



Teknisk rapport 11-11

Ekstremværdianalyse af nedbør i Danmark 1874 - 2010

Sisse Camilla Lundholm





Dmi
Klima- og Energiministeriet



Kolofon

Serietitel:

Teknisk rapport 11-11

Titel:

Ekstremværdianalyse af nedbør i Danmark 1874 - 2010

Undertitel:**Forfatter(e):**

Sisse Camilla Lundholm

Andre bidragsydere:

John Cappelen, Claus Kern-Hansen

Ansvarlig institution:

Danmarks Meteorologiske Institut

Sprog:

Dansk

Emneord:

Ekstremværdianalyse, nedbør, ekstremnedbør

Url:

www.dmi.dk/dmi/tr11-11

ISSN:

1399-1388

Versions dato:**Link til hjemmeside:**

www.dmi.dk

Copyright:

Danmarks Meteorologiske Institut. Det er tilladt at kopiere og uddrage fra publikationen med kildeangivelse.

Forsidebillede:

Grafik: Michael Scharling



Indhold:

Abstract	5
Resumé.....	5
1 Indledning	6
2 Dataforberedelse	7
3 Ekstremværdianalysen	8
3.1 Bestemmelse af tærsklen.....	8
3.2 Gentagelsesniveauer.....	9
3.3 Sammenligning mellem denne og den tidligere analyse.....	10
4 Konklusion	13
4.1 Analyse af de lange tidsserier	13
4.2 Sammenligning med den tidligere analyse	13
Referencer	14
Tidligere rapporter	14



Abstract

This report describes an extreme value analysis of five time series of daily precipitation in Denmark for the years 1874 - 2010 (137 years).

Resumé

Nærværende rapport præsenterer en ekstremværdianalyse af fem tidsserier af daglig nedbør i Danmark for årene 1874 - 2010 (137 år).



1 Indledning

I 2010 udarbejdede Lundholm et al. en ekstremværdianalyse af daglig nedbør ved 112 af DMI's manuelle nedbørsstationer dækkende perioden 1961 – 2010. Analysen og resultaterne er beskrevet i [5]. For at udnytte arbejdsindsatsen med opsætning af analysemetoden er der efterfølgende foretaget ekstremværdianalyse af DMI's fem lange nedbørsserier.

De tre af de fem nedbørsserier, der er behandlet her, består af 137 års målinger af daglig nedbør ved samme nedbørsstation, mens de to øvrige serier er sammensat af kortere serier af daglig nedbør ved forskellige, tætliggende stationer, der sammenlagt dækker perioden 1874 – 2010¹.

De tidsserier af døgnnedbør, der ligger til grund for denne rapport, kan ses i tabel 1, og dataene er hentet fra [1].

Site and period	Station	Start	End
Vestervig 1874-2010	21100 Vestervig	1. januar 1874	31. december 2010
Nordby/Fanø 1874-2010	25140/06088 Nordby	1. januar 1874	31. december 2010
Tranebjerg 1872-2010	27080 Tranebjerg 27082 Tranebjerg Øst	1. december 1872 2. august 2001	1. august 2001 31. december 2010
København 1874-2010	30380 Landbohøjskolen 30210 Meteorologisk Institut 30210 Meteorologisk Institut 30370 Botanisk Have	1. januar 1874 1. januar 1875 1. januar 1961 1. januar 1961	1. oktober 1996 30. juni 1922 31. december 1984 31. december 2010
Hammer Odde 1874-2010	32030 Sandvig 32020 Hammer Odde Fyr 06193 Hammer Odde Fyr	1. januar 1874 1. januar 1961 1. januar 1984	31. december 1970 30. juni 1987 31. december 2010

Tabel 1. DMI's fem lange tidsserier af daglig nedbør [1].

Nedbørsserierne stammer fra filer vedhæftet [1]. Af de vedhæftede filer drejer det sig om

p21100_1874_2010.dat
p06088_1874_2010.dat
p27080_1872_2001.dat
p27082_2001_2010.dat
p30210_1875_1922.dat
p30210_1961_1984.dat
p30370_1961_2010.dat
p30380_1874_1996.dat
p32030_1874_1970.dat
p32020_1961_1987.dat
p06193_1984_2010.dat

¹ Det er ikke ualmindeligt for en nedbørsstation at blive flyttet lidt rundt i løbet af dens levetid, og for "Tranebjerg" har stationen i forbindelse med en flytning fået nyt nummer, mens nedbørsmålingerne har fortsat uafbrudt. Disse to nedbørsserier behandles her som en enkelt nedbørsserie målt ved sammen station.

2 Dataforberedelse

De to sammensatte nedbørsserier er konstrueret på følgende måde: Hvis nedbørsmålingerne ved to eller flere af de involverede stationer overlapper tidsmæssigt, findes nedbøren for den sammensatte nedbørsserie som den maksimale nedbør af de overlappende målinger. Denne fremgangsmåde er vurderet mest hensigtsmæssig for ekstremværdianalysen. Med de sammensatte nedbørsserier er der i alt fem serier med 137 års målinger af daglig nedbør.

Da høje nedbørsmængder har betydning for ekstremværdianalysen, er de 10 højeste nedbørsmålinger for hver station fundet – også for de stationer, der indgår i de sammensatte nedbørsserier – inden selve ekstremværdianalysen. For de nedbørsserier, der indgår i denne analyse, har det ikke resulteret i suspekterede værdier, det har været nødvendigt at undersøge nærmere. Der er derfor uden korrektioner af de fem nedbørsserier foretaget ekstremværdianalyse. Den geografiske placering af stationer med lange nedbørsserier kan ses i figur 1.



Figur 1. Geografisk placering af DMI's stationer med 137 års målinger af daglig nedbør.
Grafik: Michael Scharling

3 Ekstremværdianalysen

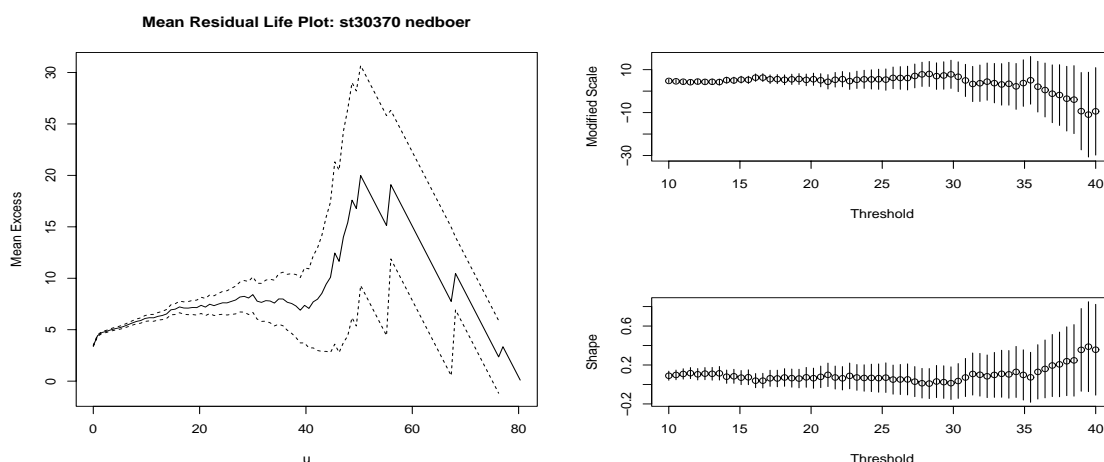
Ekstremværdianalysen er udført ved hjælp af ”extRemes”-pakken [3], der hører til det frit tilgængelige statistik-program ”R” [2].

En metode til ekstremværdianalyse er at definere en nedbørshændelse som ekstrem, hvis den overstiger en valgt, høj tærskel og at kigge udelukkende på de værdier, som overstiger tærsklen. Denne metode af ekstremværdianalyse er særligt anvendelig ved daglige målinger af en given parameter [4] – for eksempel daglig nedbør, og den metode er anvendt her. De værdier, der ligger over en høj tærskel – ekstremværdierne – antages at følge en bestemt, kendt fordeling, der kaldes en Generalized Pareto (GP) fordeling.

3.1 Bestemmelse af tærsklen

”extRemes”-pakken tilbyder et sæt af figurer ud fra hvilke, den mest optimale tærskel – en central del af denne type ekstremværdianalyse – kan findes². Den mest optimale tærskel er den mindste tærskelværdi, der samtidig opfylder, at ekstremværdierne er GP-fordelte. Hvis vi sætter tærsklen for lavt, holder antagelsen om at ekstremværdierne er GP-fordelte ikke. Hvis vi sætter tærsklen for højt, får vi meget få ekstremværdier, hvilket giver store usikkerheder på estimererne af gentagelsesniveauerne.

For den optimale tærskel og alle tærskler over, vil parametrene ”Mean Excess”, ”Modified Scale” og ”Shape” (se figur 2) være uafhængige af tærskelværdien og dermed mere eller mindre konstante (i den ene figur anvendes symbolet u for tærsklen). For mange datasæt vil denne tærskel ikke være klart og entydigt bestemt, og derfor kan fastsættelsen af denne størrelse være tidskrævende. Det kan heller ikke udelukkes, at forskellige analytikere kan komme frem til forskellige optimale tærskelværdier. Estimering af en passende tærskel ud fra ”R’s” figurer var for tidskrævende for de mange nedbørsserier fra den tidligere analyse beskrevet i [5], men for de fem lange tidsserier har det været muligt. I figur 2 ses ”extRemes”s to figurer til estimering af den optimale tærskel for nedbørsserien for København.



Figur 2. Figurer til estimering af den optimale tærskel for nedbørsserien for København. (I ”R” er datasættet for den sammensatte tidsserie benævnt ”st30370”, hvilket fremgår af titlen på figuren til venstre.)

² Følgende referencer giver mere information om valg af tærskel: kapitel 4 i [4] og kapitel 5 i <http://www.isse.ucar.edu/extremevalues/tutorial.pdf>

De estimerede tærskler er for

Station	Tærskel	Antallet af ekstremværdier
Vestervig (Vesterv):	23 mm	301
Nordby:	20 mm	406
Tranebjerg (Traneb.):	20 mm	263
København (Kbh):	17 mm	617
Hammer Odde (HammerO):	18 mm	394

med angivelse af det medfølgende antal af værdier, der overstiger tærsklen.

3.2 Gentagelsesniveauer

”extRemes”-pakken giver mulighed for at estimere et gentagelsesniveau med tilhørende konfidensinterval, og tabel 2 viser estimererne af 10-, 20-, 50-, 100-, 500- og 1000-års hændelserne med de tilhørende 95 % konfidensintervaller. Et 95 % konfidensinterval er et interval indenfor hvilket, vi med 95 % sikkerhed kan sige, at den sande hændelse – som vi ud fra det givne datasæt forsøger at estimere – ligger. Som forklaring af tabellen er Xyr.l estimeret af en X-års hændelse (yr: year; l:level) og X.ci1, X.ci2 er henholdsvis nedre og øvre grænse af 95 % konfidensintervallet. For station Vestevig (21100) er estimeret af 10-års hændelsen 43.5 mm og 95 % konfidensintervallet er [41,2 mm; 46,9 mm]. Vi er dermed 95 % sikre på, at den sande 10-års hændelse ligger mellem 41,2 og 46,9 mm.

Navn	station	10yr.l	10.ci1	10.ci2	20yr.l	20.ci1	20.ci2	50yr.l	50.ci1	50.ci2	100yr.l	100.ci1	100.ci2
Vester.	21100	43.5	41.2	46.9	48.2	45.0	53.6	54.5	49.7	63.4	59.3	52.9	71.9
Nordby	06088	46.8	43.6	51.6	53.6	48.9	61.3	63.5	56.1	76.7	71.7	61.6	90.7
Traneb.	27082	43.4	40.1	48.2	50.6	45.7	58.5	61.3	53.5	75.9	70.5	59.6	92.5
Kbh.	*	45.5	42.7	49.5	51.3	47.4	57.3	59.3	53.5	68.8	65.7	58.1	78.6
HammerO	*	45.9	42.6	50.8	52.9	48.1	60.5	62.8	55.3	75.8	70.9	60.8	89.5

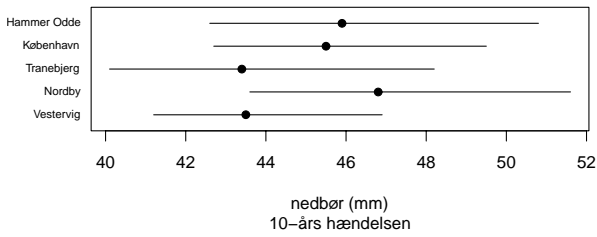
* sammensatte tidsserier

500yr.l	500.ci1	500.ci2	1000yr.l	1000.ci1	1000.ci2
70.4	59.7	95.1	75.3	62.2	106.9
93.3	74.7	133.3	103.8	80.4	157.0
95.8	74.5	146.8	108.8	81.2	179.1
81.3	68.6	105.4	88.4	72.8	118.9
91.7	73.4	130.2	101.8	78.8	152.4

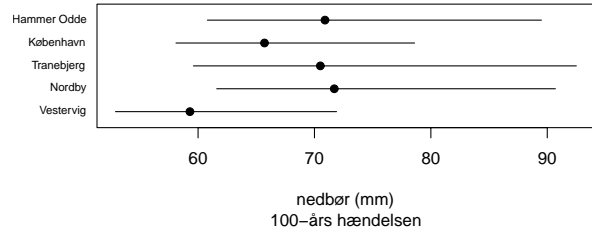
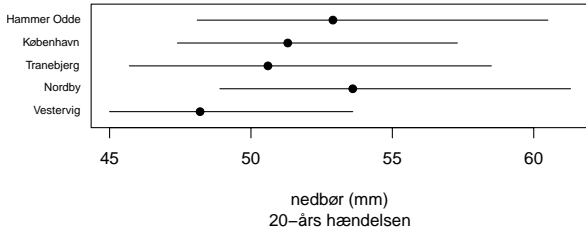
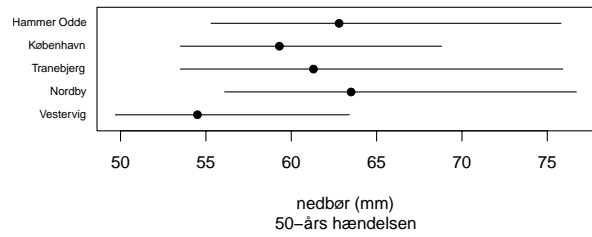
Tabel 2. Estimeret af 10-, 20-, 50-, 100-, 500- og 1000-års hændelser med tilhørende konfidensintervaller. Stationsnavnene er til dels forkortede. De fulde navne findes ovenfor.

Det fremgår af tabel 2 og figur 2, at jo længere ekstremværdianalysen skal ekstrapolere i tid (for eksempel estimeret af en 1000-års hændelse ud fra 137 års nedbørsdata), jo mere usikkert bliver estimeret, og jo bredere bliver det tilhørende konfidensinterval.

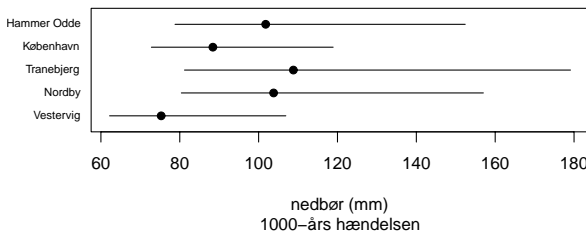
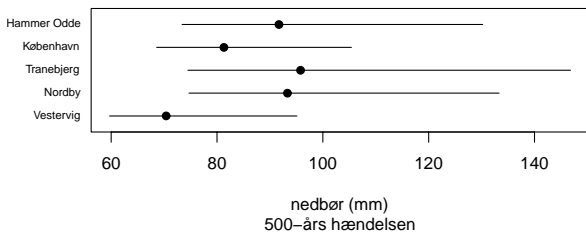
Estimater og konfidensintervaller for gentagelsesniveauer



Estimater og konfidensintervaller for gentagelsesniveauer



Estimater og konfidensintervaller for gentagelsesniveauer



Figur 2. Illustration af dataene i tabel 2 med estimater af 10-, 20-, 50- og 100-års hændelser og deres tilhørende konfidensinterval (linje)

De nøje estimater af tærsklen i ekstremværdianalysen har givet et højt antal ekstremværdier, som opfylder, at de med god tilnærmelse er GP-fordelte. Det giver grund til at tro, at estimaterne af 10-, 20-, 50- og 100-, 500- og 1000-års hændelserne og deres konfidensintervaller er yderst fornuftige. Proceduren med estimering af en passende tærskel kan have bevirket, at gentagelsesniveauerne udviser lille variation mellem de forskellige stationer – se tabel 2 og figur 2. Det er mest udtalt for gentagelsesperioder på op til 100 år.

3.3 Sammenligning mellem denne og den tidligere analyse

Der er en generelt god overensstemmelse mellem estimaterne af 10-, 20-, 50- og 100-års hændelserne fra analysen af de fem lange tidsserier og den tidligere analyse i [5] – se tabel 3.

Bemærk at for de to sammensatte nedbørsserier bygger sammenligningen enten på den af de sammensatte stationer, der findes blandt stationerne i [5] (30370, Botanisk Have) eller på en anden nærliggende station (32280, Poulsker (Bornholm)). Både længden af nedbørsserien og de involverede stationer i analysen af de lange, sammensatte tidsserier adskiller sig fra den tidligere analyse.

eva station	10yr.l	10.ci1	10.ci2	20yr.l	20.ci1	20.ci2	50yr.l	50.ci1	50.ci2	100yr.l	100.ci1	100.ci2
KFT 21100	42.7	39.9	47.7	46.4	42.8	53.8	51.0	46.2	62.6	54.3	48.3	69.9
fem 21100	43.5	41.2	46.9	48.2	45.0	53.6	54.5	49.7	63.4	59.3	52.9	71.9
KFT 25140	46.2	41.3	56.4	52.6	45.4	69.9	61.9	50.5	93.8	69.6	54.1	117.9
fem 06088	46.8	43.6	51.6	53.6	48.9	61.3	63.5	56.1	76.7	71.7	61.6	90.7
KFT 27082	44.6	39.1	55.5	52.8	44.5	71.6	65.4	52.1	101.4	76.7	58.1	132.8
fem 27082	43.4	40.1	48.2	50.6	45.7	58.5	61.3	53.5	75.9	70.5	59.6	92.5
KFT 30370	44.9	41.0	52.1	50.1	44.9	61.2	57.0	49.7	75.1	62.1	52.9	87.4
fem* Kbh.	45.5	42.7	49.5	51.3	47.4	57.3	59.3	53.5	68.8	65.7	58.1	78.6
KFT 32280	48.3	42.6	59.6	56.1	48.0	74.6	67.4	55.0	100.5	76.8	60.3	126.1
fem* HamerO	45.9	42.6	50.8	52.9	48.1	60.5	62.8	55.3	75.8	70.9	60.8	89.5

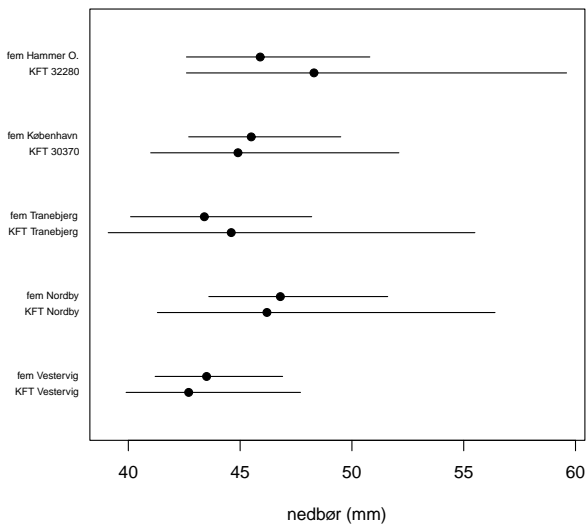
*sammensatte tidsserier

Tabel 3. Estimer af 10-, 20-, 50- og 100-års hændelser fra den tidligere analyse [5] benævnt "KFT" og fra analysen af de fem lange tidsserier benævnt "fem".

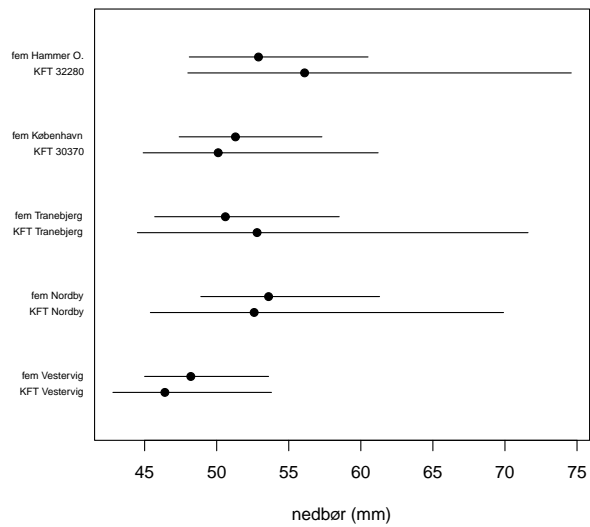
En sammenligning mellem analysen af de hele, lange tidsserier med den tidligere analysen for samme stationer viser, at denne analyses konfidensintervaller overordnet er blevet betydeligt smallere – som forventet med de lange nedbørsserier. Sammenlign for eksempel gentagelsesniveauer og konfidensintervaller fra denne analyse (angivet med "fem") med den tidligere analyse (angivet med "KFT") for station 21100 (Vestervig), 25140/06088 (Nordby) og 27082 (Tranebjerg) i tabel 3 og i figur 3, som tjener som illustration af dataene i tabellen. De sorte prikker indikerer estimatet af gentagelsesniveauet og de vandrette linjer det tilhørende konfidensinterval.

Hvis der findes relativt mange høje – eller ekstreme – værdier blandt nedbørsdataene før 1961, så bliver de nye estimater lidt højere end estimaterne i [5], og hvis der findes relativt få høje – eller ekstreme – værdier blandt nedbørsdataene før 1961, så bliver de nye estimater lidt lavere end estimaterne i [5]. De nye estimater ligger indenfor konfidensintervallerne fra den tidligere analyse.

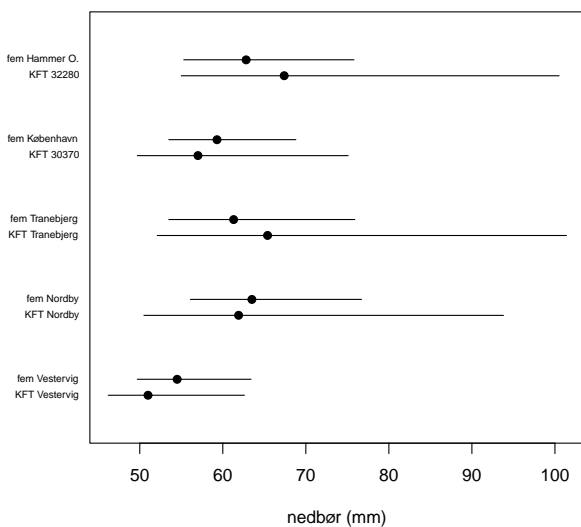
Estimater og konfidensintervaller for 10-års hændelsen



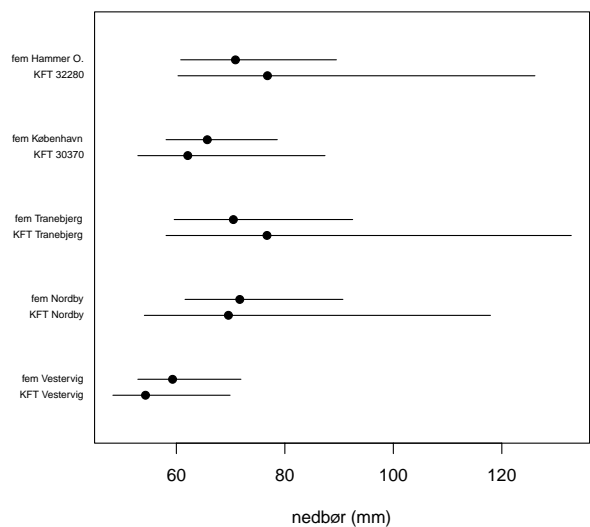
Estimater og konfidensintervaller for 20-års hændelsen



Estimater og konfidensintervaller for 50-års hændelsen



Estimater og konfidensintervaller for 100-års hændelsen



Figur 3. Illustration af dataene i tabel 3; Hammer O. er Hammer Odde.

Det manglende sammenfald mellem Hammer Odde-stationerne og stationerne i [5] afspejler sig i estimaterne af 10-, 20-, 50- og 100-års hændelserne. Der er generelt større afvigelse mellem Hammer Odde-estimatet og estimatet for station 32280 end tilfældet er for de øvrige sammenlignelige estimater – se figur 3.

Betydningen af de lange tidsserier i ekstremværdianalysen er tydelig: For fire stationer (Fanø (06088), Tranebjerg (27082), København (30370) og Hammer Odde (06193)) svarer konfidensintervallets bredde for en 500-års hændelse til en 100-års hændelse for tidsserierne i [5] (se eventuelt tabel 2 og 3). Da tidsserien for Poulsker (32280) ikke er en del af den lange, sammensatte tidsserie for Hammer Odde, er sammenligningen af konfidensintervallernes bredde tentativ. For en station (Vestervig (21100)) er der overensstemmelse i bredden af konfidensintervallet mellem denne analyse og KFT-analysen.



4 Konklusion

4.1 Analyse af de lange tidsserier

Den grundige ekstremværdianalyse med estimering af den optimale tærskel for hver af de lange tidsserier kan være årsag til, at gentagelsesniveauerne udviser lille variation mellem de forskellige stationer (det er mest gældende for gentagelsesperioder på op til 100 år). De lange tidsserier har bevirket, at konfidensintervallernes bredde er betydeligt reduceret i forhold til den tidligere analyse, og det taler for en stor tiltro til gentagelsesniveauerne og deres konfidensintervaller.

4.2 Sammenligning med den tidligere analyse

Der er en fin overensstemmelse af estimerne af 10-, 20-, 50- og 100-års hændelser mellem analysen af DMI's lange nedbørsserier og – for de samme stationer – analysen af 50-års daglig nedbør beskrevet i [5]. Det er tilfredsstillende og understøtter kvaliteten af den landsdækkende ekstremværdianalyse af op til 50 års daglig nedbørsmålinger [5].

De lange nedbørsserier giver betydeligt mere sikkerhed i estimerne af gentagelsesniveauerne i forhold til de op til 50 års daglige værdier i den landsdækkende ekstremværdianalyse beskrevet i [5]. Det viser sig som væsentligt smallere konfidensintervaller på estimerne af 10-, 20-, 50- og 100-års hændelserne i forhold til analysen i [5].

Referencer

- [1] Cappelen, J. (2011): DMI Daily Climate Data Collection 1873-2010, Denmark, The Faroe Islands and Greenland – including Air Pressure Observations 1874-2010 (WASA Data Sets), DMI Technical Report 11-06.
- [2] R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- [3] Eric Gilleland, Rick Katz and Greg Young (2010). extRemes: Extreme value toolkit. R package version 1.62. URL <http://CRAN.R-project.org/package=extRemes>
- [4] Stuart Coles (2001): An Introduction to Statistical Modeling of Extreme Values, Springer-Verlag 2001.
- [5] Lundholm, S. C., Cappelen, J. (2010): Ekstremnedbør i Danmark 1961-2010 – leverance til Koordineringsenhed for Forskning i klimaTilpasning (KFT), DMI Teknisk Rapport 10-17.

Tidligere rapporter

Tidligere rapporter fra Danmarks Meteorologiske Institut kan findes på adressen:
<http://www.dmi.dk/dmi/dmi-publikationer.htm>