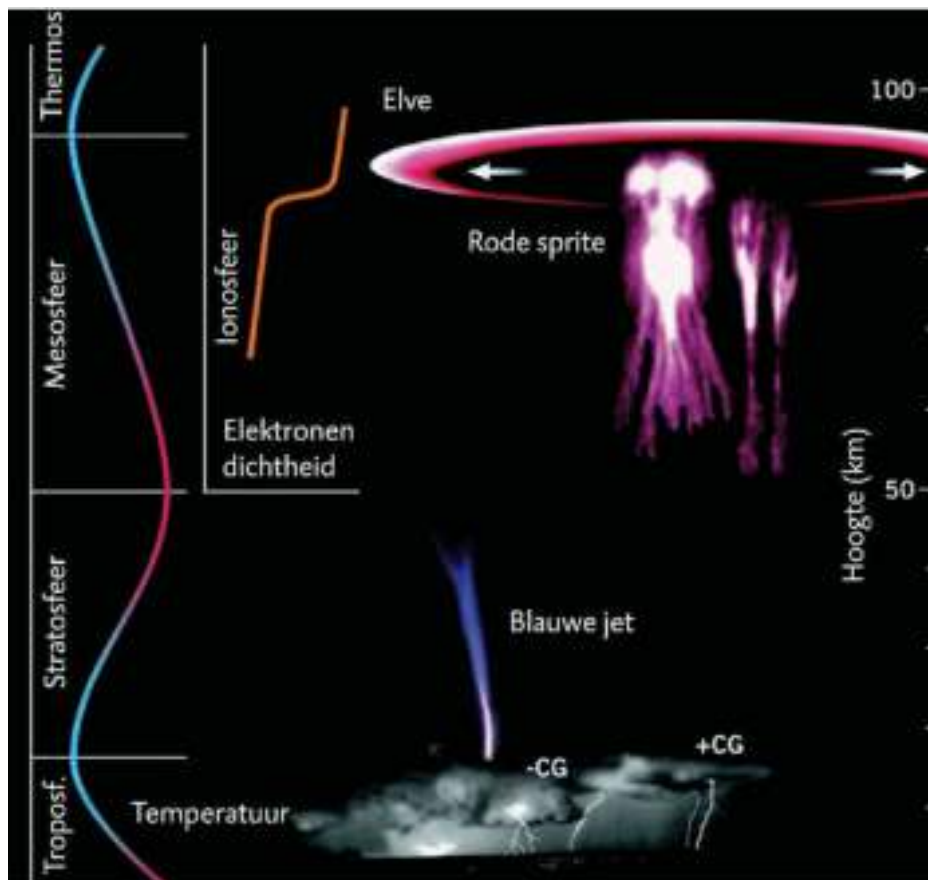
deeltjes oftewel quantumpakketjes van elektromagnetische straling) met energieën van enkele MeV met luchtmoleculen botsen. Fotonen in gammastralingsflitsen uit onweerswolken kunnen inderdaad zulke hoge en zelfs hogere energieën bereiken, zoals we sinds 1994 uit studies van *terrestrial gamma-ray flashes* (TGFs) [6] uit onweerswolken geleerd hebben.

Onweer is een belangrijk verschijnsel op onze aardbol. Daarvan kunnen we ons in Nederland geen goed beeld vormen, want dicht bij de evenaar kan de bliksemintensiteit wel meer dan honderd keer zo groot zijn als in ons land, zie Figuur 1. Ten eerste vormt onweer in tropische gebieden een sterke bron van stikstofoxide, dat een belangrijk broeikasgas is [12]. Ten tweede kunnen bliksemschichten schade aanrichten aan gebouwen, windmolens of vliegtuigen, en aan onze elektrische en elektronische apparatuur. Daarnaast kunnen zij zelfs dodelijk zijn.

Naast bliksems zijn er ook ontladingen, zoals *sprites* (reuzenbliksems, die uit de ionosfeer omlaag schieten), *elves* (uitdijende gloeiende ringen aan de onderkant van de ionosfeer) en *jets* (bliksems, die uit een wolk omhoog schieten), zie Figuur 2. Deze vormen van ontladingen zijn naar figuren uit *Een mid-*



Figuur 1 De verdeling van bliksems op aarde gemiddeld over een jaar. De zwarte gebieden langs de evenaar zijn gebieden waar bijzonder vaak bliksems worden gemeten [13].



Figuur 2 Verschillende soorten ontladingen boven onweerswolken tot een hoogte van 100 km: blauwe jets, rode sprites en als hoogste elves. Tot ongeveer 16 km komen de gewone bliksemontladingen (CG = cloud to ground = van wolk naar aarde) voor, die zowel positieve als negatieve ladingen van een wolk naar de aarde transporteren [11].

zomernachtsdroom van William Shakespeare genoemd. Dit type ontladingen vindt echter niet tussen een wolk en de aarde plaats, maar boven de wolken. Op een hoogte tussen 20

en 40 km kunnen blauwe jets worden geobserveerd. Tot een hoogte van 80 of 90 km komen er rode sprites voor en daarboven de zogenaamde elves. Niet alleen worden deze op een veel grotere hoogte dan bliksems waargenomen, zij zijn ook veel langer. Zo kan een sprite een lengte van 40 km hebben. Dit komt doordat de intrinsieke lengteschaal van een ontlading bepaald wordt door de afstand die een elektron kan afleggen tussen botsingen met luchtmoleculen, en die is ordes van grootte hoger in de zeer ijle lucht onder de ionosfeer. Sprites zijn in feite de eerste atmosferische ontladingen die gedeeltelijk door simulaties verklaard kunnen worden [5, 10].

De beweging van elektronen in lucht

Om een ontlading zoals een bliksem tot stand te brengen en in beweging te houden, moet er aan verschillende fysische omstandigheden voldaan zijn. Ten eerste zijn er vrije elektronen nodig die zich door een achtergrondgas voortplanten. In het geval van bliksem is dat lucht, maar in een laboratorium kunnen ook andere gassen worden gebruikt. Ook kennen we bliksems op andere planeten [3], bijvoorbeeld op Jupiter of Saturnus waar de atmosfeer uit waterstof en helium bestaat. Verder is

er een elektrisch veld nodig waarin de elektronen worden versneld. Figuur 3 demonstreert de beweging van elektronen in lucht, die afwisselend vrij bewegen en met luchtmoleculen botsen.

Tussen twee botsingen volgen vrije elektronen de bewegingsvergelijkingen in een elektrisch veld:

$$\frac{d\vec{p}}{dt} = e\vec{E}(\vec{r}), \quad \vec{p} = \frac{m_e \vec{v}}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad (1)$$

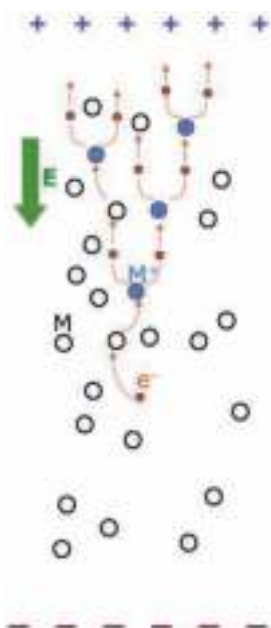
waar $\vec{p}$ de relativistische impuls, $\vec{v}$ de snelheid, m_e de massa en e de lading van een elektron zijn. $\vec{E}(\vec{r})$ is het elektrische veld waarin de elektronen worden versneld en c is de lichtsnelheid.

Botsingen zijn toevallige processen waarvan de waarschijnlijkheid wordt gegeven door

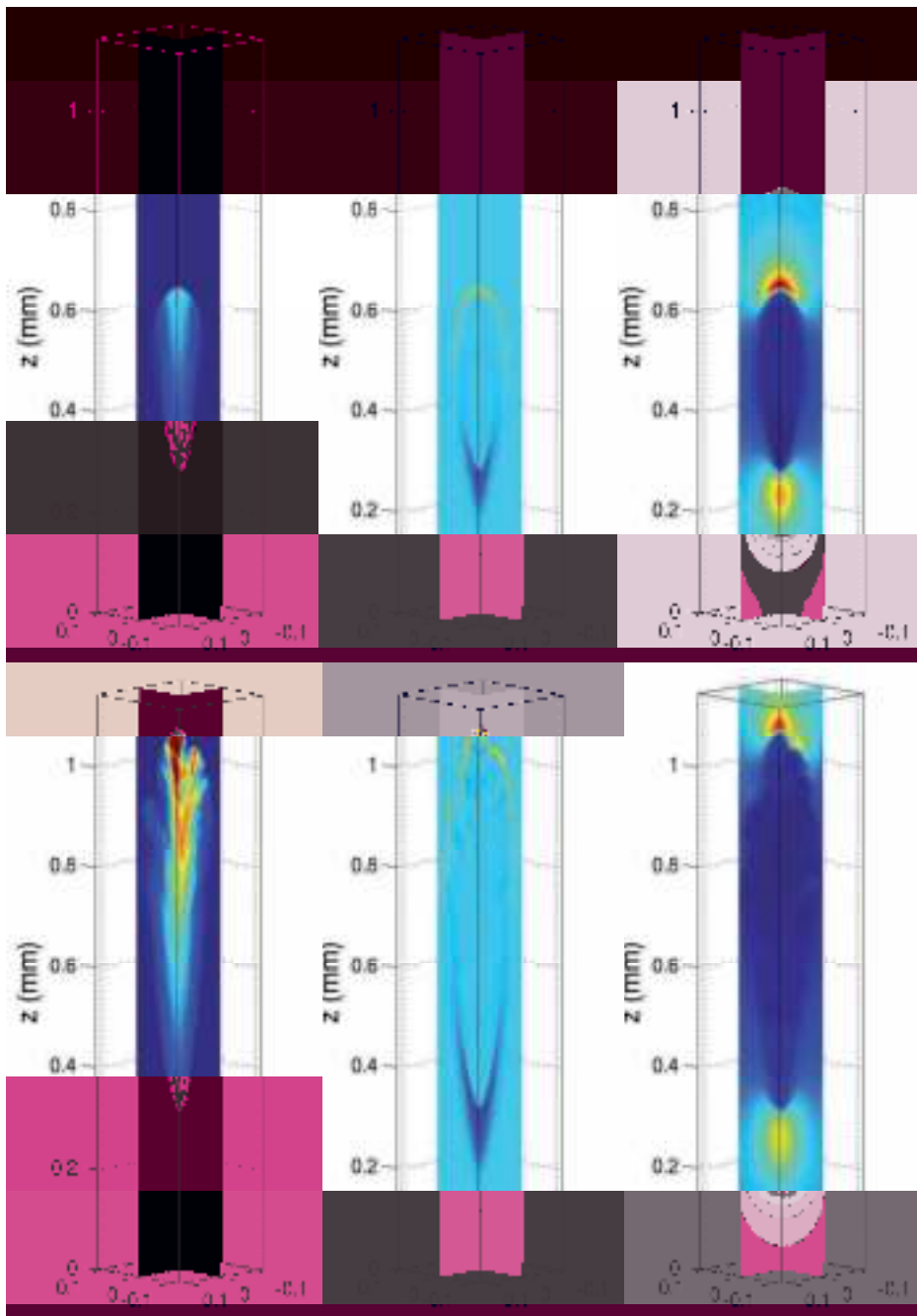
$$P = 1 - \exp(-\sigma_n(|\vec{v}|) \cdot |\vec{v}| \cdot n_B(z) \cdot \Delta t) \approx \sigma_n(|\vec{v}|) \cdot |\vec{v}| \cdot n_B(z) \cdot \Delta t. \quad (2)$$

Hierbij is $n_B(z)$ de dichtheid van het achtergrondgas op hoogte z en Δt het tijdsinterval waarbinnen de beweging van een elektron wordt bekeken. $\sigma_n(|\vec{v}|)$ is de werkzame doorsnede (*cross section*), dat is het botsingsoppervlak van twee deeltjes dat van de snelheid $\vec{v}$ van het elektron afhangt. n is een index voor de verschillende types botsingen die kunnen plaatsvinden. Een botsing kan volledig elastisch zijn, of het molecuul kan in verschillende toestanden aangeslagen worden, toestanden waarbij het molecuul trilt of roteert, of waarbij een elektron in een molecuul op een hogere energetische schil zit dan normaal. Hierdoor verliest het botsende elektron energie. Een heel belangrijk soort botsing is de ionisatie, waarbij een elektron door een inkomend elektron uit de elektronenschil van een molecuul naar buiten gestoten wordt en een positief ion achterlaat. Op deze manier groeit de hoeveelheid vrije elektronen. Daar staat een verliesmechanisme tegenover, waar een elektron aan een zuurstofmolecuul vast plakt. Als het elektrische veld hoog genoeg is (meer dan 3,2 MV/m in lucht op zeeniveau), neemt het aantal elektronen met de tijd toe en ontstaat er een elektronenwolk die zich door de lucht beweegt.

Deze beweging van elektronen door de lucht kan het beste door een Monte Carlo-deeltjesmodel worden beschreven. Dit model volgt vrije elektronen volgens (1), die aselect botsen met de luchtmoleculen die als toevallige achtergrond gemodelleerd worden.



Figuur 3 Als een elektron door gas beweegt, botst het vaak met de moleculen. Afhankelijk van de elektronenenergie vinden verschillende elastische, inelastische en ioniserende botsingen plaats.



Figuur 4 De linkerkolom toont de elektronendichtheid, de middelste kolom de negatieve ruimtelading en de rechterkolom het elektrische veld op de tijdstippen 0,72 ns en 0,90 ns. De resultaten zijn alle verkregen met het hybride model, een combinatie van het deeltjes- en dichtheidsmodel. Na ongeveer 0,46 ns is namelijk het aantal elektronen meer dan 2×10^7 en dan treedt het dichtheidsmodel in werking binnen de streamer. Aan de randen blijven de elektronen gevolgd worden met het deeltjesmodel [9].

Dit kan echter meer elektronen opleveren dan vandaag de dag gesimuleerd kunnen worden. Hoe hiermee om te gaan, bespreken we verderop.

Een belangrijk mechanisme is dat overal evenveel positieve als negatieve deeltjes ontstaan op grond van het behoud van lading, maar dat ze — voornamelijk door de beweging van de snelle en lichte elektronen — in een uitwendig elektrisch veld van elkaar gescheiden worden. Daardoor ontstaat er een ladingsver-

deling $\varrho(\vec{r}, t)$, die het elektrostatische veld $\vec{E}$ verandert volgens de Poisson-vergelijking

$$\Delta\phi(\vec{r}, t) = -\frac{\varrho(\vec{r}, t)}{\epsilon_0}, \quad \vec{E}(\vec{r}, t) = -\vec{\nabla}\phi(\vec{r}, t). \quad (3)$$

Streamers: de solitonen van de ontleding

Door de combinatie van de ionisatiereactie in het uitwendige veld en de Poisson-vergelijking ontstaan er onmiddellijk niet-lineaire structuren in een ontleding, die ook

wel *streamers* genoemd worden. Net als een soliton bewegen deze door de ruimte zonder dat hun ruimtelijke structuur veel verandert. Figuur 4 toont streamers die met behulp van een hybride model zijn gesimuleerd (zie verderop).

Streamers worden voortgestuwd door een uitwendig elektrisch veld, maar dat veld kan wel onder de kritische waarde van 3,2 MV/m liggen. Omdat het binnenste van de streamer voldoende sterk geïoniseerd is, schermt deze zich af van het uitwendige elektrische veld. Dit wordt veroorzaakt door een dunne schil van ruimtelading aan de buitenkant, die het veld voor de streamerkop sterk verhoogt. Hier krijgt het veld waarden tot hoog boven het kritische veld, en vindt botsingsionisatie plaats, waardoor de streamer groeit. Hier worden ook de moleculen aangeslagen, waardoor uiteindelijk broeikasgassen ontstaan zoals eerder genoemd.

Streamers vormen de allereerste stap van een gepulste ontleding, en ze effenen de weg voor een latere bliksemschicht.

Veelschalige ruimtelijke structuur

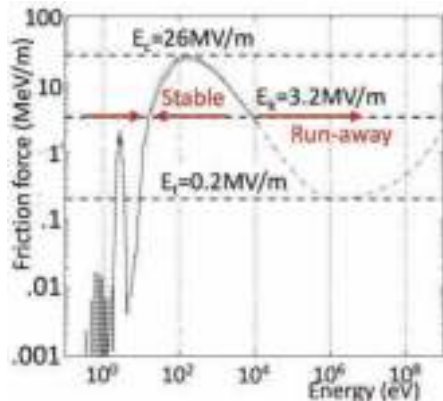
Een enkele streamer heeft al een veelschalige inwendige structuur, en vaak treden streamers in sterk vertakte boomstructuren op. We bespreken hier nog niet de strategie hoe de boomstructuur aan te pakken, maar beperken ons tot rekenmodellen voor enkele streamers.

Het hierboven besproken *deeltjesmodel* bootst direct de boven besproken elektrodynamica in lucht na. Wordt het aantal deeltjes echter te groot voor het computergeheugen, dan worden *superdeeltjes* geïntroduceerd, die uit meer dan één deeltje bestaan. Het rekenen met deze superdeeltjes kan echter tot gevolg hebben dat zich niet-fysische verschijnselen voordoen in de oplossing.

De traditionele aanpak van berekeningen bestaat uit een dichtheidsbenadering voor de elektronen, die echter fluctuaties van dichtheden en energieën van elektronen niet goed kan weergeven. Om de rekenaars efficiëntie van de dichtheidsbenadering te combineren met een deeltjesaanpak, hebben we een *hybride-in-ruimte*-model ontwikkeld.

Dichtheidsmodel

Dit model is een benadering van het deeltjesmodel, maar veel details zoals fluctuaties kunnen niet goed in dit model opgenomen worden. Het is bekend dat in een constant elektrisch veld de elektronenenergie varieert binnen de elektronenzwerm. Aan de voorkant



Figuur 5 De wrijvingskracht in lucht als functie van de elektronenenergie.

van de zwerm hebben de elektronen een hogere energie en zullen eerder ioniseren dan de elektronen aan de achterkant van de zwerm. Bovendien hebben deze laatste ook een lagere snelheid. De klassieke dichtheidsmodellen houden hier geen rekening mee en zullen minder nauwkeurige resultaten leveren. Vandaar dat wij het klassieke dichtheidsmodel hebben uitgebreid. Door de eerste term van de gradiëntexplosie in de elektronendichtheid op te nemen in de botsingsionisatiesnelheid, worden zowel de locatie van de deeltjeszwerm als de planaire ionisatie aan de voorzijde van de streamer goed benaderd.

Verder wordt gebruik gemaakt van adaptieve roosterverfijning, zodat in die gebieden waar de oplossing snel fluctueert, zoals aan de kop van de streamer, op een fijn numeriek rooster de simulatie berekend wordt, terwijl elders waar de oplossing glad is, op een veel grover rooster gerekend kan worden. Dit reduceert zowel de simulatietijd als het geheugengebruik ten opzichte van een fijn rooster in het gehele gebied.

In de nabije toekomst zullen de resultaten van onze collega's Dujko e.a. [4], die een hogere-orde-model voor streamers afgeleid hebben, toegepast gaan worden. Gebruikmakend van de impulsoverdrachtstheorie, worden de dichtheidsvergelijkingen verkregen als snelheidsmomenten van de Boltzmann-vergelijkingen. In het oude model worden twee behoudsvergelijkingen opgelost, namelijk die van de deeltjes- en de stroomdichtheid. In [4] wordt dit systeem uitgebreid naar vier vergelijkingen, namelijk met die voor energie en energiestroomdichtheid. Als de Boltzmann-theorie zonder benaderingen toegepast zou worden, dan zijn er oneindig veel vergelijkingen nodig. Hier beperkt men zich tot vier vergelijkingen, die een veel nauwkeurigere oplossing bieden ten opzichte van het oude model met slechts twee vergelijkingen.

Hybride model

Dit model koppelt het deeltjesmodel aan het dichtheidsmodel door middel van een (mee)bewegende modelinterface. Bij een streamerontlading bevinden de meeste elektronen zich juist in gebieden waar het elektrisch veld laag is, in het binnenste van de streamer. Dit gebied laat zich goed beschrijven door een dichtheidsmodel.

Relatief weinig elektronen bevinden zich aan de voorkant van de streamer of buiten de streamer. Precies deze elektronen moeten afzonderlijk gevolgd worden en niet samen genomen worden in superdeeltjes, zoals in de deeltjesmethode. Dit *hybride-in-ruimte*-model, in plaats van een willekeurige hybride methode, maakt gebruik van een bufferzone, een overgangsgedebied tussen deeltjes en dichtheden.

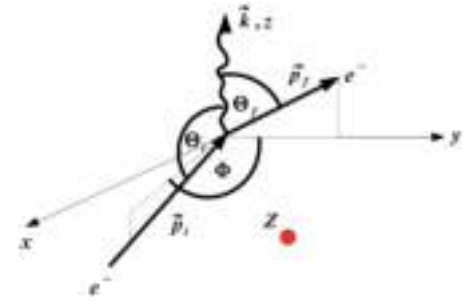
Wanneer de elektronenstroom over de interface tussen deeltjes- en dichtheidsmodel wordt berekend, moet dit fysisch consistent zijn. Daarom moeten aan beide kanten van de interface dezelfde *coëfficiënten* gebruikt worden. Gebeurt dit niet, dan kan er op de interface een discontinuïteit optreden in de elektronendichtheid. Een belangrijk voordeel van dit *hybride-in-ruimte*-model is dat het de beweeglijke deeltjes alleen volgt in de dynamisch meest interessante gebieden. De noodzaak om over te gaan tot superdeeltjes speelt dus veel later in de simulatie een rol.

Gamma- en antimaterieflitsen

In 1994 werden de eerste gammaflitsen gemeten vanuit een satelliet [6]. Volgens recente metingen kan de energie van enkele lichtdeeltjes in deze *Terrestrial Gamma-ray Flashes* (TGFs) tot 40 MeV of mogelijk zelf tot 100 MeV oplopen. Veel aardse gammaflitsen zijn gecorreleerd met sterke negatieve ontladingen, die binnen de wolk naar boven gaan. Zoals eerder vermeld, zijn twee jaar geleden ook bundels van positronen gemeten. De verklaring van deze verschijnselen moet gezocht worden in een uitbreiding van ontladingsmodellen naar hoge energieën. We bespreken nu belangrijke elementen ervan.

Weglopende elektronen

Door inelastische en ioniserende botsingen verliezen de elektronen voortdurend energie. De gemiddelde impulsverliezen kunnen door een wrijvingskracht $\vec{F}_w(\mathcal{E}, \vec{r})$ benaderd worden, die van de positie $\vec{r}$ en de energie $\mathcal{E}$ van de elektronen afhangt. De gemiddelde beweging van een elektron met een gemiddelde energie $\mathcal{E}$ wordt dan door de niet-lineaire



Figuur 6 Een elektron met een impuls $\vec{p}_i$ botst met een molecuul met atoomnummer Z en emitteert een foton met een impuls $\hbar\vec{k}$ zodat het elektron hierna een impuls $\vec{p}_f$ heeft. De hoeken Θ_i , Θ_f en Φ beschrijven de volledige geometrie van dit proces [8].

vergelijking

$$\frac{d\vec{p}}{dt} = e\vec{E}(\vec{r}) - \vec{F}_w(\mathcal{E}, \vec{r}) \quad (4)$$

benaderd, waar $e\vec{E}(\vec{r})$ de gemiddelde versnellingskracht in het elektrische veld is. Figuur 5 toont de wrijvingskracht in lucht als functie van de elektronenenergie [7]. Ook is de versnellingskracht in een veld van 3,2 MV/m als horizontale lijn aangeduid.

Bij elektronenenergieën onder 0,01 eV komen alleen elastische botsingen voor; dan is er geen wrijving en de elektronen kunnen alleen maar versneld worden. Boven 0,1 eV neemt de wrijving snel toe door de verschillende types inelastische botsingen, tot aan een maximum bij 200 eV. Als de elektronenenergie verder groeit, nemen de botsingsdoorsneden en dus de wrijving af (want de elektronen zijn bij wijze van spreken te snel voor een intensieve wisselwerking). De wrijving stijgt weer boven 1 MeV, waar relativistische effecten belangrijk worden, zoals de emissie van gammastralen.

Als de elektronen tegelijkertijd versneld worden door een veld van 3,2 MV/m, dan is de totale kracht (4) positief voor elektronen met energieën $\mathcal{E} < 10$ eV en $\mathcal{E} > 10$ keV en negatief daartussen. Dit betekent dat elektronen met $\mathcal{E} < 10$ eV gemiddeld versneld worden en dat elektronen met $10 \text{ eV} < \mathcal{E} < 10 \text{ keV}$ geremd worden, tot ze een energie van gemiddeld 10 eV bereikt hebben. Elektronen met een energie groter dan 10 keV daarentegen worden verder en verder versneld en lopen weg. Ze heten daarom *run-away* elektronen.

Ontladingsmodellen proberen dus te ontfaan of elektronen tot in het wegloopgebied versneld kunnen worden en welke elektronen welke energieën kunnen bereiken.

Remstraling, gammaflitsen en antimaterie

De weglopende energetische elektronen zijn

ook nu weer de oorzaak van de gammaflitsen. Als de elektronen met moleculen botsen, kunnen ze een deel van hun energie uitstralen, de zogenaamde remstraling. Remstraling bestaat uit fotonen, en als de elektronen energieën boven 1 MeV hebben, kunnen enkele fotonen zich ook binnen het energiegebied van gammastraling bevinden.

De effectieve doorsnede voor remstraling in lucht werd in 1934 berekend door Bethe en Heitler [1]. Zij leidden een uitdrukking af voor

$$\mathcal{H} := \frac{d^4\sigma}{d\omega d\Omega_i d\Omega_f d\phi}(E_{kin,i}, \omega, \Theta_i, \Theta_f, \phi), \quad (5)$$

die zowel van de energie van het inkomende elektron en van het uitlopende foton alsmede van de volledige geometrie afhangt (Figuur 6), waar de ruimtehoeken gedefinieerd worden door $d\Omega_{i,f} = \sin\Theta_{i,f} d\Theta_{i,f}$. De volledige expressie is echter te lang om hier uit te schrijven.

Door $\mathcal{H}$ te integreren over de hoeken die het uitlopende elektron parametriseren,

$$\begin{aligned} \frac{d^2\sigma}{d\omega d\Omega_i}(E_{kin,i}, \omega, \Theta_i) \\ = \int_{\phi=0}^{2\pi} \int_{\Theta_f=0}^{\pi} \mathcal{H} d\Omega_f d\phi, \end{aligned} \quad (6)$$

hebben wij een uitdrukking afgeleid die slechts afhangt van de energiewaarden en de hoek tussen de richting van het inkomende elektron en van het uitgezonden foton [8]. De integratie kan volledig analytisch in twee stappen worden uitgevoerd: De integralen over ϕ hebben de vorm $\int_{\phi=0}^{2\pi} d\phi a_{nm} \cos^{m-1} \phi / (\alpha \cos \phi + \beta)^n$, met $n, m \in \{1, 2\}$, waar de waarden van a_{nm} niet van ϕ afhangen. Dit soort integralen kan met behulp van de residustelling worden berekend. De resulterende integraal, die slechts van Θ_f afhangt, kan dan eenvoudig worden berekend uit de primitieve functie met de bijbehorende grenzen.

De integratie levert twee voordelen op: in de Monte Carlo-procedure zijn minder toevalsgetallen nodig, en we krijgen meer inzicht in de richting van de uitgezonden fotonen [8]. Zullen ze als een bundel een sa-

telliet bereiken of vanuit de wolk alle kanten op gaan? Onze analyse laat duidelijk zien dat elektronen met een energie boven 500 keV, fotonen overwegend in de voorwaartse richting uitzenden. Daaronder wordt de straling gelijkmatiger in alle richtingen geëmitteerd.

Als de ontlading naar boven loopt, zullen de zeer energetische remstraling-fotonen ook naar boven lopen. Onderweg kunnen ze weer een wisselwerking hebben met de lucht-moleculen. Fotonen met energieën boven 1 MeV kunnen daarbij een elektron-positron-paar genereren. Deze bundels van materie-antimaterie zijn in 2011 door de Fermi-satelliet waargenomen.

De vergelijkingen die de paarproductie beschrijven, tonen een fysische symmetrie met de vergelijkingen voor remstraling. Daarom konden we de analyse en integratie voor remstraling overdragen op paarproductie.

Hiermee hebben we enkele belangrijke fysische processen in ontladingen besproken. Er is echter nog veel werk nodig om een meer gedetailleerd beeld van de verschijnselen te ontwikkelen.

Referenties

- 1 H.A. Bethe en W. Heitler, On the stopping of fast particles and on the creation of positive electrons, *Proc. Phys. Soc. London* 146 (1934), pp. 83–112.
- 2 M.S. Briggs et al., Electron-positron beams from terrestrial lightning observed with Fermi GBM, *Geophys. Res. Lett.* 38 (2011), L02808.
- 3 D. Dubrovina et al., Sprite discharges on Venus and Jupiter-like planets: a laboratory investigation, *J. Geophys. Res.*, 115 (2010), A00E34.
- 4 S. Dujko, A.H. Markosyan, R.D. White en U. Ebert, 2013. High order fluid model for streamer discharges: I. Derivation of model and transport data, submitted to *J. Phys. D: Appl. Phys.*, preprint: <http://arxiv.org/abs/1302.4115>.
- 5 U. Ebert, De bliksem op het spoor, *Nederlands Tijdschrift voor Natuurkunde*, augustus 2009, pp. 300–305.
- 6 G.J. Fishman et al., Discovery of intense gamma-ray flashes of atmospheric origin, *Science* 264 (1994), pp. 1313–1316.
- 7 A.V. Gurevich, On the theory of runaway electrons. *Sov. Phys. JETP-USSR* 12 (1961), pp. 904–912.
- 8 C. Köhn en U. Ebert, 2012. Angular distribution of Bremsstrahlung photons and of positrons for calculations of terrestrial gamma-ray flashes and positron beams, te verschijnen in: *Atmospheric Research*, *ATMOS* 2874, preprint: arxiv.org/abs/1202.4879.
- 9 Chao Li, Jannis Teunissen, Margreet Nool, Willem Hundsdoerfer en Ute Ebert, A comparison of 3D particle, fluid and hybrid simulations for negative streamers, *Plasma Sources Sci. Technol.* 21 (2012), 055019(14pp).
- 10 A. Luque en U. Ebert, Emergence of sprite streamers from screening-ionization waves in the lower ionosphere, *Nature Geoscience* 2 (2009), pp. 757–760, Published online: 25 October 2009, doi:10.1038/ngeo662.
- 11 T. Neubert, On Sprites and their Exotic Kin, *Science* 300 (2003), pp. 741–749.
- 12 Vladimir A. Rakov en Martin A. Uman, *Lightning: Physics and Effects*, Cambridge University Press, 2007.
- 13 www.nasa.gov/vision/earth/lookingatearth/lightning_safety2005.html.

Chris Stolk

Korteweg-de Vries Instituut voor Wiskunde

Universiteit van Amsterdam

c.c.stolk@uva.nl

De wiskunde van reflectieseismiek

Voor de bestudering van de aardkorst wordt veelvuldig gebruik gemaakt van reflectieseismiek. Aan het aardoppervlak worden met behulp van akoestische bronnen golven opgewekt die zich voortplanten en gereflecteerd worden in de aardbodem. Trillingen ten gevolge van de gereflecteerde golven worden aan het oppervlak gemeten en geven informatie over de bodemstructuur. Het construeren van die structuur uit de gemeten reflecties kan beschouwd worden als een *invers probleem*. Chris Stolk gaat in op de wiskundige theorie en oplossingsmethodiek van dit inverse probleem.

Reflectieseismiek is de belangrijkste methode in de olie- en gasindustrie om fossiele brandstoffen op te sporen in de ondergrond. Varianten van de methode worden gebruikt in bodemonderzoek voor engineering toepassingen en in de academische geofysica voor de bestudering van de aardkorst en de mantel. Reflectieseismiek werkt met akoestische (seismische) golven. Deze planten zich over grote afstanden voort in de aarde en met de informatie die ze bevatten kunnen vaak beelden met hoge resolutie worden gemaakt. De

geschiedenis van het vakgebied gaat terug tot de jaren twintig van de vorige eeuw. De eerste olievondst mogelijk gemaakt door reflectieseismiek was in 1928 in de Verenigde Staten.

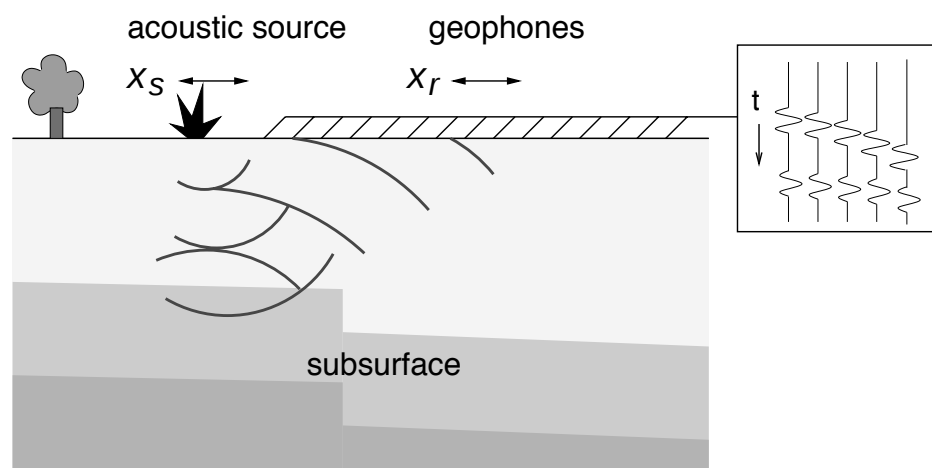
De opzet van een experiment is schematisch weergegeven in Figuur 1. Bronnen aan het aard- of zeeoppervlak genereren golven die zich in de aarde voortplanten en reflecteren. Aan het oppervlak worden de trillingen gemeten met een groot aantal speciale microfoons, geofoons of hydrofoons genaamd. Dit

experiment wordt vele malen herhaald, met bronnen op verschillende locaties.

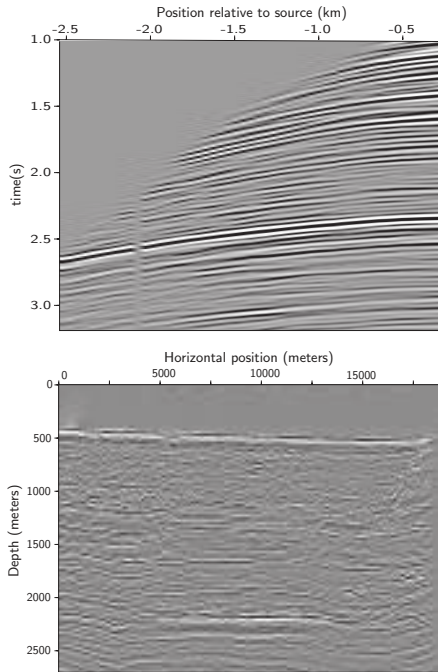
Golfvoortplanting wordt beschreven door tijdsafhankelijke partiële differentiaalvergelijkingen die lineair zijn aangezien de trillingen in de bodem een kleine uitslag hebben (behalve in de directe omgeving van de bron). De mechanische eigenschappen van het gesteente, zoals de dichtheid en compressibiliteit, vormen positieafhankelijke coëfficiënten in dit stelsel van partiële differentiaalvergelijkingen. De verwerking van reflectieseismische data kan dus worden gezien als een *invers probleem*. De data van dit inverse probleem zijn de metingen van oplossingen van partiële differentiaalvergelijkingen aan het oppervlak. De oplossing is een collectie van coëfficiëntfuncties, en de impliciete relatie tussen data en oplossing wordt gegeven door een stelsel partiële differentiaalvergelijkingen.

Een voorbeeld van seismische data en het resultaat van de dataverwerking zijn weergegeven in Figuur 2 (op de data is een filter toegepast om de gereflecteerde golven beter te laten zien). De volgende stap na de reconstructie van de coëfficiëntfuncties is seismische interpretatie, waarbij het doel is om de geologische structuur zo goed mogelijk te begrijpen.

In dit artikel bespreken we twee aspecten van het probleem. Het belangrijkste onderwerp is de wiskundige theorie voor imaging van enkelvoudig gereflecteerde golven. De theorie hiervan is grotendeels ontwikkeld in de laatste dertig jaar. Deze heeft geleid tot een veel beter begrip van (de prestaties van)



Figuur 1 Schematische weergave van een seismisch experiment op land



Figuur 2 Voorbeeld van seismische data, ingezoomd op een deel van de data van één experiment (boven); seismische afbeelding van de ondergrond ergens in de Golf van Mexico (onder). (Met dank aan The Rice Inversion Project).

de methodes, tot algoritmes om voorheen ontoegankelijke informatie uit de data te verkrijgen en tot nieuwe benaderingen van het probleem. Voordat we zover zijn, bespreken we de partiële differentiaalvergelijkingen met discontinue coëfficiënten die het voorwaartse probleem vormen. Goede referenties zijn het review artikel van Symes [14] voor de wiskundige aspecten, en boeken van Sheriff en Geldart [10] en Yilmaz [16].

Voorwaarts model

We zullen niet lang stilstaan bij de keuze van een model voor seismische golfvoortplanting, hoewel hier wel degelijk veel over te zeggen valt. Lineaire elastodynamica wordt algemeen beschouwd als een goed model. In tweede-orde-vorm zijn de vergelijkingen

$$\rho \frac{\partial^2 u_i}{\partial t^2} - \sum_{j,k,l} \frac{\partial}{\partial x_j} (C_{ijkl} \frac{\partial}{\partial x_l} u_k) = f_i. \quad (1)$$

Hierin is $u_i(x)$ een vectorveld dat de uitwijking (richting en grootte) van het medium beschrijft, $\rho(x)$ de dichtheid, $C_{ijkl}(x)$ de elastische tensor en $f_i(x)$ een vectorveld dat de krachtverdeling beschrijft die op het medium werkt [4].

Vergelijking (1) beschrijft golfvoortplanting in vaste stoffen, en is daarmee op het eerste gezicht het meest voor de hand liggende model. In de praktijk is ook de scalaire akoestische

golfvergelijking van groot belang, gegeven door

$$\frac{1}{\rho c^2} \frac{\partial^2 u}{\partial t^2} - \nabla \cdot \frac{1}{\rho} \nabla u = g. \quad (2)$$

Hierin is $u(x)$ de akoestische druk, $\rho(x)$ de massadichtheid, $c(x)$ de golfsnelheid en $g(x)$ een combinatie van brontermen. Net als (1) wordt deze vergelijking aangevuld met geschikte begin- en randvoorwaarden. Strikt genomen beschrijft (2) de voortplanting van akoestische golven in een vloeistof en niet in een vaste stof zoals de ondergrond van de aarde. Dat deze toch wordt gebruikt, komt omdat het essentiële gedrag van voortplantende golven al bevat is in de vergelijking en deze vaak een redelijke beschrijving van de longitudinale golven in een vaste stof geeft. Vanuit het oogpunt van inversie is het een logische keuze omdat het moeilijk is om meer parameters dan de functies c en ρ bevat in (2) te reconstrueren. Voor de verdere discussie beperken we ons daarom tot de scalaire akoestische vergelijking. Voor constante ρ en c heeft (2) de bekende uitgaande golfoplossingen

$$u(x, t) = \frac{w(t - r/c)}{4\pi r}, \quad r = |x - x_s|. \quad (3)$$

Uit metingen blijkt dat de eigenschappen van het aardmateriaal sterk variëren, zowel over korte als langere lengteschalen. De cartoonachtige afbeelding in Figuur 1 is niet juist: ook binnen de afgebeelde lagen treden bijvoorbeeld veel variaties op terwijl de laagovergangen vaak veel minder duidelijk zijn. De vraag is daarom hoe de golfvergelijking geïnterpreteerd moet worden in het geval van sterk variërende, bijvoorbeeld discontinue coëfficiënten.

In de klassieke, sterke formulering zijn beide kanten van de partiële differentiaalvergelijking continue functies, en geldt de vergelijking voor alle (x, t) in het domein, noem dit $\Omega \times [0, T]$ met $\Omega \subset \mathbb{R}^3$. Dit is niet mogelijk met discontinue coëfficiënten. De oplossing is om de vergelijking in de *zwakke vorm* te schrijven. We eisen dat

$$\iint \left[\frac{1}{\rho c^2} \frac{\partial^2 u}{\partial t^2} - \nabla \cdot \frac{1}{\rho} \nabla u - g \right] \cdot v(x, t) dx dt = 0 \quad (4)$$

voor alle $v \in C_0^\infty(\Omega \times [0, T])$. Na partiële integratie wordt dit

$$\iint \left[\frac{1}{\rho c^2} \frac{\partial u}{\partial t} \frac{\partial v}{\partial t}(x, t) - \frac{1}{\rho} \nabla v \cdot \nabla u + g v \right] dx dt = 0, \quad (5)$$

voor alle v in $C_0^\infty(\Omega \times [0, T])$. Deze vergelijking heeft zin als u , ∇u , $\frac{\partial u}{\partial t}$ en g kwadratisch integreerbaar zijn, en ρ en c meetbaar en begrensd. Voor ρ en c eisen we bovendien dat ze van 0 weg begrensd zijn, dus

$$\begin{aligned} \rho, c &\in L^\infty(\Omega), \quad 0 < \alpha < \rho(x) < \beta \\ \text{en } 0 < \alpha < c(x) < \beta. \end{aligned} \quad (6)$$

De goedgesteldheid van het probleem is behandeld in [7], met verdere ontwikkelingen in [3, 11], met als resultaten onder andere:

- Existentie en eenduidigheid van oplossingen en continue afhankelijkheid van de coëfficiënten.
- Differentieerbare afhankelijkheid van de coëfficiënten met een verlies van Sobolev-regulariteit van een orde. Deze informatie is van belang voor de keuze van optimalisatiemethodes in het inverse probleem.

Het hierboven geschetste kader is nog niet direct toepasbaar. We nemen namelijk aan dat de bron g van de vorm

$$g(x, t) = \delta(x - x_s) w(t) \quad (7)$$

is, waarin δ de Dirac-deltafunctie is, x_s de positie van de bron en $w(t)$ een korte, smalle puls, zonder zeer lage frequenties. Dit is geen kwadratisch integreerbare functie. Ook geldt dat de oplossingen bevat zijn in ruimtes van kwadratisch integreerbare functies en dus niet zonder meer continu zijn in x , wat nodig is om puntmetingen te doen. Met de extra aanname dat c en ρ extra regulariteit hebben rond de positie van de bron en de ontvangers, bijvoorbeeld lokaal constant zijn, kan ook dit geval worden behandeld [14].

De verzameling van paren van bron- en ontvangerposities (x_s, x_r) waarvoor metingen worden gedaan, noteren we met Y . Deze liggen aan het oppervlak, zodat $Y \subset \mathbb{R}^2 \times \mathbb{R}^2$. Metingen worden gedaan gedurende een tijdsinterval $[0, T]$. De *voorwaartse afbeelding* is dan de afbeelding van de functies c en ρ naar de waarden van de oplossing u van (2) in de meetpunten, genoteerd met

$$\begin{aligned} \mathcal{A}[c, \rho](x_r, t; x_s) &= u(x_r, t; x_s), \\ (x_r, x_s) &\in Y, \quad 0 \leq t \leq T. \end{aligned} \quad (8)$$

Lineaire inverse verstrooiing

De huidige techniek van imaging van seismische data is bijna uitsluitend gebaseerd op het gebruik van de linearisatie van de operator $\mathcal{F}$ in c . Een fysische omschrijving is dat gebruik gemaakt wordt van enkelvoudig gereflecteerde golven. De betekenis van ‘enkelvoudig gereflecteerd’ is in dit geval echter niet zo duidelijk: We kennen allemaal het principe van reflectie van licht in bijvoorbeeld een ruit. Daarbij is het duidelijk wat de gereflecteerde golf is, en dat de reflectie precies aan de overgang tussen lucht en glas optreedt. In een variabel medium kunnen reflectie en refractie echter overal optreden.

Zoals gezegd lineariseren we $\mathcal{F}$ in c , waarbij we aannemen dat ρ een gegeven constante is. Dat betekent dat we een kleine verstoring $\delta c(x)$ toevoegen aan de snelheid $c(x)$, en vervolgens een vergelijking voor de resulterende verstoring δu van u afleiden. Het blijkt eenvoudig te zijn om te werken met de relatieve verstoring $r = \delta c/c$, het achtergrondmodel c noteren we ook wel met v . Dit resulteert in de volgende vergelijking voor de perturbatie δu :

$$\left(\frac{1}{v^2} \frac{\partial^2}{\partial t^2} - \nabla^2 \right) \delta u = \frac{2r}{v^2} \frac{\partial^2 u}{\partial t^2}, \quad (9)$$

$$\delta u = 0 \quad \text{voor } t < 0.$$

Deze vergelijking is lineair in r en de gelineariseerde voorwaartse afbeelding wordt gedefinieerd door

$$F[v]r = \delta u|_{Y \times [0, T]}. \quad (10)$$

De reconstructie van r gegeven δu wordt het gelineariseerde inverse probleem genoemd. Daarbij is de functie v , het zogenaamde snelheidsmodel, gegeven. Uiteindelijk moet ook v bepaald worden. Ook wordt ervan uitgegaan dat u en δu kunnen worden gescheiden in de gemeten data. In de praktijk zijn de niet-gereflecteerde golven te herkennen aan de hoek waaronder ze aankomen aan het oppervlak en aan de aankomsttijd.

Een belangrijke aanname in het model (9) is dat v en r verschillende delen van het snelheidsveld c beschrijven: v is een gladde functie en bevat alleen laag-frequente componenten, en r bevat alleen hoog-frequente, oscillerende componenten (scheiding van schalen). Met de scheiding van schalen geeft de linearisatie een beschrijving van de data die opvallend veel beter is dan wanneer r ook een laag-frequent deel heeft, zo blijkt uit numerieke berekeningen [14].

Vergelijking (9) voor δu kunnen we interpreteren met het principe van Huygens. Aan de rechterkant staat een bronterm die het product van r en $\frac{\partial^2 u}{\partial t^2}$ bevat. Een bron van golven ontstaat dus voor (x, t) waar r en $\frac{\partial^2 u}{\partial t^2}$ allebei ‘groot’ zijn. Dit is in overeenstemming met het principe van Huygens, dat zegt dat een secundaire bron van golven ontstaat wanneer het golfvront u een reflector bereikt. (Inderdaad is $r(x)$ groot waar de variatie in het medium groot is en het is bekend dat met name daar de reflecties optreden.)

Wiskundigen legden de link tussen seismische imaging algoritmes en het gelineariseerde inverse probleem eind jaren zeventig en in de jaren tachtig. Een belangrijk artikel was dat van G. Beylkin [1], met bijdragen op twee vlakken:

- De precisering in welke zin een inverse afbeelding van F wordt geconstrueerd, gegeven het feit dat het geen exacte inverse is.
- De constructie van een dergelijke inverse in de vorm van een gegeneraliseerde Radon-transformatie.

Het blijkt dat de geconstrueerde inverse een reconstructie geeft van de *discontinuïteiten* in r .

We geven de belangrijkste ingrediënten van dit resultaat. De afleiding van de voorwaartse afbeelding maakt gebruik van de *functie van Green*, de oplossing $G(x, t; x_s)$ van

$$\left(\frac{1}{v^2} \frac{\partial^2}{\partial t^2} - \nabla^2 \right) G(x, t; x_s) = \delta(t) \delta(x - x_s). \quad (11)$$

Met een geometrische optica-aanname geldt

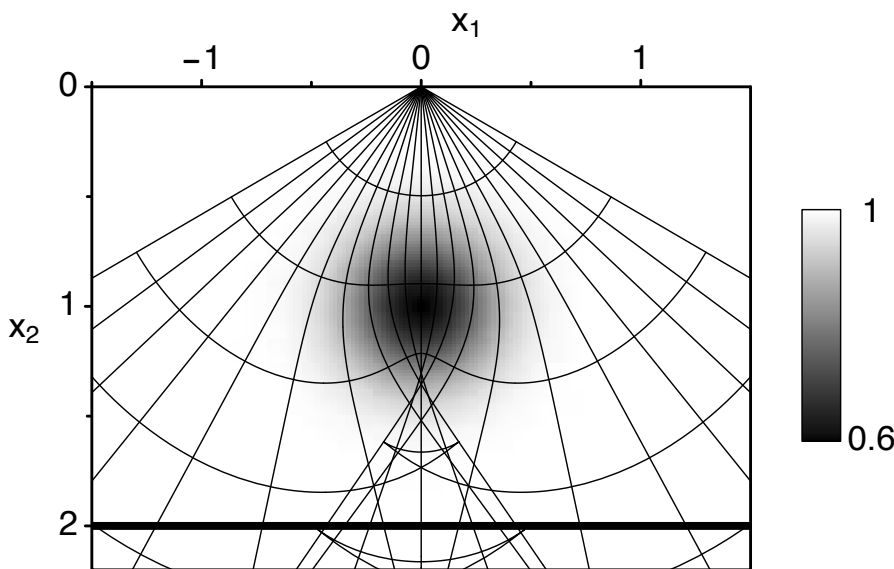
$$G(x, t; x_s) \simeq a(x; x_s) \delta(t - \tau(x, x_s)). \quad (12)$$

Hierin is $\tau(x, x_s)$ de reistijd functie. Voor constante v wordt deze gegeven door $\frac{|x - x_s|}{v}$, in het algemene geval door een oplossing van de eikonaalvergelijking. De amplitude a voldoet aan de zogenaamde transportvergelijking. Beide zijn gladde functies weg van x_s en kunnen op diverse manieren numeriek worden bepaald. Het symbool $\simeq$ in (12) betekent ‘modulo een gladdere fout’ en de vergelijking is geldig voor kleine tijd, waarbij de aanname is dat er voor elk paar (x, x_s) er precies één ‘lichtstraal’ is die de twee punten verbindt. De situatie dat de gekromde stralen focuseren, en dat meerdere golfvronten ontstaan, zie Figuur 3, wordt daarmee uitgesloten.

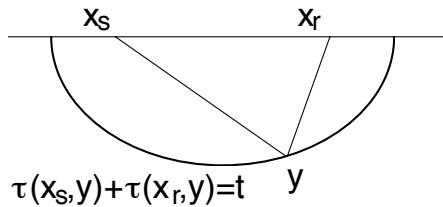
Gebruikmakend van (12) vinden we

$$F[v]r(x, t; x_s) \simeq \frac{\partial^2}{\partial t^2} \int dx \frac{2r(x)}{v(x)^2} a(x, x_r) a(x, x_s) \cdot \delta(t - \tau(x, x_r) - \tau(x, x_s)). \quad (13)$$

Dit is de zogenaamde gegeneraliseerde Radon-transformatie: een integraal over gekromde oppervlakken (of tweedimensionale krommes) gegeven door de vergelijking $\tau(x, x_r) + \tau(x, x_s) = t$, de zogenaamde isochronen, zie Figuur 4.



Figuur 3 Lichtstralen uit de oorsprong in een medium met een lage-snelheidszone rond $(0, 1)$. Buiging van de stralen leidt tot focuseren van de stralen rond $(0, 1.3)$ en tot meerdere golfvronten daaronder. Hiermee wordt de geometrische optica-aanname geschonden.



Figuur 4 Isochronen: Integratie-oppevlakken in de gegeneraliseerde Radon-transformatie.

In [1] werd een inverse van (13) gegeven. De precieze beschrijving in welke zin de geconstrueerde operator een inverse is, vereist kennis van de *microlokale analyse*, een verzameling technieken die voortbouwt op distributietheorie en wordt gebruikt in de studie van lineaire partiële differentiaaloperatoren [5–6]. Belangrijk hierbij is dat singulariteiten een positie en een richting hebben. Voor een discontinuïteit die optreedt langs een oppervlak wordt de ermee geassocieerde richting gegeven door de normaal op het oppervlak. In het algemeen kan deze bepaald worden door te bestuderen in welke richtingen de Fourier-getransformeerde al dan niet snel afneemt. Voor een distributie u wordt de verzameling van paren (positie, richting) in $\mathbb{R}^n \times \mathbb{R}^n$ waar singulariteiten optreden, de golffrontverzameling $WF(u)$ genoemd.

Singulariteiten in r leiden tot singuliere bijdrages in het gereflecteerde golfveld δu . Echter, niet alle gereflecteerde golven worden waargenomen aan het oppervlak. Alleen het deel van de singulariteiten in r waarvan de reflecties worden waargenomen, kan worden gereconstrueerd. Dit leidt tot een *microlokale*

reconstructie. Zij Γ de verzameling van singulariteiten die we op grond hiervan kunnen reconstrueren. Beylkin liet zien dat er een functie $b(x_r, t; x_s)$ is zo dat

$$F[v]^\dagger d(x) = \iiint dx_r dt b(x_r, t; x_s) \cdot \delta(t - \tau(x, x_s) - \tau(x, x_r)) d(x_r, t; x_s) \quad (14)$$

geeft dat

$$F[v]^\dagger F[v]r \approx r \quad (15)$$

als $WF(r)$ bevat is in Γ .

Het bovenstaande resultaat betreft imaging algoritmes die gebruik maken van de gegeneraliseerde Radon-transformatie in de situatie dat $\dim Y = 2$, en heeft als beperking de aanname bij formule (12). In een groot aantal artikelen is het gebruik van gelineariseerde inversie verder onderzocht. Dat betrof bijvoorbeeld inversie op basis van het elastische model, situaties met minder sterke aannames op de stralen, met andere bronontvangerverzamelingen ($\dim Y = 3$ of 4), en over speciale imaging algoritmes die bijvoorbeeld hoekafhankelijke reflectiecoëfficiënten proberen te reconstrueren, zie bijvoorbeeld [2, 9] en verder de referenties in [14].

Slotopmerkingen

We sluiten af met een korte discussie van twee aspecten die nog weinig aan bod zijn gekomen: De bepaling van v en niet-lineaire

kleinste-kwadrateninversie. In de laatste methode is het doel om de coëfficiënten zo te kiezen dat het verschil tussen gemodelleerde data en gemeten data minimaal wordt in de L^2 -norm [15]. Gezien de hoge dimensies van de ruimtes van data en modelparameters is dit een kostbare methode. Echter, het grootste probleem is dat het minimum niet makkelijk te vinden is omdat de functie die geminimaliseerd moet worden veel lokale minima heeft, zie [8] voor een benadering die dit probleem soms oplost.

Voor de bepaling van het snelheidsmodel zijn met name de aankomsttijden van de gereflecteerde golven van belang. Een discontinuïteit in het medium kan vaak worden waargenomen vanuit meerdere bronontvangervercombinaties, deze redundantie maakt het mogelijk om v te reconstrueren. Voor gelaagde media kan de zogenaamde Dix-formule worden gebruikt, in het algemeen is dit een interactief proces waarbij verschillende methodes in combinatie gebruikt worden. *Differential semblance optimization* (DSO) is een geautomatiseerde methode hiervoor. Het sterke punt hiervan is dat de te minimaliseren functie zich veel beter gedraagt dan in kleinste-kwadraten inversie. Naar deze methode, bedacht door W.W. Symes, is daarom veel onderzoek verricht. De methode lijkt echter gehinderd te worden door de linearisatiefout. De vraag is daarom of er methodes zijn die het grote aantrekkingsgebied van DSO behouden en ook de niet-lineaire effecten aankunnen, zie [12–13].

Referenties

- 1 G. Beylkin, Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized Radon transform, *J. Math. Phys.* 26(1) (1985), 99–108.
- 2 G. Beylkin en R. Burrige, Linearized inverse scattering problems in acoustics and elasticity, *Wave Motion* 12(1) (1990), 15–52.
- 3 K. D. Blazek, C. C. Stolk en W. W. Symes, A mathematical framework for inverse wave problems in heterogeneous media, *Inverse Problems* 29(6) (2013), 065001.
- 4 C.H. Chapman, *Fundamentals of seismic wave propagation*, Cambridge University Press, UK, New York, 2004.
- 5 J.J. Duistermaat, *Fourier integral operators*, Progress in Mathematics 130, Birkhäuser, Boston, MA, 1996.
- 6 L. Hörmander, *The analysis of linear partial differential operators. Vol. I–IV*, Springer, Berlin, 1983.
- 7 J.-L. Lions en E. Magenes, *Non-homogeneous boundary value problems and applications. Vol. I*, Springer, New York, 1972. Translated from the French by P. Kenneth, *Die Grundlehren der Mathematischen Wissenschaften*, Band 181.
- 8 R. Pratt, Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model, *Geophysics* 64(3) (1999), 888–901.
- 9 Rakesh, A linearised inverse problem for the wave equation, *Comm. Partial Differential Equations* 13(5) (1988), 573–601.
- 10 R.E. Sheriff en L.P. Geldart, *Exploration Seismology*, Cambridge University Press, second edition, 1995.
- 11 C.C. Stolk, On the modeling and inversion of seismic data, *Utrecht University*, PhD thesis, December 2000.
- 12 C.C. Stolk en W.W. Symes, Smooth objective functionals for seismic velocity inversion, *Inverse Problems* 19(1) (2003), 73–89.
- 13 W.W. Symes, Migration velocity analysis and waveform inversion, *Geophys. Prospect.* 56 (2008), 654–663.
- 14 W.W. Symes, The seismic reflection inverse problem, *Inverse Problems* 25(12) (2009), 123008, 39.
- 15 A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2005.
- 16 O. Yilmaz, *Seismic Data Analysis: Processing, Inversion, And Interpretation Of Seismic Data*, Society of Exploration Geophysicists, 2001.

Problemen

| Problem Section

This Problem Section is open to everyone; everybody is encouraged to send in solutions and propose problems. Group contributions are welcome.

For each problem, the most elegant correct solution will be rewarded with a book token worth €20. At times there will be a Star Problem, to which the proposer does not know any solution. For the first correct solution sent in within one year there is a prize of €100.

When proposing a problem, please either include a complete solution or indicate that it is intended as a Star Problem. Electronic submissions of problems and solutions are preferred (problems@nieuwarchief.nl).

The deadline for solutions to the problems in this edition is 1 December 2013.

Problem A (proposed by Hendrik Lenstra)

Let $a, b \in \mathbb{C}$. Show that if there exists an irreducible polynomial $f \in \mathbb{Q}[X]$ such that $f(a) = f(a+b) = f(a+2b) = 0$, then $b = 0$.

Problem B (The attribution will appear in the March issue of 2014)

Let n be a positive integer, and let e_{ij} be an integer for all $1 \leq j \leq i \leq n$. Show that there exists an $n \times n$ -matrix with integer entries such that the eigenvalues of the top left $i \times i$ -minor are $e_{i1}, \dots, e_{ii}$ (with multiplicity).

Problem C (proposed by Bart de Smit and Hendrik Lenstra)

Let A be a finite commutative unital ring. Does there exist a pair (B, f) with B a finite commutative unital ring in which every ideal is principal, and f an injective ring homomorphism $A \rightarrow B$?

Edition 2013-1 We received solutions from Wouter Cames van Batenburg (Leiden), Hao Chen (Seattle), Charles Delorme (Paris), Florian Eisele (Brussels), Alex Heinis (Hoofddorp), Pieter de Groen (Brussels), Thijmen Krebs (Nootdorp), Guido Senden (Groningen), Sep Thijssen (Nijmegen) and Traian Viteam (Punta Arenas).

Problem 2013-1/A Consider a regular n -gon $P_1P_2 \dots P_n$, and draw $n-3$ diagonals such that there are no intersection points in the interior. The polygon is now divided into $n-2$ triangles. Let t_i be the number of such triangles that have a vertex at P_i . Show that

$$t_1 - \frac{1}{t_2 - \frac{1}{\dots - \frac{1}{t_{n-1}}}} = 0.$$

Solution We received solutions from Charles Delorme, Alex Heinis, Pieter de Groen, Guido Senden, Sep Thijssen, Traian Viteam and Thijmen Krebs. The book token goes to Sep Thijssen. All solutions followed the same general strategy. The following solution is most similar to that of Charles Delorme, Pieter de Groen and Sep Thijssen.

We use $[t_1, \dots, t_{n-1}]$ to denote the continued fraction from the problem statement, and we will prove the identity by induction for all convex (not necessarily regular) n -gons with $n \geq 3$. Note that the base case $n = 3$ is trivial as $1 - \frac{1}{1} = 0$.

Now assume $n \geq 4$ and assume that the identity holds for $n-1$. The induction step works by finding a triangle with two exterior edges, that is, a triangle $P_{k-1}P_kP_{k+1}$ with $t_k = 1$, and by removing that triangle from the n -gon, which yields an $(n-1)$ -gon.

Redactie:

Johan Bosman

Gabriele Dalla Torre

Christophe Debry

Jinbi Jin

Marco Streng

Wouter Zomervrucht

Problemenrubriek NAW

Mathematisch Instituut

Universiteit Leiden

Postbus 9512

2300 RA Leiden

problems@nieuwarchief.nl

www.nieuwarchief.nl/problems

Lemma 1. *If $n \geq 4$, then there exists an integer k with $2 \leq k \leq n-1$ and $t_k = 1$.*

Proof. We prove first that there are at least two integers k with $1 \leq k \leq n$ with $t_k = 1$. Suppose that there is at most one such integer, that is, suppose that there is at most one triangle with two exterior edges. The total number of exterior edges n then is at most $2 + (\# \text{triangles} - 1) = n-1$, which is a contradiction.

Now note that no two such integers k can correspond to adjacent vertices, so that at least one of them satisfies $2 \leq k \leq n-1$. $\square$

Now let k be as in the lemma and assume for now $k \neq n-1$. Consider the $(n-1)$ -gon with the vertex P_k and the edges $P_{k-1}P_k$ and P_kP_{k+1} removed. For that $(n-1)$ -gon, the triangle count t_k is omitted, and the adjacent numbers t_{k-1} and t_{k+1} are lowered by 1. In other words, the induction hypothesis yields

$$[t_1, \dots, t_{k-2}, t_{k-1} - 1, t_{k+1} - 1, t_{k+2}, \dots, t_{n-1}] = 0.$$

In particular, the proof of the case $k \neq n-1$ is finished once we prove the following key identity.

Lemma 2. *We have*

$$\begin{aligned} & [t_1, \dots, t_{k-2}, t_{k-1}, 1, t_{k+1}, t_{k+2}, \dots, t_{n-1}] \\ &= [t_1, \dots, t_{k-2}, t_{k-1} - 1, t_{k+1} - 1, t_{k+2}, \dots, t_{n-1}]. \end{aligned}$$

Proof. From the definition, we directly get

$$[a_1, \dots, a_m, [b_1, \dots, b_n]] = [a_1, \dots, a_m, b_1, \dots, b_n]. \quad (1)$$

Moreover, it is a simple matter of writing things out to see

$$[a, 1, b] = [a - 1, b - 1]. \quad (2)$$

Indeed, we have

$$a - \frac{1}{1 - \frac{1}{b}} = a - \frac{b}{b-1} = a - 1 - \frac{1}{b-1}.$$

The result follows by taking $a = t_{k-1}$ and $b = [t_{k+1}, t_{k+2}, \dots, t_{n-1}]$ in (2) and applying (1) repeatedly. $\square$

This leaves only the case $k = n-1$, for which we use the identity

$$[t_1, \dots, t_{k-2}, t_{k-1}, 1] = [t_1, \dots, t_{k-2}, t_{k-1} - 1],$$

which is obvious as the tail reads $t_{k-1} - \frac{1}{1} = t_{k-1} - 1$.

Remark. One might worry about division by zero. However, if one takes $1/0 = \infty$ and $1/\infty = 0$, then all the identities in the proof above continue to make sense even with divisions by zero. Alternatively, it is possible to prove that division by zero does not occur. This can be done by showing in the induction $[t_i, \dots, t_{n-1}] \geq 0$ for $i = 1, \dots, n-1$ with equality if and only if $i = 1$, which requires a few extra case distinctions in the induction step.

Remark. If one does the induction with $k = 1$ or $k = n$, then the induction step becomes more complicated. Fortunately we did not need to do this thanks to Lemma 1. For example, with $k = 1$, one needs to prove the implication

$$[t_2 - 1, t_3, \dots, t_{n-1}] = 0 \implies [1, t_2, t_3, \dots, t_{n-1}] = 0.$$

Let $b = [t_2, t_3, \dots, t_{n-1}]$, then this reads $b - 1 = 0 \Rightarrow 1 - 1/b = 0$, which is true, but cannot be proved by an identity as in the cases $2 \leq k \leq n - 1$, because generally $b - 1 \neq 1 - 1/b$.

Problem 2013-1/B You are allowed to transform positive integers n in the following way. Write n in base 2. Write plus signs between the bits at will (at most one per position), and then perform the indicated additions of binary numbers. For example, $123_{10} = 1111011$ can get + signs after the second, third and fifth bits to become $11 + 1 + 10 + 11 = 9_{10}$; or it can get + signs between all the bits to become $1 + 1 + 1 + 1 + 0 + 1 + 1 = 6_{10}$; and so on.

Prove that it is possible to reduce arbitrary positive integers to 1 in a bounded number of steps. That is, there is a constant C such that for any n there is a sequence of at most C transformations that starts with n and ends at 1.

Solution This problem appeared originally in the Fall 2011 edition of the MSRI Emissary. We received solutions from Wouter Cames van Batenburg, Hao Chen, Pieter de Groen, Thijmen Krebs and Guido Senden. The book token goes to Guido Senden. The following solution contains ideas from various solutions.

For a positive integer n , let $I(n)$ denote the number of 1's in its binary expansion. For two positive integers a, b , we denote by $(ab)_2$ the positive integer obtained by concatenating the binary expansions of a and b . We remark that if a, b can be transformed into x, y , respectively, then $(ab)_2$ can be transformed into $x + y$, by placing a + between a and b .

Lemma 3. Every positive integer n can be transformed into every integer x with $I(n) \leq x \leq \frac{3}{2}I(n)$.

Proof. One transforms n into $I(n)$ by placing + at every position. Omitting a + after a 1 that has a + before it raises the sum by 1, as $2 = (1 + 0) + 1$ and $3 = (1 + 1) + 1$. This can be done at least $\lfloor \frac{1}{2}I(n) \rfloor$ times, by omitting the plus after every other 1. $\square$

Corollary 4. Any positive integer n such that $\frac{2}{3}2^{\lceil \log_2 I(n) \rceil} \leq I(n) \leq 2^{\lceil \log_2 I(n) \rceil}$ can be transformed into $2^{\lceil \log_2 I(n) \rceil}$.

The following proposition, in combination with Corollary 4 now shows that we only need to consider the cases where $I(n) \in \{4, 5, 8, 9, 10\}$.

Proposition 5. Suppose that all positive integers n with $8 \leq I(n) \leq 16$ can be transformed into 16. Let $k \geq 4$, and let n be a positive integer with $2^{k-1} \leq I(n) \leq 2^k$. Then n can be transformed into 2^k .

Proof. We proceed by induction. For $k = 4$, this is exactly our assumption. Suppose that the proposition is true for $k = i$, and let n be a positive integer with $2^i \leq I(n) \leq 2^{i+1}$. Write $n = (ab)_2$ with $I(a) = \lfloor \frac{1}{2}I(n) \rfloor$ and $I(b) = \lceil \frac{1}{2}I(n) \rceil$. Then $2^{i-1} \leq I(a) \leq I(b) \leq 2^i$, so a, b can both be transformed into 2^i , hence n can be transformed into 2^{i+1} , as desired. $\square$

We will now first treat the easiest of the remaining cases, i.e. $I(n) \in \{4, 8, 10\}$.

Lemma 6. Any positive integer n such that $I(n) = 4$ can be transformed into 8.

Proof. Write $n = (ab)_2$ with a consisting of the first three bits of n . If a starts with $(11)_2$, then $a = 8 - I(b)$, and b can be transformed into $I(b)$ by Lemma 3, so n can be transformed into 8. If a starts with $(10)_2$, then $a = 7 - I(b)$, and b can be transformed into $I(b) + 1$ by Lemma 3, as $I(b) \geq 2$. We deduce that n can be transformed into 8. $\square$

Corollary 7. Any positive integer n such that $I(n) \in \{8, 10\}$ can be transformed into 16.

Proof. Write $n = (ab)_2$, with $I(a) = 4$. Then both a and b can be transformed into 8 by Lemma 6 and Corollary 4, so n can be transformed into 16. $\square$

Finally, we do the case $I(n) \in \{5, 9\}$, which requires considering more cases.

Lemma 8. Any positive integer n such that $I(n) = 5$ can be transformed into a power of 2. If n does not start with $(11)_2$, then n can be transformed into 8.

Proof. First suppose that n starts with $(11)_2$. Then write $n = (ab)_2$, where a consists of the first four bits of n . If a starts with $(111)_2$, then $a = 16 - I(b)$, and b can be transformed into $I(b)$ by Lemma 3, so n can be transformed into 16. If a starts with $(110)_2$, then $a = 15 - I(b)$, and b can be transformed into $I(b) + 1$ by Lemma 3, as $I(b) \geq 2$. Hence n can be transformed into 16. Otherwise n starts with $(10)_2$. In this case, write $n = (ab)_2$, where $a = (10)_2$ consists of the first two bits of n . Now $I(b) = 4$, which can be transformed into 6 by Lemma 3, so n can be transformed into 8. $\square$

Lemma 9. Any positive integer n such that $I(n) = 9$ can be transformed into 16.

Proof. First suppose that n starts with $(11)_2$. Write $n = (ab)_2$ with a consisting of the first three bits of n . Then $a = 4 + I(a) = 13 - I(b)$. As $I(b) \geq 6$, we have $3 \leq \frac{1}{2}I(b)$. Hence $I(b) \leq I(b) + 3 \leq \frac{3}{2}I(b)$, so b can be transformed into $I(b) + 3$. We deduce that n can be transformed into 16.

Otherwise n starts with $(10)_2$. Write $n = (ab)_2$ in this case, with a containing the first 5 ones of n , and b containing the last 4 ones. Then by Lemmas 6 and 8, a and b can both be transformed into 8, so n can be transformed into 16. $\square$

Remark. As Pieter de Groen and Guido Senden remarked, one can prove very quickly that any positive integer can be reduced to 1 in at most 3 steps, once one has Corollary 4. For this, note that the remaining cases are those for which $2^{\lceil \log_2 I(n) \rceil - 1} < I(n) < \frac{2}{3} 2^{\lceil \log_2 I(n) \rceil}$. The smallest such case is $I(n) = 5$. Then one can transform n into $\frac{3}{2} \cdot 2^{\lceil \log_2 I(n) \rceil - 1}$ instead by Lemma 3, as $\frac{3}{2} > \frac{4}{3}$, so

$$I(n) < \frac{2}{3} \cdot 2^{\lceil \log_2 I(n) \rceil} = \frac{4}{3} \cdot 2^{\lceil \log_2 I(n) \rceil - 1} < \frac{3}{2} \cdot 2^{\lceil \log_2 I(n) \rceil - 1} < \frac{3}{2} I(n).$$

The number $\frac{3}{2} \cdot 2^{\lceil \log_2 I(n) \rceil - 1}$ can then be transformed into 2 as $I(n) \geq 5$, which can then be transformed into 1, showing that any positive integer can be reduced to 1 in at most three steps.

Problem 2013-1/C Let R be a commutative ring with 1. Consider the set

$$S = \{(i, j) \in R^2 : i^2 = i, j^2 = j, ij = 0\}.$$

Show that the cardinality of S is a power of 3 if S is finite.

Solution We received solutions from Florian Eisele, Thijmen Krebs and Guido Senden. The book token goes to Florian Eisele. The following solution is partly based on that of Florian Eisele and Thijmen Krebs.

Let us write $S(R)$ for the set S corresponding to a (commutative unital) ring R . First note that for all idempotents $i \in R$ (i.e. elements $i \in R$ with $i^2 = i$) we have $(i, 0) \in S(R)$. Hence the set of idempotents $I(R)$ of R is finite if S is finite, so we can proceed by induction on the number of idempotents $n(R)$ of R . Note that $0, 1 \in R$ both are idempotents for all rings R , so $n(R) \geq 1$, and if $n(R) = 1$, then R is the zero ring, for which $S(R) = \{(0, 0)\}$, hence $\#S(R) = 3^0$. If $n(R) = 2$, then $I(R) = \{0, 1\}$, therefore $S(R) = \{(0, 0), (1, 0), (0, 1)\}$, and $\#S(R) = 3^1$.

Now let $k > 2$, and suppose that for all rings R with $n(R) < k$, the set $S(R)$ has cardinality a power of 3. Let R be a ring with $n(R) = k$. As $n(R) > 2$, there exists an $i \in I(R)$ with $i \notin \{0, 1\}$. Let $j = 1 - i$, then $j \in I(R)$, and $ij = 0$. Then the map $\varphi: R \rightarrow R/iR \times R/jR$, $x \mapsto (x + iR, x + jR)$ is an isomorphism of rings by the Chinese Remainder Theorem. Note that φ induces a bijection $I(R) \rightarrow I(R/iR) \times I(R/jR)$, therefore also a bijection $S(R) \rightarrow S(R/iR) \times S(R/jR)$.

As j is non-zero, it induces a non-zero idempotent in $(R/iR) \times (R/jR)$. It follows that $j + iR \neq 0$, so $n(R/iR) > 1$ and similarly (using $i + jR$) also $n(R/jR) > 1$, hence $n(R/iR), n(R/jR) < n(R) = k$. It follows that $\#S(R/iR), \#S(R/jR)$ are powers of 3, hence so is $\#S(R) = \#S(R/iR) \#S(R/jR)$.