

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 tijdsgecorrelleerde) 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.

Weerextremen, 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  $(\mu, \sigma, \xi)$ :

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

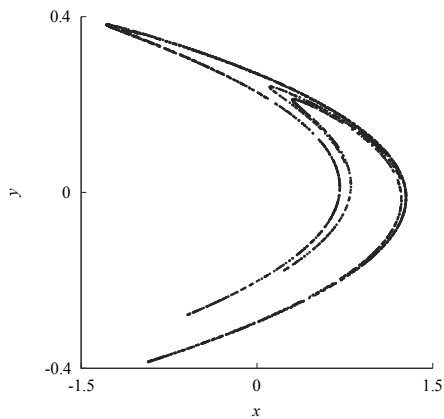
met  $\mu \in (-\infty, \infty)$  en  $\sigma \in (0, \infty)$ . Voor toepas-

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  $\mu$  en  $\sigma$  hierdoor, maar  $\xi$  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 een dichtheidspunt is van de invariante maat van het systeem, hetgeen betekent dat de relatieve maat van elke  $\epsilon$ -omgeving van  $\tilde{x}$  nadert tot 1 als  $\epsilon$  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

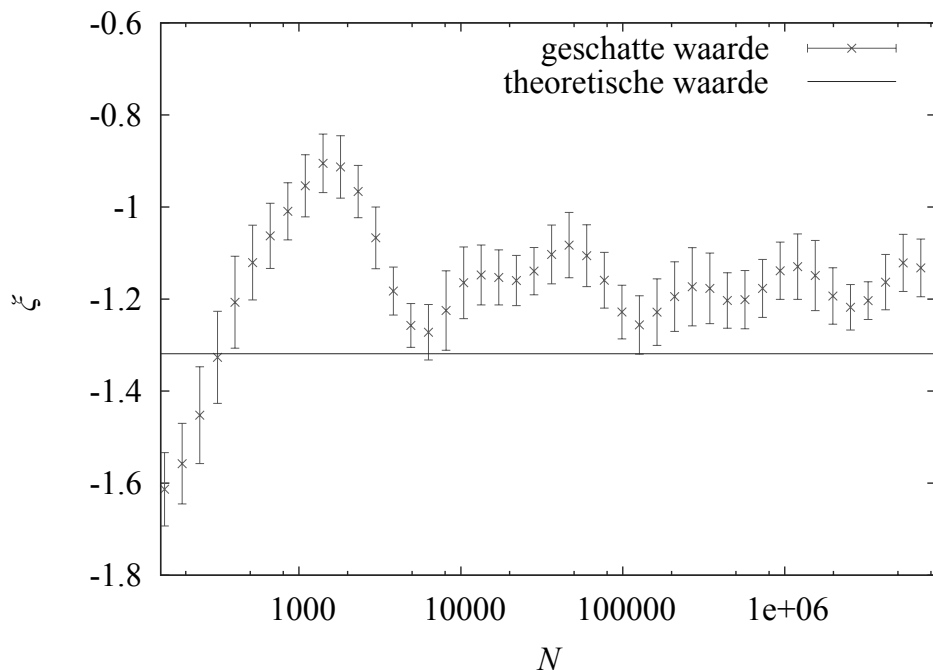
$$\omega(x) = v^\top x \quad (4)$$

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 definitie 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  $\xi$  als functie van de bloklength  $N$  berekend met behulp van de blokmaximummethode voor de Hénon-afbeelding.

zin dat de staartparameter  $\xi$  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  $\xi$  voldoet aan

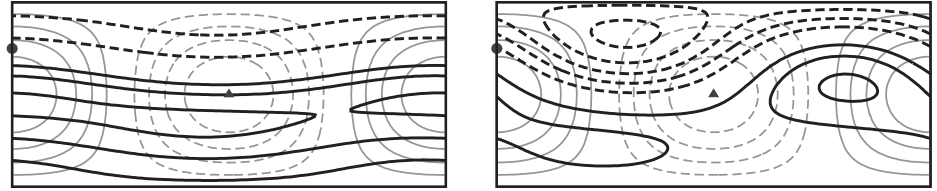
$$-\frac{1}{\xi} = \dim(A) - \frac{1}{2}, \quad (6)$$

waar  $\dim(A)$  de Hausdorff-dimensie van de Hénon-aantrekker is. Dit vermoeden is getoetst door  $\xi$  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  $\xi$  worden geschat met behulp van standaard methoden uit de statistiek zoals maximum likelihood.

Figuur 2 toont schattingen van  $\xi$  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  $\xi$  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  $\lambda$  een maat voor de exponentiële groei van fouten in de begintoestand en deze hangt af van zowel de begintoestand  $x$  als de voorspellingstijd  $\tau$ . In [16] is onderzocht of de groefactor  $\lambda$  groter of juist kleiner is voor begintoestanden  $x \in M$  die leiden naar extreme waarden van de observatie  $\omega(\Phi_\tau(x))$  na voorspellingstijd  $\tau$ .

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

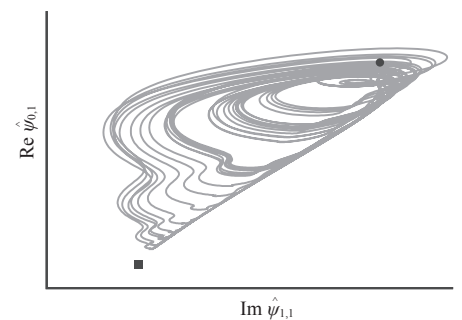
vreemde aantrekker in de 6-dimensionale toestandsruimte van het model, zie Figuur 4.

We kiezen twee observaties door de windsnelheid te berekenen op twee geografische locaties:  $\omega_{\text{zij}}$  en  $\omega_{\text{dal}}$  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  $\tau$  tijdseenheden leiden tot een overschrijding van  $\omega_{\text{zij}}$  of  $\omega_{\text{dal}}$  boven een vooraf gekozen grenswaarde  $q$ :

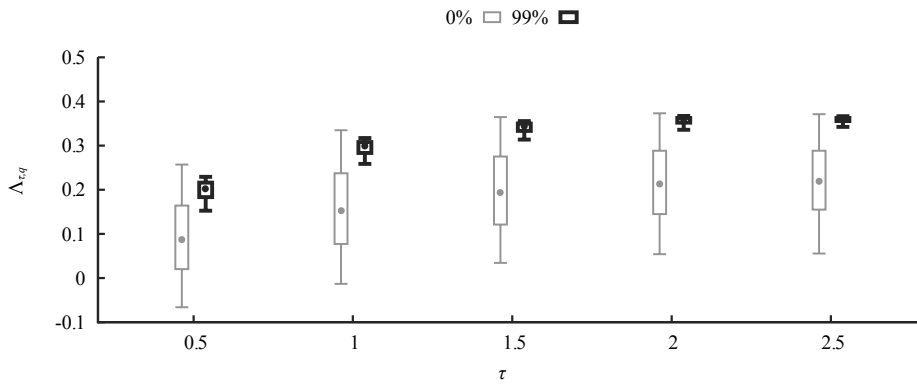
$$\Lambda_{\tau,q} = \{ \lambda(x, \tau) \mid \omega(\Phi_\tau(x)) \geq q \}.$$

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

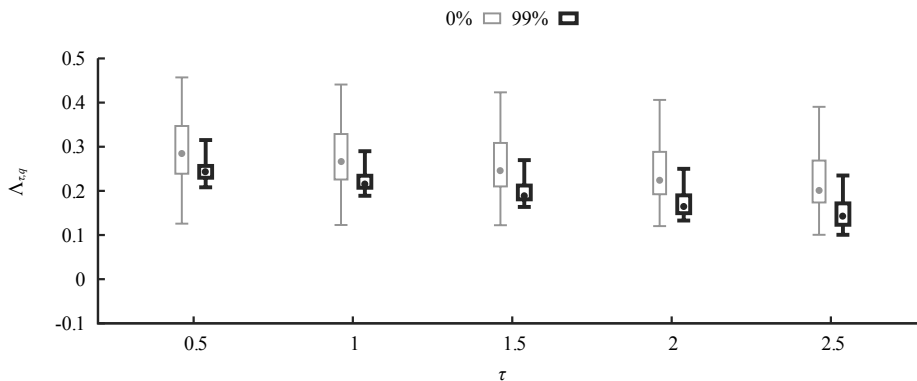
Figuur 5 toont boxplots van  $\Lambda_{\tau,q}$  als functie van de voorspellingstijd  $\tau$  voor  $\omega_{\text{zij}}$ . Voor elke waarde van  $\tau$  zijn twee boxplots weergegeven: voor grenswaarde  $q = 0$  (dat wil zeggen alle begincondities) en  $q$  gelijk aan het 99ste percentiel van de windsnelheid (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  $\tau$  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  $\tau$ -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  $\omega_{\text{dal}}$ : 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  $\tau$ .

### 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

- H.W. Broer, F. Takens. *Dynamical Systems and Chaos*. Springer, 2011.
- C. Bonatti, L. Diaz, M. Viana. *Dynamics Beyond Uniform Hyperbolicity*. Springer, 2005.
- S. Coles. *An Introduction to Statistical Modeling of Extreme Values*, Springer, 2001.
- J.G. Charney, J.G. DeVore. *Multiple flow equilibria in the atmosphere and blocking*, J. Atmos. Sci. 36, pp. 1205–1216, 1979.
- D.T. Crommelin, J.D. Opsteegh, F. Verhulst. *A mechanism for atmospheric regime behavior*, J. Atmos. Sci. 61, pp. 1406–1419, 2004.
- 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.
- 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.
- 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.
- G. Haiman. *Extreme values of the tent map process*, Statistics and Probability Letters 65, pp. 451–456, 2003.
- M.P. Holland, M. Nicol, A. Török. *Extreme value theory for non-uniformly expanding dynamical systems*, Transactions of the AMS, pp. 661–688, 2012.
- J.R. Holton. *Introduction to Dynamic Meteorology*. Elsevier Academic Press, 2004.
- M.R. Leadbetter, G. Lindgren, H. Rootzén. *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag, 1980.
- M.R. Leadbetter. *Extremes and local dependence in stationary sequences*, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 65, pp. 291–306, 1983.
- 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.
- E.N. Lorenz. *Deterministic non-periodic flow*, J. Atmos. Sci. 20, pp. 130–141, 1963.
- 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.
- H.E. de Swart. *Analysis of a six-component atmospheric spectral model: chaos, predictability and vacillation*, Physica D 36, pp. 222–234, 1989.
- L.-S. Young. *What are SRB measures, and which dynamical systems have them?*, J. Stat. Phys. 108, pp. 733–754, 2002.