

## BILAGA IV

### Klusteranalyser för regional indelning

För att få fram en regional indelning av skyfallsstatistik för Sverige användes K-means-klustring (MacKay, 2003) på SMHI:s automatstationer med initialt fyra kluster. Detta resulterade i ett index för varje station, från 1 till 4, för vilken region den ansågs tillhöra.

Varje station modellerades som en punkt i ett 5-dimensionellt vektorrum. De första 3 dimensionerna sattes till GEV-fördelningens ML-parametrar på årsmax-data. Dimensioner 4 och 5 sattes till RT90-koordinaterna för stationen. Eftersom detta gjordes på alla 128 stationer så har vi 128 värden för varje dimension. Stationer med färre än 15 tillgängliga årsmax togs bort från analysen, då dessa ansågs ge för osäker parameterskattning.

Efter detta steg normerades alla fem dimensioner, d.v.s. först subtraherades medelvärdet, och sedan dividerades detta med standardavvikelsen. Detta gjorde att varje dimension blev enhetslös, vilket gör att vi kan jämföra avstånd mellan parametervärden med avstånd i rummet (RT90-koordinaterna).

För att ha större kontroll över hur tungt geografiskt avstånd viktas jämfört med parameteravstånd så skalades avståndskoordinaterna med en faktor 3. Detta för att få bättre balans mellan hur lika två stationer är i sina GEV-parametrar relativt deras inbördes avstånd.

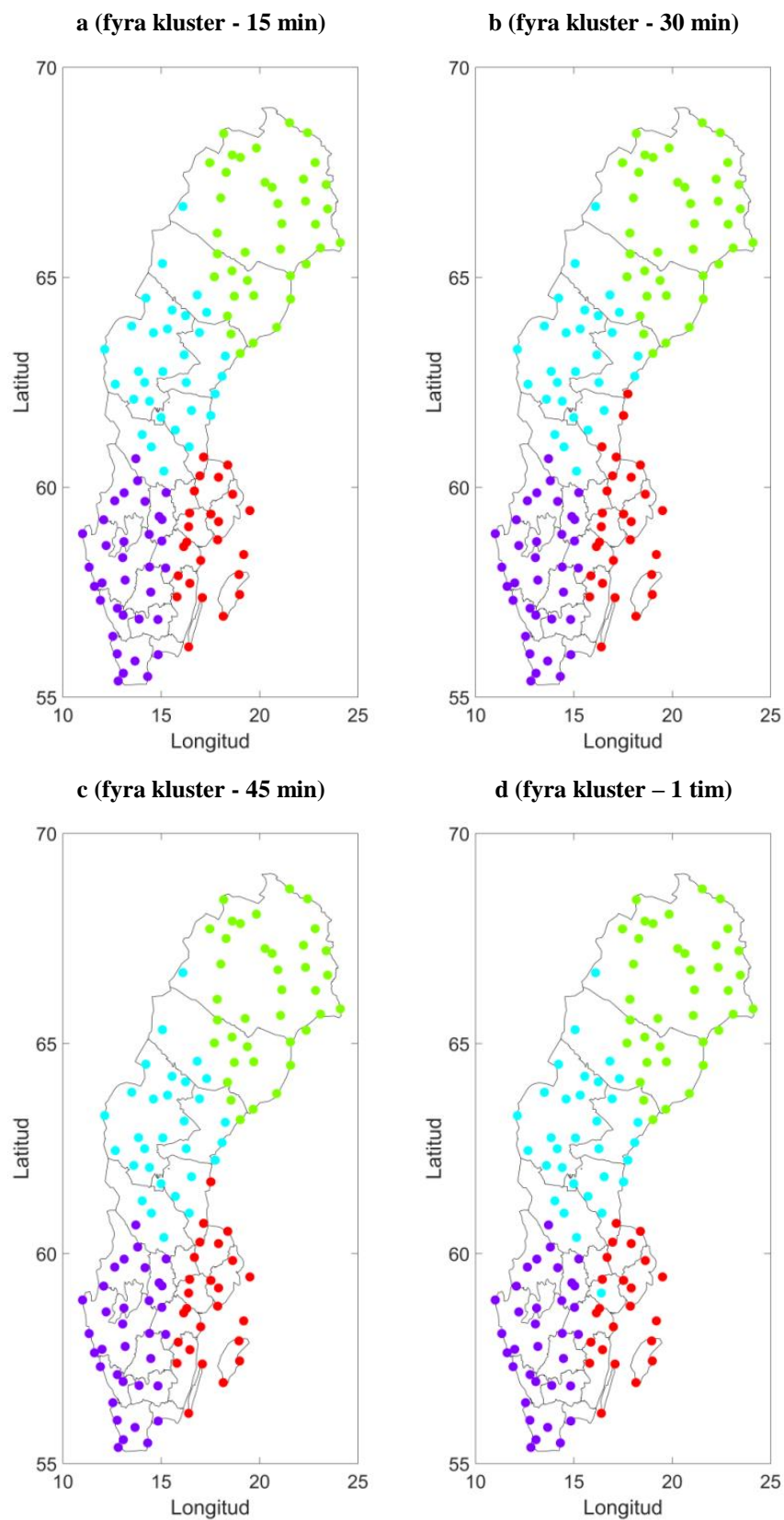
Resultaten av denna klusteranalys presenteras i form av kartor nedan (Figur 1a-f) som indikerar att regionerna är överlag stabila för varaktigheter mellan 15 min och 6 tim.

För att undersöka inverkan av antalet kluster gjordes ett antal tester, vilka gav följande resultat:

- Vid indelning i tre kluster slås de två sydliga regionerna samman till en.
- Vid indelning i fem kluster behålls i princip gränserna från fyra kluster, men det nordliga klustret delas in i en sydvästlig och en nordöstlig del (Figur 1g).
- Vid indelning i sex kluster delas det nordliga klustret enligt ovan och dessutom delas det sydvästliga klustret in i en västlig och en östlig del (Figur 1h).

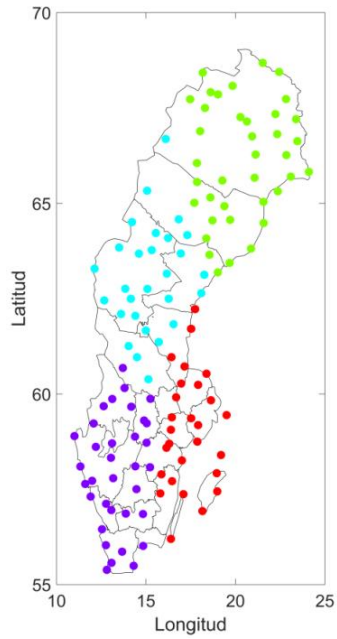
### Referenser

MacKay, David (2003). "Chapter 20. An Example Inference Task: Clustering" (*PDF*). Information Theory, Inference and Learning Algorithms. *Cambridge University Press*. pp. 284–292. ISBN 0-521-64298-1. MR 2012999.

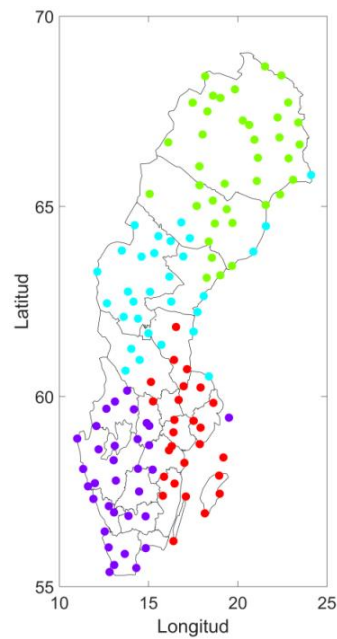


Figur 1. Resultat av klusteranalys med fyra kluster, varaktigheter 15 min (a), 30 min (b), 45 min (c) och 1 tim (d).

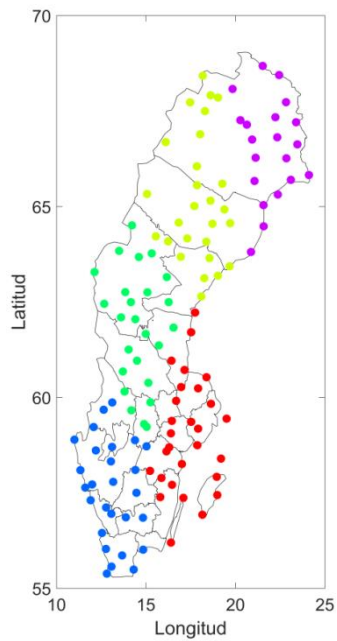
**e (fyra kluster – 2 tim)**



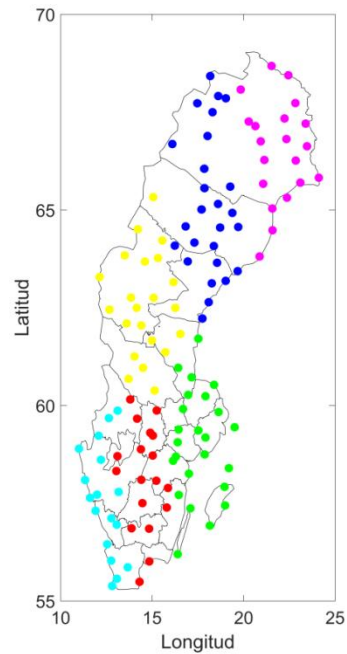
**f (fyra kluster – 6 tim)**



**g (fem kluster -30 min)**



**h (sex kluster – 15 min)**



*Figur 1 (forts.). Resultat av klusteranalys med fyra kluster, varaktigheter 2 tim (e) och 6 tim (f), fem kluster, varaktighet 30 min (g) och sex kluster, varaktighet 15 min (h).*