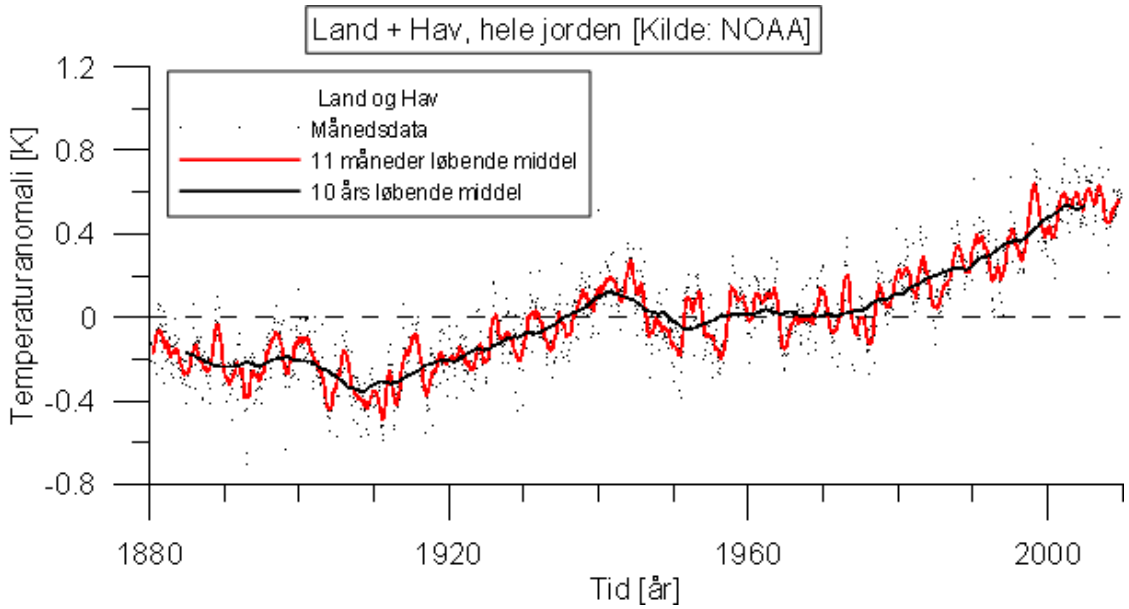
Temperaturdata - Statistiske betragtninger

Af Thomas Wernberg

Lidt på baggrund af diskussionen om hvorvidt temperaturen 2010 er stigende eller faldende og forskellig fra perioden 1990-2000, synes jeg det kunne være interessant at gennemgå data ud fra nogle statistiske betragtninger.

Datakilde: monthly.land_ocean.90S.90N.df_1901-2000mean.dat fra NOAA.

Jeg vil ikke her ind på diskussionen om datakvaliteten og metoden data er behandlet.

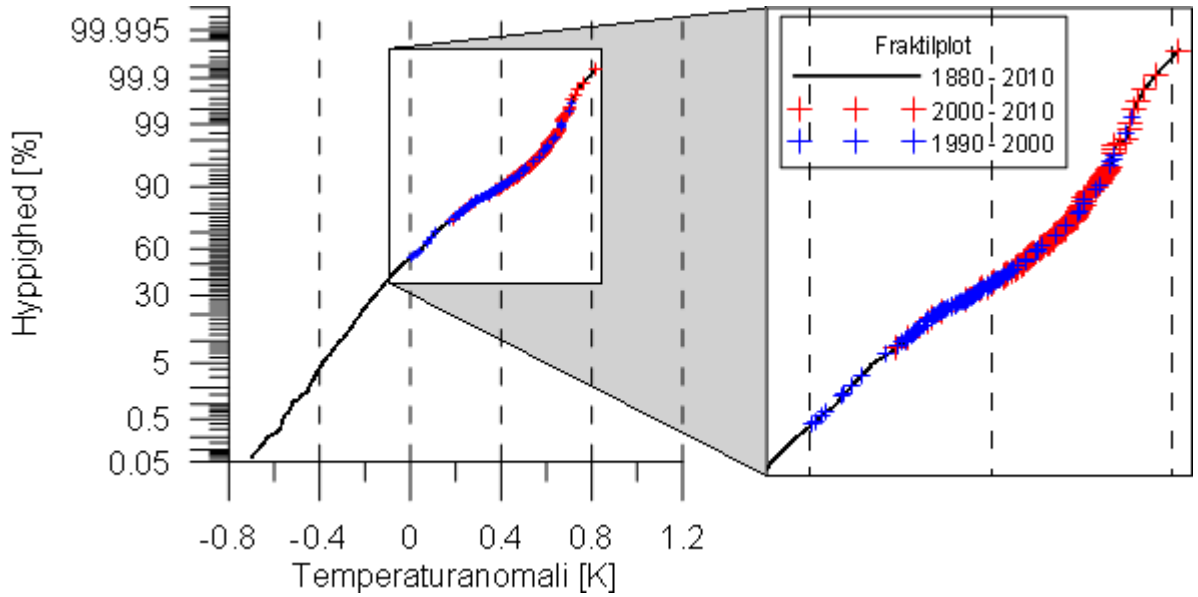


Figur 1:Hav og Land - Hele tidsserien - 1880 - 2010

Figur 1 viser hele tidsserien med de kendte karakteristika.

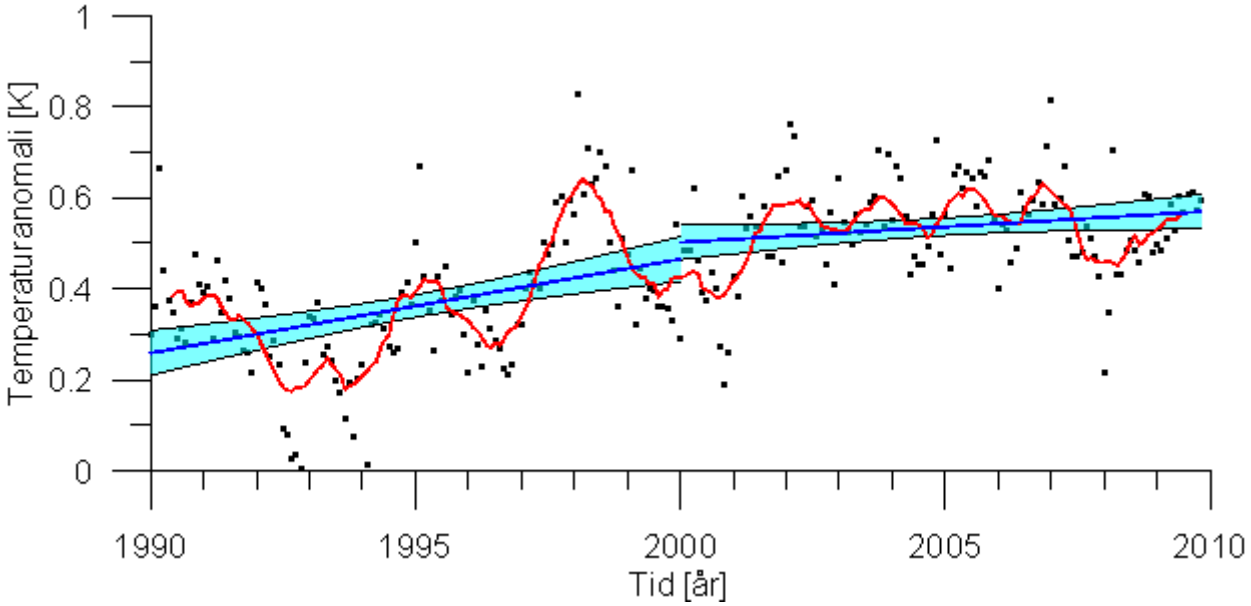
For at kunne behandle data statistisk med den tilgang jeg har tiltænkt, må vi først undersøge om data er normalfordelt.

Figur 2 Viser fraktilplot over alle temperaturdata. Data er tilnærmelsesvis normaltfordelt (de tilpasses en ret linje). Det ses også at de seneste 20 års data ligger i den varme del af datasættet, samt at de sidste 10 års data falder markant ud. Det antages at data er normaltfordelt. Der er dog en 'bule' på kurven i intervallet 0.2 – 0.7 K. Årsagen hertil kan man tage op til diskussion, men en sådanne ændring af data kan tyde på at de høje temperaturer falder i et nyt statistisk system (der er nogle nye forhold, som gør at data ikke længere følger samme fordeling). Vi antager dog her at data er normaltfordelt.



Figur 2:Fraktilplot over alle temperaturdata i perioden 1880 – 2010. Med blå er markeret data i perioden 1990 – 2000 og med rød er vist data for perioden 2000 – 2010.

I 1990'erne er der en en markant stigning, som ser ud til at flade ud i perioden 2000-2010. Figur 3



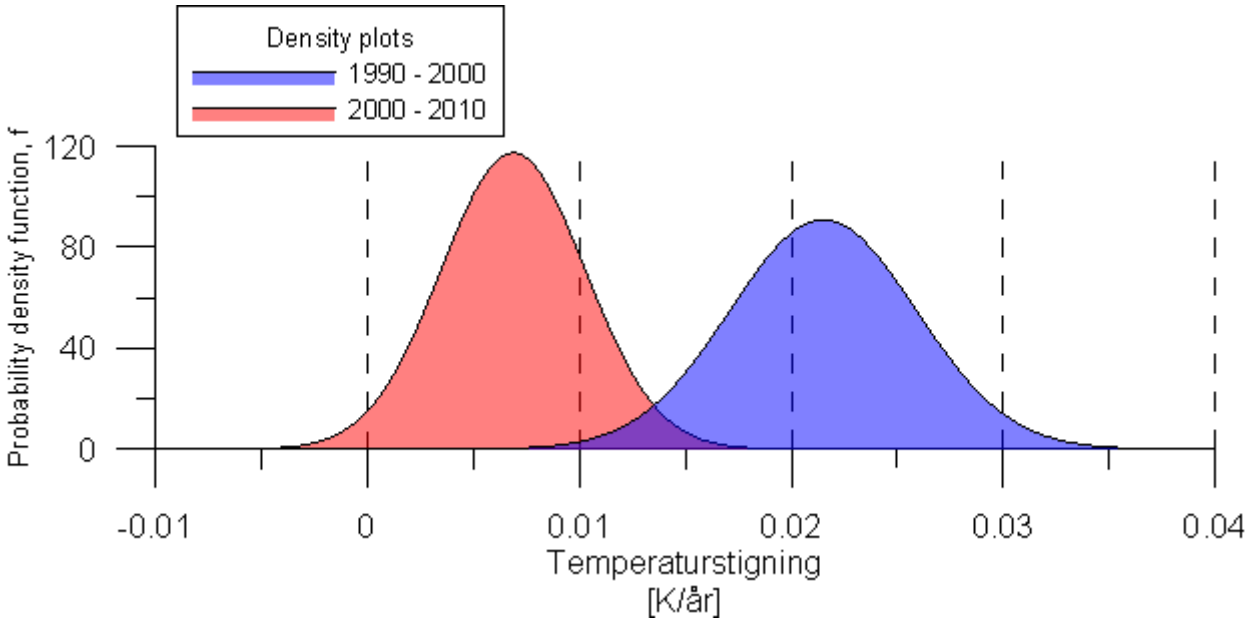
Figur 3:Temperaturdata fra 1990-2010. Lineær regression af data 2 10årsperioder. Med rød ses 11 måneders løbende middel.

viser perioden 1990 – 2010 i detaljer. Der er foretaget en lineær regression af data for de to perioder. Med blå er markeret 95 konfidensinterval på regressionen. Det betyder at man med 95 % sandsynlighed kan placere en ret linje inden for det skraverede område. Det ses tydeligt at perioden 1990 – 2000 er stejlere end perioden 2000-2010. Det ses også at man ikke kan have en linje med negativ hældning i perioden 1990-2000 mens man i perioden 2000 – 2010 godt kan have en negativ hældning.

Tabel 1 viser regressionsanalysen af temperaturdata. Det ses at hældningen i perioden 1990 -2000 er ca. 3 gange større end perioden 2000-2010. Det ses også at variansen på hældningen er lidt større i den første periode.

Tabel 1:Hældning af den lineære regression for de to perioder

Periode	*Hældning, b [K / år]*	*Varians på hældning σ_b^2*
1990-2000	0,0210	0,0043
2000-2010	0,0069	0,0034



Figur 4:Tæthedsfunktion for hældningen af den lineære regression for 1990 - 2000 og 2000 - 2010

Hypoteser:

Hældningen 1990 – 2000 er forskellig for hældningen 2000 – 2010

Her er foretaget en simpel Student's T-test på hældningsdata. Resultatet er at der er 7 % sandsynlighed for at hældningen i de to perioder er identisk.

Temperaturen har været faldende i perioden 1990 – 2000

Denne værdi er så godt som 0

Temperaturen har været faldende i perioden 2000 – 2010

Der er lidt over 2.2 % chance for at temperaturen har været faldende i perioden. Den er altså ikke signifikant.