

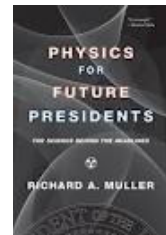
**Problemen met De Hockeystick:
statistische overwegingen bij het
gastcollege van Marcel Crok**

Peter Grünwald
HOVO 31-10 2011



“The Hockey Stick Story is a Tale that
Every Future President Needs to Know”
Richard A. Muller

Beroemde natuurkundige Muller
denkt dat de opwarming van de
aarde door de mens veroorzaakt
wordt, maar staat toch geheel
achter McIntyre en McKittrick in
hun kritiek op de hockey stick



Menu

1. Methodologie Mann, Bradley & Hughes 1998 (MBH98), het “hockeystick paper”
 - mini-cursus lineaire regressie
2. Technische Problemen met MBH98 zoals gevonden door McIntyre en McKittrick 2003, 2005 (M&M)
3. Meer Problemen met MBH98

Hockeystick Methodologie

1. Fitten van 122 (of 159?) “Proxies”
 - 1e gevaar: **overfitten**
 - 2e gevaar: **“brittleness”**
2. Calibratie vs. Verificatieperiode
 - MBH weten 1e gevaar te voorkomen
3. Principale Componentenanalyse
 - maar 2e gevaar niet!

Spreadsheet Demo

- [Hier wordt een demo gegeven tijdens het college]

Overfitten

- Als je veel 'proxy' reeksen hebt, is er altijd kans dat je gaat 'overfitten':
door puur toeval zit er een reeks bij die je heel goed kunt 'fitten' op alle punten die je hebt gemeten, maar daarbuiten helemaal niet (de reeks heeft geen voorspellende kracht)
- Vgl. "Pietje heeft de lotto gewonnen dus Pietje is een fraudeur"
- Dit probleem gaan MBH98 *op zich* goed te lijf door een **verificatieperiode** in te voeren en via een 'significance level' te kijken hoe goed hun data op de verificatieperiode past
 - Alleen berekenen ze dat significance level dan verkeerd...

De "Verificatieperiode"

- MBH's studie loopt tot 1980
- Data over temperatuur gaat terug tot 1854
- De proxies worden **gefit** op de periode 1902-1980. Dit levert functie op:

$$Y = a_0 + a_1 \cdot \text{proxy}_1 + a_2 \cdot \text{proxy}_2 + \dots + a_{122} \cdot \text{proxy}_{122}$$

- Deze functie wordt **getest** (geverifieerd) op periode 1854-1901: er wordt gekeken hoe goed de temperatuur in die jaren (die immers bekend is) voorspeld wordt door de gevonden functie. Hoe beter dat gaat, hoe meer **vertrouwen** men kan hebben in de gevonden functie"
- vervolgens gebruikt men functie om de temperatuur van de jaren 1000-1853 te voorspellen (*reconstrueren*)

De R² statistic

- De R² statistic is een standaard maat voor het vertrouwen in de gevonden voorspel-functie:

$$R^2 = \frac{\sum_{i=1}^n (\text{voorspelling voor } y_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

\bar{y} = gemiddelde van alle n uitkomsten

- In ons geval geldt $n=48$, y_1 is temp. in 1854, y_2 is temp. in 1855, enz.

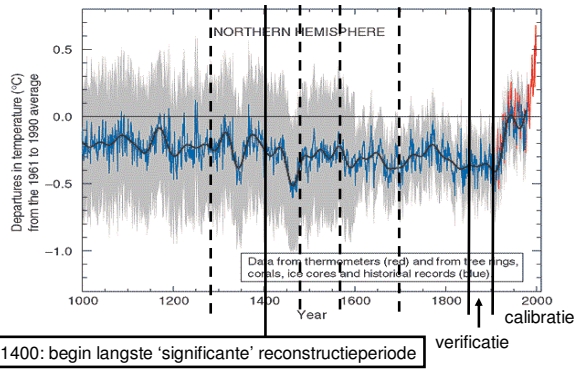
Verificatie

- MBH '98 hebben dus verschillende betrouwbaarheidssindicatoren: R^2 -1000, R^2 -1400, R^2 -1450, etc.
- Ze gebruiken niet alleen de R^2 statistic, maar ook de veel minder bekende RE statistic
 - Op zich is het goed om meer dan 1 statistic te gebruiken, maar let op wat volgt...
- De 1400 stap is de **vroegste stap die een zo grote RE-statistic geeft dat hij volgens MBH serieus te nemen is.**

Om het nog moeilijker te maken...

- MBH '98 doen dit hele proces niet één keer, maar een aantal keren
- Ze hebben nl. maar 22 proxies die tot 1400 teruggaan in de tijd, 57 die tot 1450 gaan, enz. enz.
- Om dit probleem op te lossen bepalen ze de voorspellingen voor
 - 1400-1450: fit de 22 beschikbare proxies op 1902-1980, en test gevonden functie op 1854-1901. Dit levert de R^2 en RE waarde voor de '1400 stap'
 - 1450-1500: fit de 57 beschikbare proxies voor 1550-1450 op 1902-1980, en test gevonden functie op 1854-1901. Dit levert de R^2 en RE waarde voor de '1500' stap
 - enz...

Calibratie, Verificatie en "Reconstructie"



Principale Componenten - I

- Van sommige gebieden zijn er veel meer 'proxies' beschikbaar dan van anderen (bijvoorbeeld de NoAmer database).
- MBH98 besluiten die proxies 'samen te vatten' zodat er van elke regio ongeveer evenveel proxies zijn
- Hoe doe je dat?
 - 1e idee: neem de gemiddelde waarden van alle proxies in NoAmer. Dat werkt niet, o.a. omdat sommige boomsoorten veel harder groeien dan andere, en sommigen veel minder reageren op temperatuur dan anderen
 - MBH nemen daarom een sophisticated soort gemiddelde: ze nemen een aantal *principale componenten*
 - dat is op zich o.k ; maar we zullen zometeen zien dat ze de principale componenten verkeerd bepalen

Principale Componenten - II

- Op dezelfde manier nemen MBH niet de “gemiddelde temperatuur op het noordelijk halfrond” in jaar 1854, 1855, ... als Y-waarden, maar nemen ze de eerste 4 principale componenten van de reeks van alle temperatuurmetingen
- Niet van belang voor wat volgt

Technische Problemen

Probleem 1: De Data

- Data zijn **rommelig**. Mann legt niet uit hoe hij hier mee omgaat
- sommige jaargangen van sommige proxies ontbreken. Die worden door Mann zelf “ingevuld”, op een niet consequente manier (maar à la...)
 - sommige proxies komen zowel voor “op zichzelf” als als onderdeel van een principale componenten proxy (dit is wél een serieus probleem in combinatie met probleem 2)
 - de keuze van het aantal te behouden principale componenten is niet op enig principe gebaseerd (we komen hier op terug)
 - (nog meer gedoe met de data)

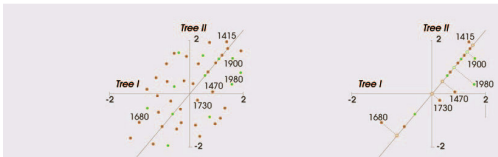
2. PCA is fout geprogrammeerd

- Het computerprogramma dat PCA implementeert bevat een fout met serieuze consequenties in combinatie met probleem 1
 - **als** er een proxy met een hockeystickvorm onderdeel is van de PCA, dan wordt dat de 1e principale component!!!
- Belangrijk: het feit dat met deze fout hockeysticks worden gefabriceerd suggereert dat deze fout met opzet is gemaakt. Maar dat is niet zo: dit is geen fraude, dit is echt gewoon een fout

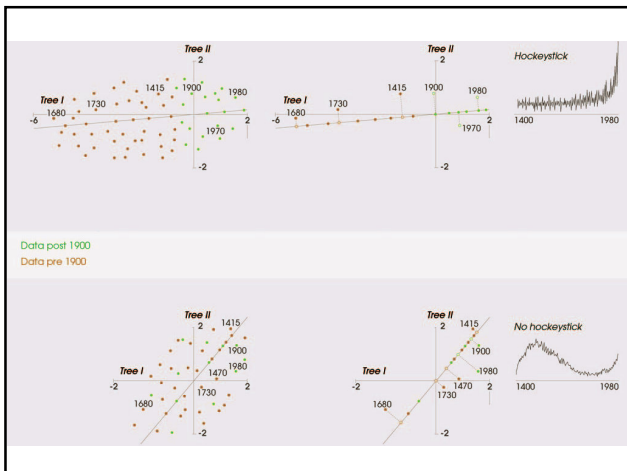
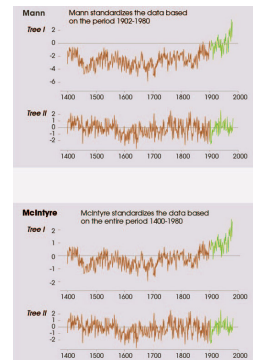
Principale Componenten

(overgenomen uit NWT 2005, zie HOVO website)

PCA wordt gebruikt om een "slim" gewogen gemiddelde te bepalen van een reeks proxies



Principal Component Analysis



Probleem 1. en 2., gecombineerd

- Als de stekelden (bristlecone pine) en de Gaspé reeks uit de data worden gehaald, dan verdwijnt de hockeystick en wordt de grafiek een rommeltje!
- De hele hockeystickgrafiek is feitelijk gebaseerd op **twee** proxies
 - Van de stekelden is zelfs bekend dat hij niet zou moeten worden gebruikt in temperatuurreconstructies omdat hij in de 20e eeuw veel harder is gaan groeien dan andere boomsoorten!

3. Resultaat is *niet bepaald significant*

- Mann vermeldt RE-statistic voor al zijn reconstructies, zegt dat hij R^2 berekend heeft voor al zijn reconstructies, maar geeft de R^2 waarde alleen voor de 1820-1980 stap
- Voor zover na te gaan blijkt: de 1820-1980 stap is de enige die hoge R^2 geeft. Alle andere stappen geven R^2 van ongeveer 0.02 (waardeloos!)
- Dit is wat mij betreft op het randje...

Resultaat niet significant - 2

- MBH98 noemen alleen de RE statistiek voor de 1400 stap
 - die komt lekker hoog uit (0.51) en is volgens hun 'significant' op 5% level
- M&M laten zien dat de significantieberekening niet klopt – volgens hun is 0.51 nog niet significant
 - Dit is tricky omdat je allerlei extra aannames moet maken om een 'statistic' (data, heeft niets met kansen te maken!) in een significance level te veranderen (daarvoor met je een 'nulhypothese'- kansverdeling postuleren)
 - Maar wat M&M doen lijkt alleszins redelijker dan wat Mann doet
- Hoe dan ook is lage R^2 statistiek zeer verdacht!

4. Resultaat niet robuust ('brittle')

- **Sparsity Effect:** bij regressieanalyse is het altijd mogelijk dat Y (temperatuur) afhankelijk is van slechts een paar van de Xen (proxies).
- Soms is dat niet erg, maar in geval van temperatuur-reconstructies wél – maakt ze betekenisloos
- Bij zo'n heroïsch project als MBH98 moet je zowieso *altijd* een **robuustheidsanalyse** doen: wat gebeurt er als je bepaalde proxies uit de dataset haalt? Krijg je iets heel anders?
- Dat is niet gebeurd.
- Dubbel gebruik van proxy (op zichzelf en als onderdeel van PCA proxy), ad hoc kiezen van aantal te behouden PCA componenten, verkeerde PCA implementatie enz. **werken sparsity effect juist sterk in de hand**

“Weerlegging”

- Deze en nog veel meer problemen blootgelegd door McIntyre en McKittrick in een aantal paper
- MBH verdedigen zich op niet serieus te nemen manier (“verkeerde PCA methode heeft geen invloed op resultaat wanneer je Preisendorfer's Regel gebruikt om aantal principale componenten te bepalen” (5!) ...enz.)
- Wahl & Amman 2007a,b: “problemen 2,3 en 4 hebben geen invloed op eindresultaat mits je alles op 'juiste' manier berekent/interpreteert”
 - “ R^2 statistic **werkt niet goed voor klimaat-reconstructies**”, “confidence levels RE statistic moeten berekend worden op basis **van een mix tussen de verificatie en de calibratiewaarden...**”
- Ik ben niet bepaald overtuigd...

“Probleem” met R² statistic

- “The combination of high RE and low R² means that the reconstruction identified the change in mean levels between the calibration and validation periods reasonably well but failed to track the variations within the validation period”
(*Surface Temperature Reconstructions for the Last 2000 Years*, National Academies Press, 2006)

$$RE = 1 - \frac{\sum_{i=1854..1901} (y_i - \text{voorspelling } i)^2}{\sum_{i=1854..1901} (y_i - \text{gemiddelde } 1902-1980)^2}$$

Problemen met de Betekenis

Divergentieprobleem – al behandeld door Marcel

Problemen met het Grote Vertrouwen in Wetenschap

Replicatie

- Hockeystick is “gerepliceerd” maar dit is omstreden (“onafhankelijke” data eigenlijk niet zo onafhankelijk...)
- MBH98 laat extreem veel details onvermeld (niet echt hun schuld: Nature papers *moeten* kort zijn...)
- M&M hebben jaren moeten vechten om onderliggende data te verkrijgen – nog *steeds* niet 100% ‘letterlijk reproduceerbaar’
- Dit laat zien dat, terwijl de grafiek vele malen voorkomt in IPCC 2001 rapport, niemand had geprobeerd MBH te repliceren!
- **Dit is een zeer zorgwekkende gang van zaken**
 - MBH deed statistiek helemaal verkeerd.
 - zelfs al doe je de statistiek goed, dan kan er nog van alles misgaan (publicatiebias, verkeerde data)
 - **Onafhankelijke replicatie blijft essentieel!**

Replicatie

- Hockeystick is "gerepliceerd" maar dit is omstreden ("onafhankelijke" data eigenlijk niet zo onafhankelijk...)
- MBH98 laat extreem veel details onvermeld (niet echt hun schuld: Nature papers *moeten* kort zijn...)
- M&M hebben jaren moeten vechten om onderliggende data te verkrijgen – nog *steeds* niet 100% 'letterlijk reproduceerbaar'
- Dit laat zien dat, terwijl de grafiek vele malen voorkomt in IPCC 2001 rapport, niemand had geprobeerd MBH te repliceren!
- **Dit is een zeer zorgwekkende gang van zaken**
 - MBH deed statistiek helemaal verkeerd.
 - zelfs al doe je de statistiek goed, dan kan er nog van alles misgaan (publicatiebias, verkeerde data)
 - **Onafhankelijke replicatie blijft essentieel!**

En als je er wat van zegt, ben je "fout"...

Meer weten?

- Het is erg moeilijk om goede informatie op het internet te vinden
 - zelfs wikipedia is "biased"
- Blogs aan beide zijden in het debat zijn vaak uiterst subjectief (en onaangenaam)
- De blog van McIntyre zelf is inhoudelijk goed, maar erg moeilijk te volgen
- Veruit de beste referentie (wel gekleurd van toon, maar inhoudelijk grotendeels correct, wel erg dik) is *The Hockeystick Illusion* van A.W. Montford (2010)
- De beste **korte referentie** (wellicht wereldwijd!) is nog steeds Marcel Crok's NWT artikel op de HOVO website

