NORGES VASSDRAGS- OG ENERGIVERK
VASSDRAGSDIREKTORATET
HYDROLOGISK AVDELING

JÄMFÖRELSE AV NÅGRA METODER FÖR INTER- OCH EXTRAPOLATION

OPPDRAGSRAPPORT
17 - 87

OPPDRAGSRAPPORT 17-87

Rapportens tittel:	Dato: 1987-11-10
JÄMFÖRELSE AV NÅGRA METODER FÖR INTER- OCH EXTRAPOLATION	Rapporten er: Åpen
	Opplag: 30

Saksbehandler/Forfatter:	Ansvarlig:
Torbjørn Jutman Datakontoret	Nils Roar Sælthun

Oppdragsgiver:

SAMKJØRINGEN AV KRAFTVERKENE I NORGE

Sammanfattning:

Ny metodik för inter- och extrapolation av avrinningsserier har jämförts med de metoder som idag är i användning inom NVE, - Samkjøringen och Statkraft.

Regressionssamband används för närvarande vid extrapolation av serier och skalering vid interpolation.
Den nya metodiken innebär utnyttjande av empiriska ortogonala funktioner vid båda tillämpningarna.

Undersökningen är av begränsad natur och främst ägnad att spåra indikationer på att metoderna skiljer sig åt vad det gäller prestanda och hanterbarhet.
Serier från den nordligaste av regionerna enligt Samkjøringens regionindelning har använts för analys och verifiering. Speciellt har önskan varit att kontrollera de simulerade volymerna för olika säsonger.

Resultaten visar på små skillnader prestandamässigt sett. Generellt är överensstämmelsen relativt dålig vid låga avrinningsvärden. Detta gäller alla de prövade metoderna.

Stor skillnad metoderna emellan föreligger när det gäller det praktiska handhavandet. Utnyttjande av empiriska ortogonala funktioner ger väsentligt förenklade rutiner samtidigt som resultaten blir konsistenta.

NORGES
VASSDRAGS- OG ENERGIDIREKTORAT
BIBLIOTEKET

Forord

Metoder for ekstrapolasjon og interpolasjon er viktige for å utnytte maksimalt informasjon fra hydrologiske målestasjoner. Operative rutiner bør revideres fortløpende i forhold til nye landevinninger. Denne rapporten behandler bruk av såkalte Empiriske ortogonale funksjoner (EOF) for ekstrapolasjon av måleverdier i tid og rom. Metoden er sammenliknet med andre metoder som er i bruk idag ved NVE, Samkjøringen og Statkraft.
Prosjektet er et samarbeidsprojekt mellom Samkjøringen og NVE, Hydrologisk avdeling.

Oslo, oktober 1987

Arne Tollan
avdelingsdirektør

INNEHÅLL

		Sida
1.	ALLMÄNT	3
2.	EMPIRISKA ORTOGONALA FUNKTIONER	3
3.	FÖRLÄNGNING AV SERIER	6
4.	INTERPOLATION	8
5.	LIKHETER OCH OLIKHETER MELLAN METODERNA	14
6.	VILKA ÅTGÄRDER KRÄVS FÖR ATT GÖRA EOF-METODEN OPERATIV?	16
7.	REFERENSER	22

1. ALLMÄNT

I samband med en presentation av Empiriska Ortogonala Funktioner (EOF) som redskap vid förlängning/komplettering av serier alternativt interpolation, initierades ett projekt för utvärdering av metodens förmåga i jämförelse med idag använda metoder vid NVE, Statkraft och Samkjøringen.
Projektet är ett samarbete mellan Samkjøringen och NVE.

Verifiering av metodernas förmåga har skett via applicering på "tilsigsdata" (veckovärden) från region 4 enligt Samkjøringens regionindelning.
1941 - 1960 valdes som testperiod och maximala antalet serier som täcker hela den perioden är omkring 20. Ytterligare ett antal stationer som ej täcker hela perioden har använts för verifiering.

Presentationen inleds med en kortfattad presentation av EOF - metoden.

Metodernas förmåga med avseende på extrapolation respektive interpolation studeras dels via jämförelser av simulerade serier med observerade verifieringsserier, dels genom kontroll av volymer över olika säsonger för dessa serier.
I formuleringen av projektet uttrycktes speciellt intresse för kontroll av volymerna under vinterperioden.

EOF-metoden är för närvarande ej i operativt skick. Avslutningsvis kommenteras därför nödvändiga åtgärder för att uppnå detta.

2. EMPIRISKA ORTOGONALA FUNKTIONER

Metoden innebär en transformering av existerande serier , exempelvis avrinningsserier, till en uppsättning nya serier.

Den teoretiska delen av metodbeskrivningen såväl som ytterligare exempel på praktiska applikationer kan inhämtas från bland andra Bartlein (1982), Holmström (1969) och Jutman (1985).

Antag att vår utgångspunkt är serierna $Q_i(t)$, $i=1,\ldots,N$ för en given tidsperiod (0,T). En expansion i EOF har då formen :

$$Q_i(t) = \sum_{n=1}^{M} h_{ni}\, B_n(t) \ , \ i=1,\ldots,N \qquad (1)$$

där h_{ni} är viktkoefficienter och $B_n(t)$ är de nya tidsserierna,här kallade amplitudfuntioner.

Amplitudfunktionerna är inbördes ortogonala tack vare de krav som ställs på expansionen och empiriska i den meningen att de bygger på de ursprungliga funktionerna = tidsserierna.

Expansionsformeln ovan ger att till varje ursprunglig serie hör en uppsättning viktkoefficienter med en koefficient för varje amplitudfunktion. Viktkoefficienterna kan sägas motsvara rumsvariationen i de ursprungliga serierna.

Uppsättningen av amplitudfunktioner är däremot gemensam för alla de ursprungliga serierna och kan sägas beskriva tidsvariationen. Analyserar vi tillgängliga serier i en viss regimregion, så kommer vi att få en uppsättning amplitudfunktioner som då kommer att gälla för varje punkt i den regionen för den analyserade tidsperioden.

Exempel på amplitudfuntioner visas i figur 1.

Här redovisas de tre första amplitudfunktionerna baserade på EOF-analys av veckovärden för perioden 1941 - 1944.

Ytterligare en egenskap hos de empiriska ortogonala funktionerna framgår här, nämligen koncentrationen av av information.
Den största delen av de ursprungliga seriernas varians finns som synes i den första funktionen, en mindre del i den andra o.s.v. Man kan alltså med ett litet antal av EOF beskriva en stor del av variansen i det ursprungliga materialet.

Man bör vara försiktig med att lägga in alltför långtgående tolkningar av de enskilda amplitudfunktionernas utseende.
Det framgår dock relativt klart att den första funktionen i stort beskriver säsongsvariationen. Man ser i figuren att viktkoefficienterna för den första amplitudfunktionen genomgående har negativt förtecken.
Med lite god vilja kan man säga att den tredje funktionen främst påverkar recessionens utseende allt efter det tecken som motsvarande viktkoefficient har.

Generellt gäller att EOF-analysen ej nödvändigtvis behöver bygga på tidsserierna i ursprunglig form. Det kan många gånger visa sig fördelaktigt och mera effektivt att använda sig av olika transformer av serierna.

De transformer som här ligger närmast till hands är :

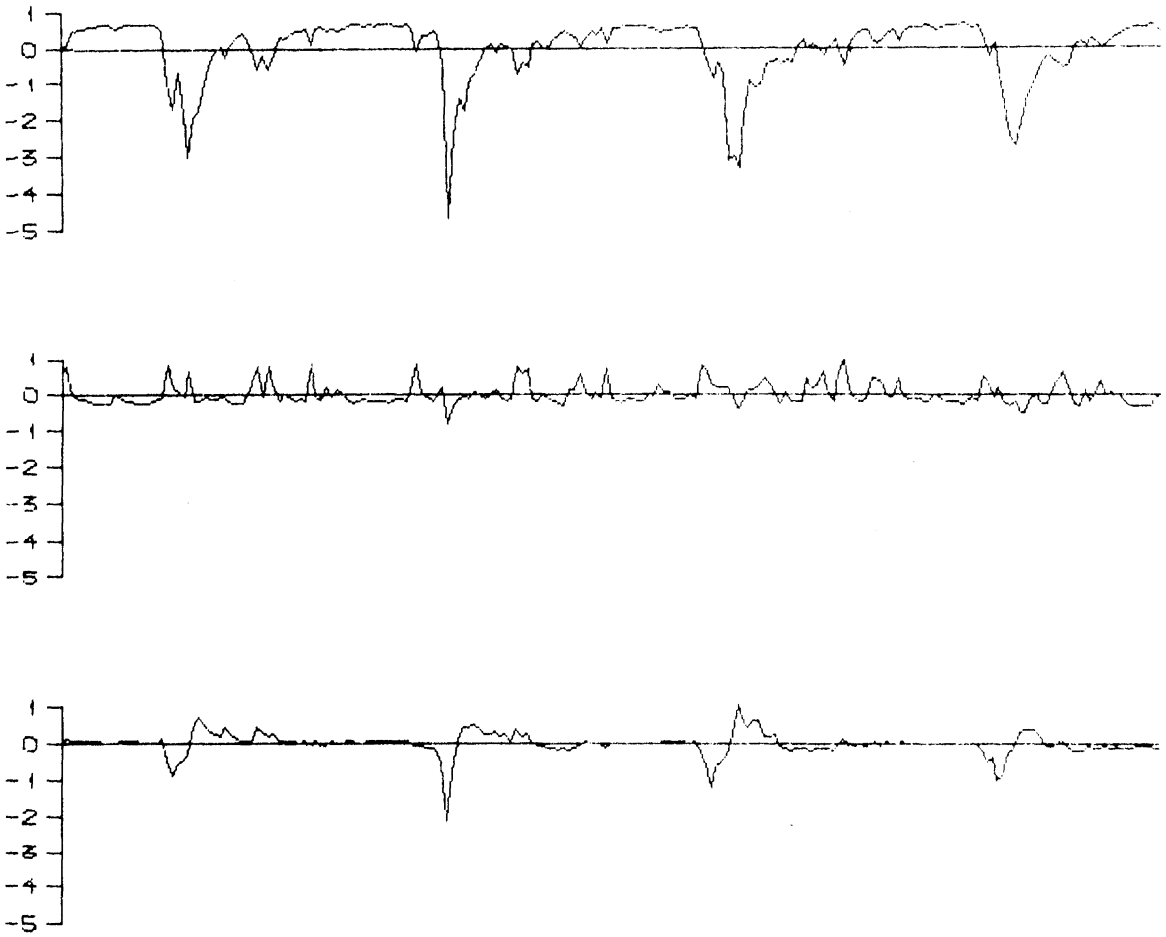
1. Subtraktion med respektive series medelvärde.

2. Standarisering av serierna, d.v.s. subtraktion av medelvärdet och division med standardavvikelsen.

3. Logaritmtransformering (eventullt i kombintion med någon av ovanstående former).

I exemplet i figur 1 har alternativ 2 använts.

Vid användning av EOF för inter- eller extrapolationsändamål bör man begränsa sig till en rumskala som innefattar avrinningsområden större än 100 km^2 och upp till 2000 km^2. Angivna gränser är naturligtvis mycket ungefärliga.



Figur 1. Exempel på amplitudfunktioner.

Begränsningen vad det gäller områdesstorlek grundas på två faktorer:

- Avrinningens struktur för ett litet område präglas i mycket hög grad av områdets specifika egenskaper (geologi,topografi,etc.).

- Ett stort område kan ofta sträcka sig över två eller flera regimregioner, vilket gör en beskrivnig med EOF ineffektiv.

I projektet har därför framför allt verifieringsstationerna i största möjliga utsträckning hållits inom angivna gränser.

3. FÖRLÄNGNING AV SERIER

Förutsättningen är att vi har en serie som ska förlängas till att omfatta en viss tidsperiod. Vi förutsätter att EOF-analysen är genomförd för regionen och då har vi amplitudfunktionerna, vilka är gemensamma för regionen och vi behöver skattningar på motsvarande viktkoefficienter.

Man kan teoretiskt visa att viktkoefficienterna i det här fallet är lika med korrelationskoefficienterna mellan den existerande serien och respektive amplitudfunktion.
Ett begränsat antal amplitudfunktioner används. Här har antalet satts till 4 , men generellt bör detta antal begränsas av huruvida korrelationskoefficienten är signifikant eller ej.

EOF-analys genomfördes på de 21 serier av "tilsigsdata" från region 4 för perioden 1941 - 1960 .

De amplitudfunktioner som man erhåller kan så användas för förlängning av serier till att omfatta hela denna period. En förutsättning är dock att den aktuella serien täcker någon del av perioden. Denna del används för bestämning av viktkoefficienterna.
Hur lång den existerande serien måste vara för att en förlängning ska bli meningsfull med den här metoden beror bl.a. på variabiliteten inom regionen, men en c:a 5 år lång utgångsserie bör vara tillräcklig.

Det visade sig vid försök med tidigare nämnda former av transformeringar att den som gav de bästa resultaten innebar subtraktion av medelvärdet i kombination med logaritmtransformerig. Detta medför att man för den serie som ska förlängas måste ha en skattning av seriens medelvärde i tillägg till skattningen av viktkoefficienterna enligt ovan.

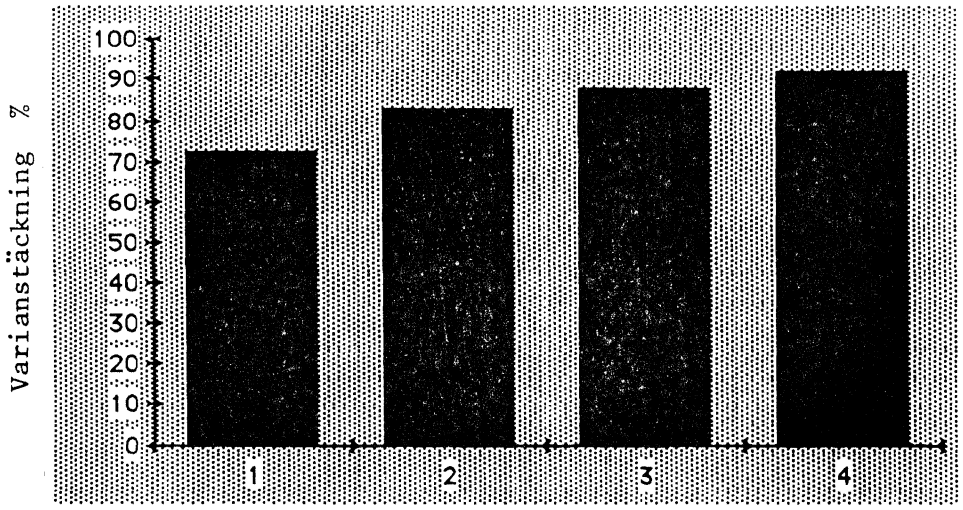
Det har här förutsatts att vi känner medelvärdet för den förlängda serien.

Var och en av de ursprungliga 21 serierna kan återskapas fullständigt om vi använder oss av samtliga 21 amplitudfunktioner. Vi är dock intresserade av att utnyttja ett mindre antal.

Andel förklarad varians utgör ett mått på hur väl vi kan återskapa de ursprungliga serierna.

Figur 2 ger varianstäckningen i % då vi använder upp till fyra amplitudfuntioner. Väl att märka gäller detta analys av serierna med transformation enligt ovan och utfallen är genomsnitt av återskapad varians för samtliga ursprungliga serier.

Det framgår av figuren att en dominerande del av den ursprungliga variansen ligger i den första amplitudfunktionen.



Figur 2. Varianstäckning (%) i relation till använt antal amplitudfunktioner.

Tre serier, som ej inkluderats i själva EOF-analysen, har använts för verifiering. För beräkning av korrelationskoefficienterna har använts en period av fem år och förlängning av respektive serie görs enligt :

$$S_k(t) = m_k + r_{1k}*B_1(t) + r_{2k}*B_2(t) + r_{3k}*B_3(t) + r_{4k}*B_4(t) \quad (2)$$

där $S_k(t)$ är simulerad serie nr k, m_k medelvärdet för den period som avses för serie k, r_{ik} korrelationskoefficienten beräknad på serie S_k och amplitudfunktion B_i och slutligen $B_i(t)$ är de för regionen gemensamma amplitudfunktionerna.
Förlängningen kan omfatta den period som EOF-analysen täcker men för tydlighetens skull visas endast 4 år i figurerna.

Resultaten visas i figurerna 3a - 3c.

I figurerna visas också resultaten från ordinarie extrapolationsmetod, d.v.s. regression mot ett antal stationer. Som input till regressionen har använts samma uppsättning stationer som till EOF-analysen. Maximalt antal stationer i regressionsekvationen sattes till fem.

Figurererna ger att skillnaderna mellan de två metodernas förmåga i dessa tre exempel är minimal.

Utvärdering har också skett via kontroll av simulerade volymer för olika säsonger samt för kalenderår. Följande säsongsindelning har använts:

* 1/5 - 31/7

* 1/8 - 30/9

* 1/10- 30/4

Figur 4a visar jämförelse mellan simulerad respektive observerad volym för de tre verifieringsstationerna vid förlängning av serierna med EOF. Enbart vintervolymerna redovisas för två av stationerna.
Figur 4b visar motsvarande resultat vid användning av regression för förlängning.

Vintervolymerna har vid användning av EOF underskattats i ett av de tre fallen och vid användning av regression har en överskattning skett i två fall. En viss skillnad mellan metoderna kan ses i det här momentet men man ska vara medveten om att båda metoderna kan uppvisa inslag av stora procentuella avvikelser vid små flöden.

Det är väl värt att påpeka att tre verifieringsstationer inte är tillräckligt antal för att statistiskt säkerställa eventuella skillnader metoderna emellan.
Avsikten har här varit att via en kort, översiktlig jämförelse få indikationer på om väsentliga skillnader föreligger.

4. INTERPOLATION

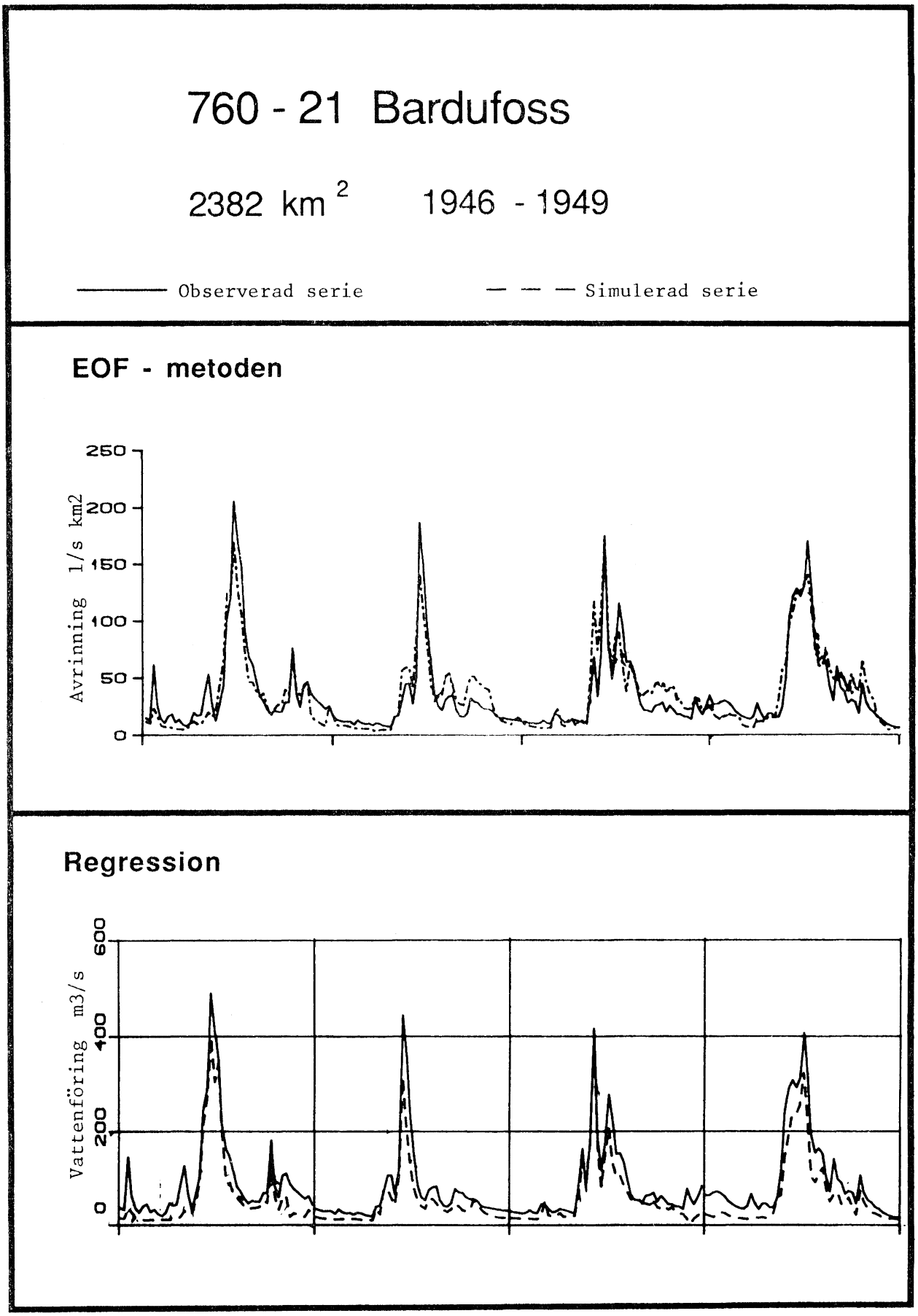
Användning av EOF för interpolation innebär att man utnyttjar att amplitudfunktionerna är gemensamma för den analyserade regionen och att man försöker få fram en rumsbeskrivning av viktkoefficienterna.

Beskrivningen av viktkoefficienternas variation i rummet kan exempelvis genomföras via regressionssamband. Man kan utnyttja information om de avrinningsområden som ingår i EOF-analysen som kan tänkas vara av betydelse för rumsbeskrivningen.

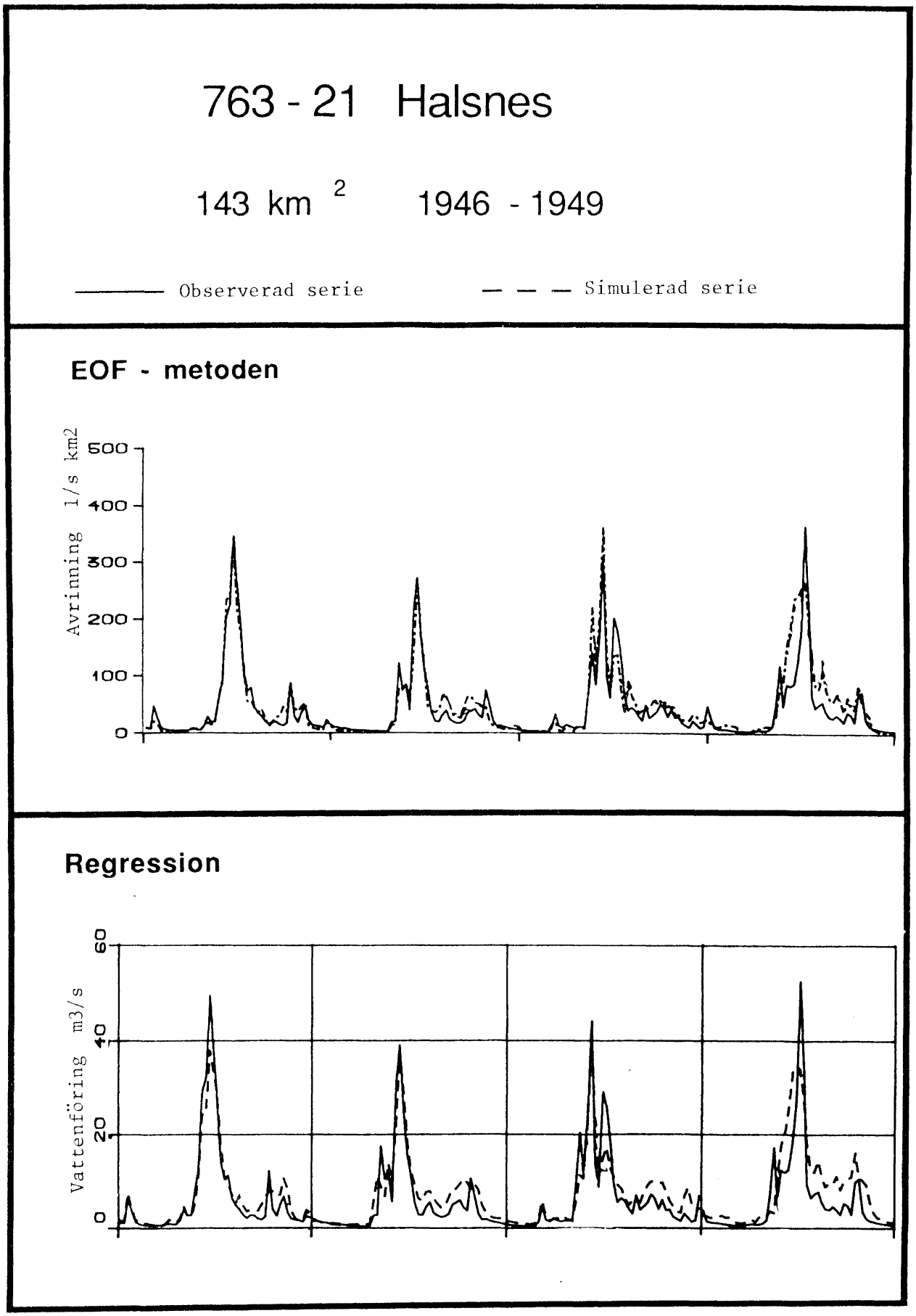
Här har EOF-analys genomförts för de 17 stationer i region 4 som har någorlunda kompletta uppgifter vad avser stationsparametrar. Som oberoende variabler i beskrivningen av viktkoefficienternas rumsvariation testades:

- Areal
- Sjöprocent
- Effektiv sjöprocent
- Relief
- Medelhöjd

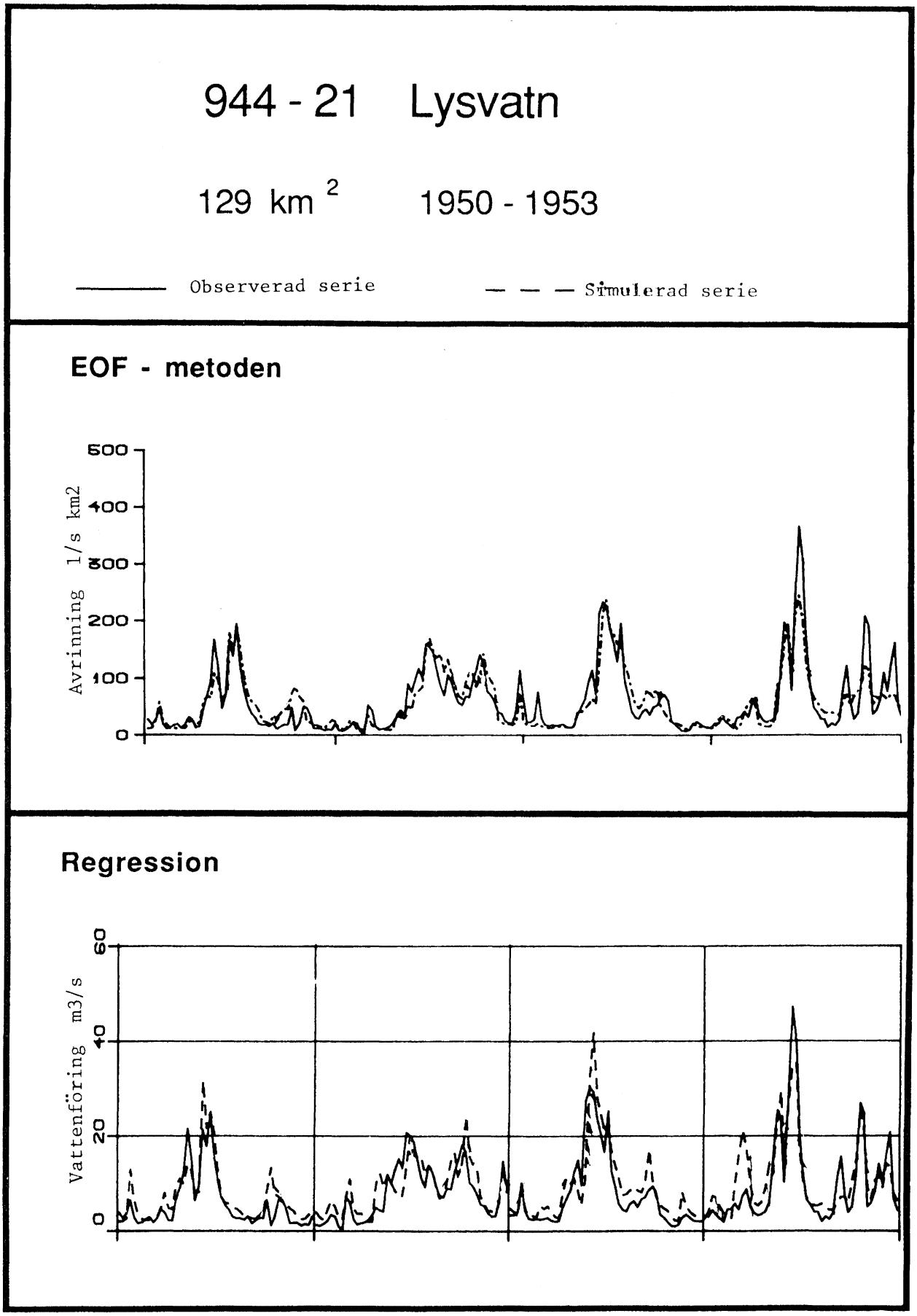
Av tidsmässiga skäl kunde ej position i form av rätvinkliga koordinater inkluderas. Dessa är angivna som longitud och latitud i stationsförteckningen. Dessutom anger givna koordinater positionen för aktuell station och ej, som önskas här, positionen för avrinningsområdets tyngdpunkt. Man kan dock vänta sig att just dessa



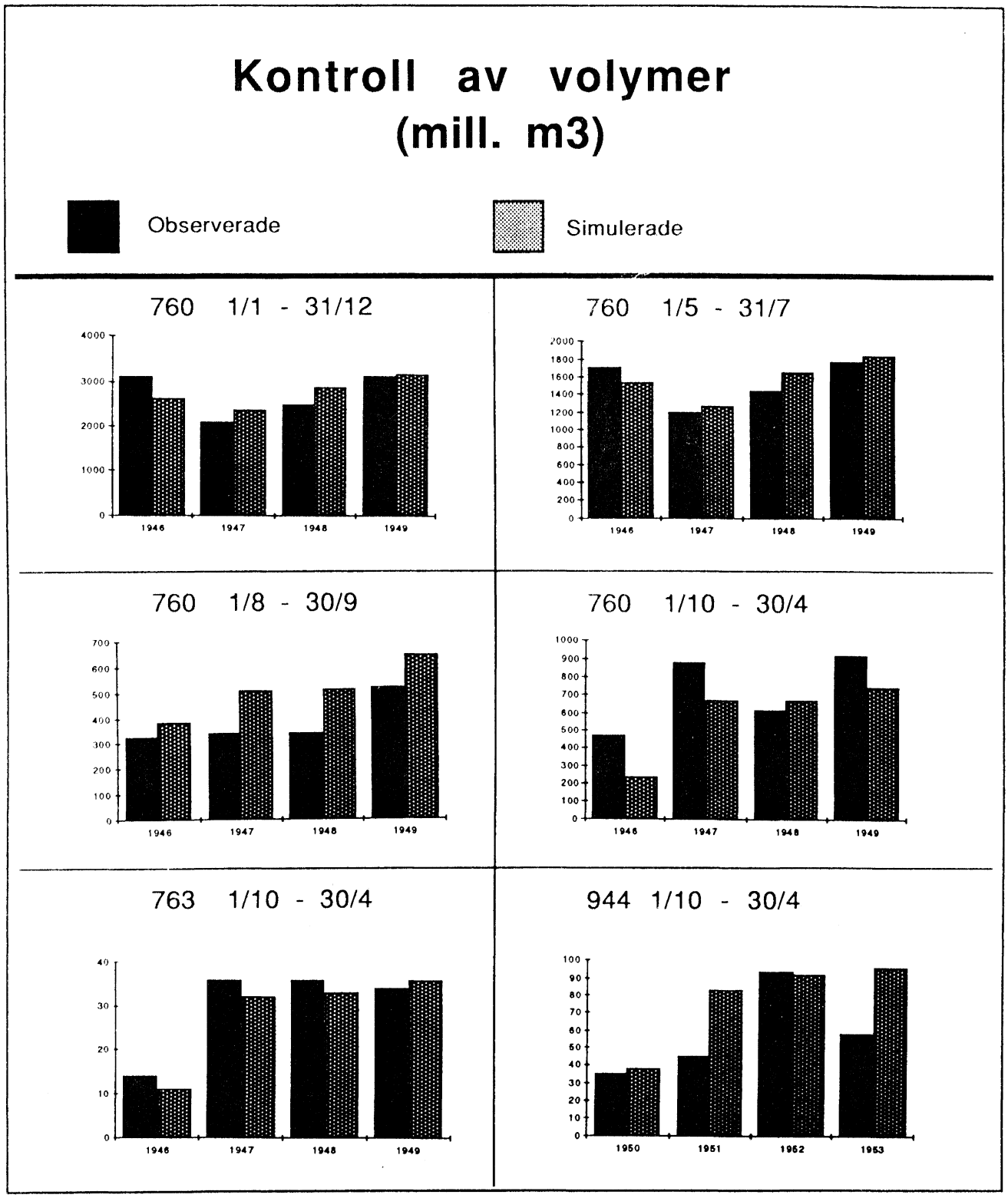
Figur 3a. Jämförelse av metoder vid extrapolation. Station 760.



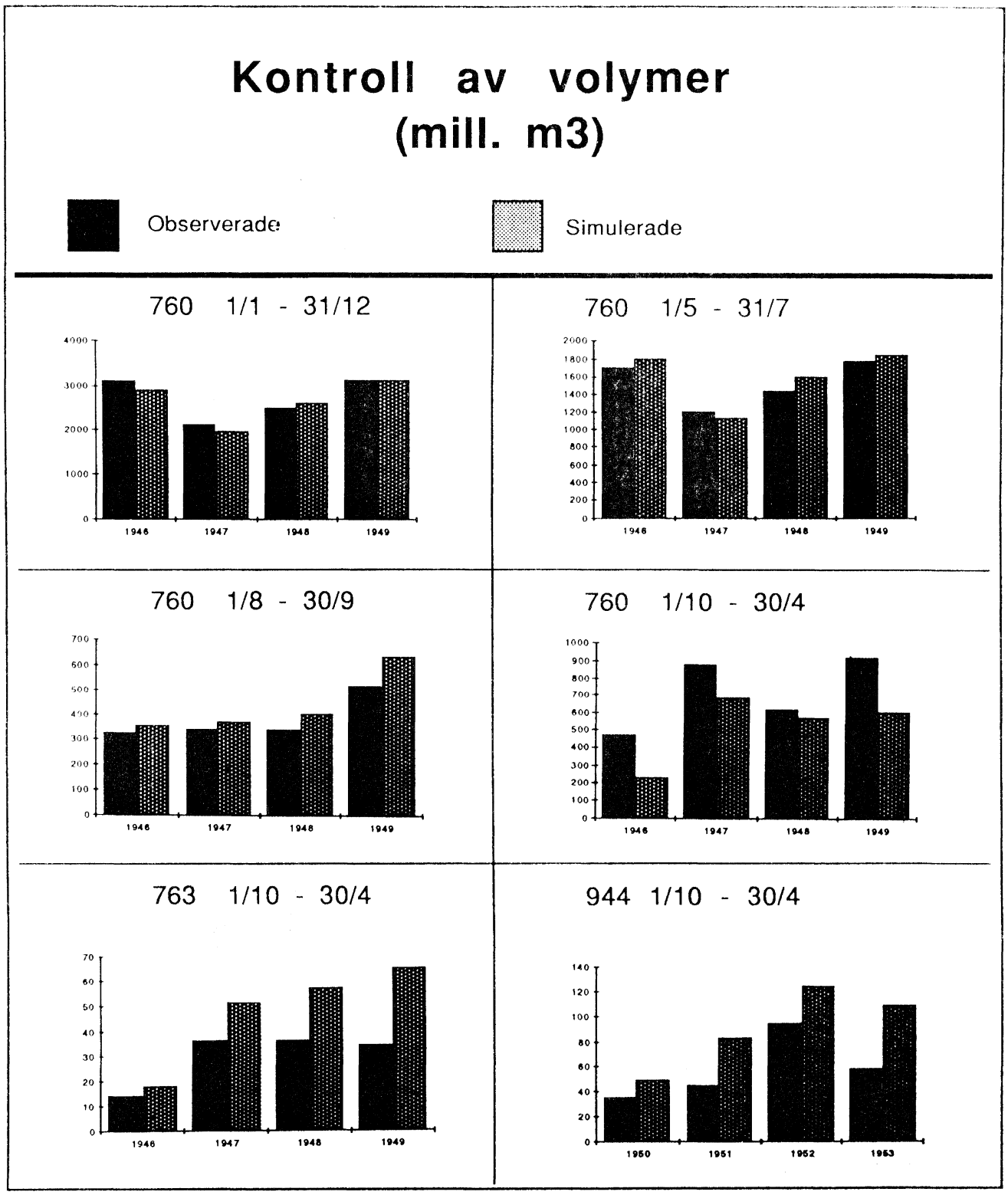
Figur 3b. Jämförelse av metoder vid extrapolation. Station 763.



Figur 3c. Jämförelse av metoder vid extrapolation. Station 944.



Figur 4a. Kontroll av volymer vid extrapolation med EOF.



Figur 4b. Kontroll av volymer vid extrapolation med regression.

blir väsentliga i beskrivningen av rumsvariationen hos viktkoefficienterna.

I interpolationen har endast tre amplitudfunktioner använts. Det är av tidigare erfarenhet från svenska förhållanden oftast inte möjligt att få en signifikant beskrivning av viktkoefficienterna för fler amplitudfunktioner.

Alla de ovan nämnda oberoende variablerna kom med i åtminstone något av regessionssambanden.

Med kännedom om medelvärdet för den önskade serien och de enligt ovan beräknade viktkoefficienterna, kan en simulering ske enligt samma mönster som vid förlängning av serier (Se kap. 3).

Resultaten av simuleringarna ges i figurerna 5a - 5c.

Interpolering enligt ordinarie metod har skett via skalering av vattenföringsdata från station med jämförbara egenskaper.Resultaten visas i den nedre halvan av de tre figurerna.

Jämförelse av volymer har utförts på samma sätt och med samma säsongsindelning som vid extrapolationen. Resultaten ges i figurerna 6a och 6b.

Det allmänna intrycket är attskalering ger något stabilare värden i de låga registren medan skillnaderna i övrigt är små.

Vintervolymerna är även här osäkra och det kan konstateras att EOF-metoden i en del fall ger kraftiga avvikelser.

5. LIKHETER OCH OLIKHETER MELLAN METODERNA

Den tänkta användningen av den extrapolerade eller interpolerade serien i hög grad är avgörande för bedömningen av den underliggande metodens lämplighet. Bedömningen har här inriktats mot kontroll av generell överensstämmelse mellan simulerad och observerad serie och mot volymjämförelser.

Det kan när det gäller förmågan att förlänga serier konstateras att de båda metoderna är i stort sett likvärdiga i sin prestationsförmåga. Regressionsmetoden visar i den begränsade undersökningen en viss tendens till överskattning av vintervolymerna.

Det finns dock skillnader mellan metoderna som i stället hänför sig till själva handhavandet av de operativa rutinerna (de tänkta rutinerna, vad det gäller EOF-metoden).

Regressionsmetoden kräver alltid tillgång till samtliga serier som kan tänkas komma i användning, d.v.s. samtliga serier i det område för vilket beräkningar kan komma i fråga, inbegripet ett visst randområde.

Användaren måste själv göra göra viss bedömning av vilka serier som kan vara lämpliga att ta med i analysen.
Separat företas sedan en simulering på grundval av resultaten från regressionsanalysen.

Vid användning av EOF-metoden behöver man enbart tillgång till tre eller fyra amplitudfunktioner. Dessa tas fram vid analyser som i stort sett genomförs en gång för alla.

Amplitudfunktionerna kräver litet lagringsutrymme.

Användaren behöver i stort sett som input ge den serie som ska förlängas tillsammmans med information om önskad tidsperiod och tidsupplösning.
Metoden blir mycket enkel att använda och maskintiden blir kort.
En annan fördel är att alla användare får samma resultat med samma serie och förutsättningar med avseende på tidsperioder. Detta senare är inte givet vid användning av regressionsmetoden.

Med avseende på interpolering jämförs EOF-metoden med resultat från skalering av serier.
I stort har båda metoderna även här gett likartade resultat i testsimuleringarna.
I något av fallen har EOF-metoden gett sämre resultat. Under vinterperioden med låga flöden är precisionen generellt sämre. Skalering har här gett något bättre resultat.

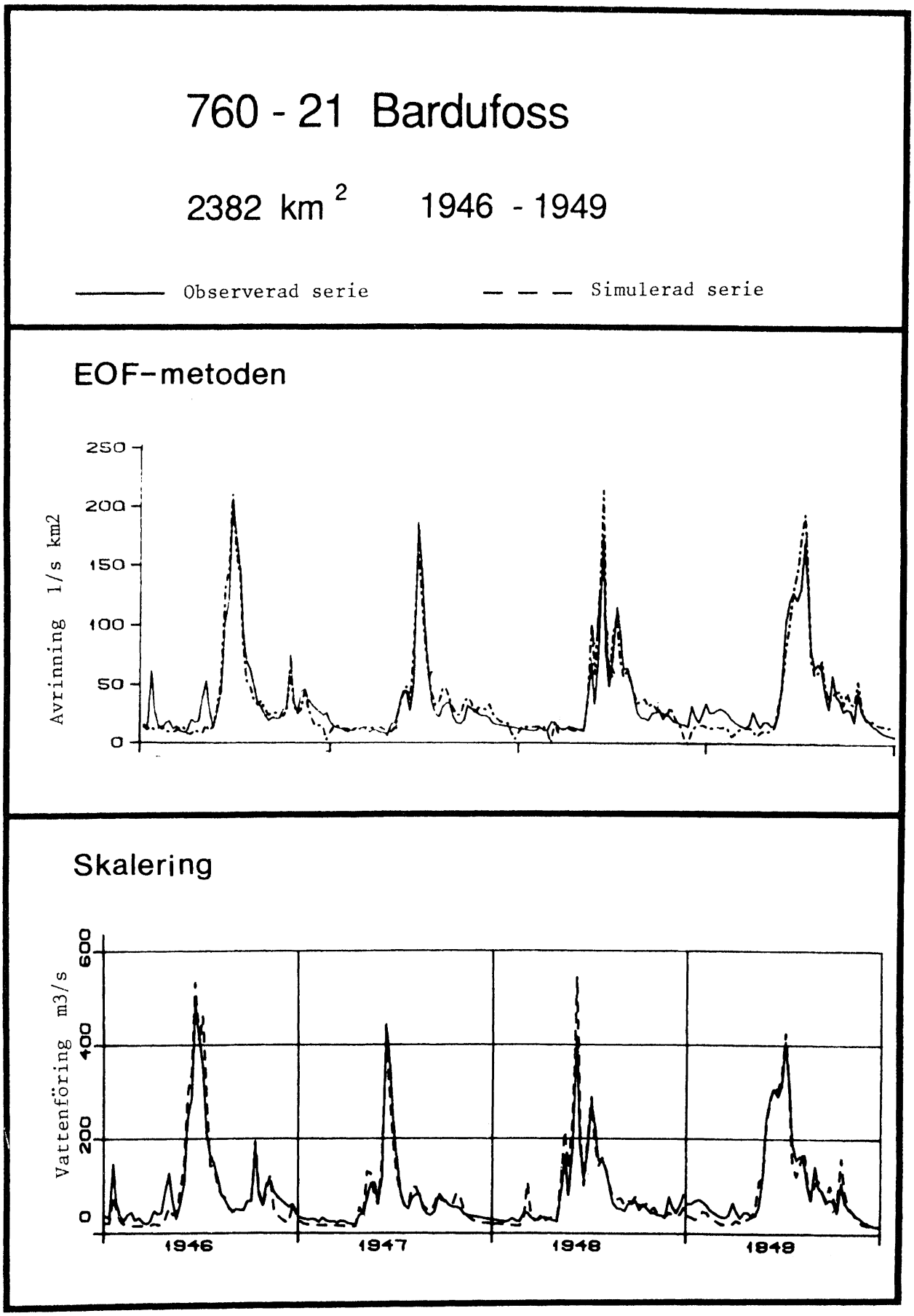
En svensk undersökning baserad på en analys av månadsvärden gav emellertid inte några problem vid användning av EOF för interpolation. Det är dock klart att en lägre grad av tidsupplösning inte är lika utslagsgivande och att variationen i rummet är mindre på den svenska sidan.

Det allmänna intrycket från jämförelsen är att metoderna presterar ungefär lika goda resultat vid förlängning liksom vid interpolation.

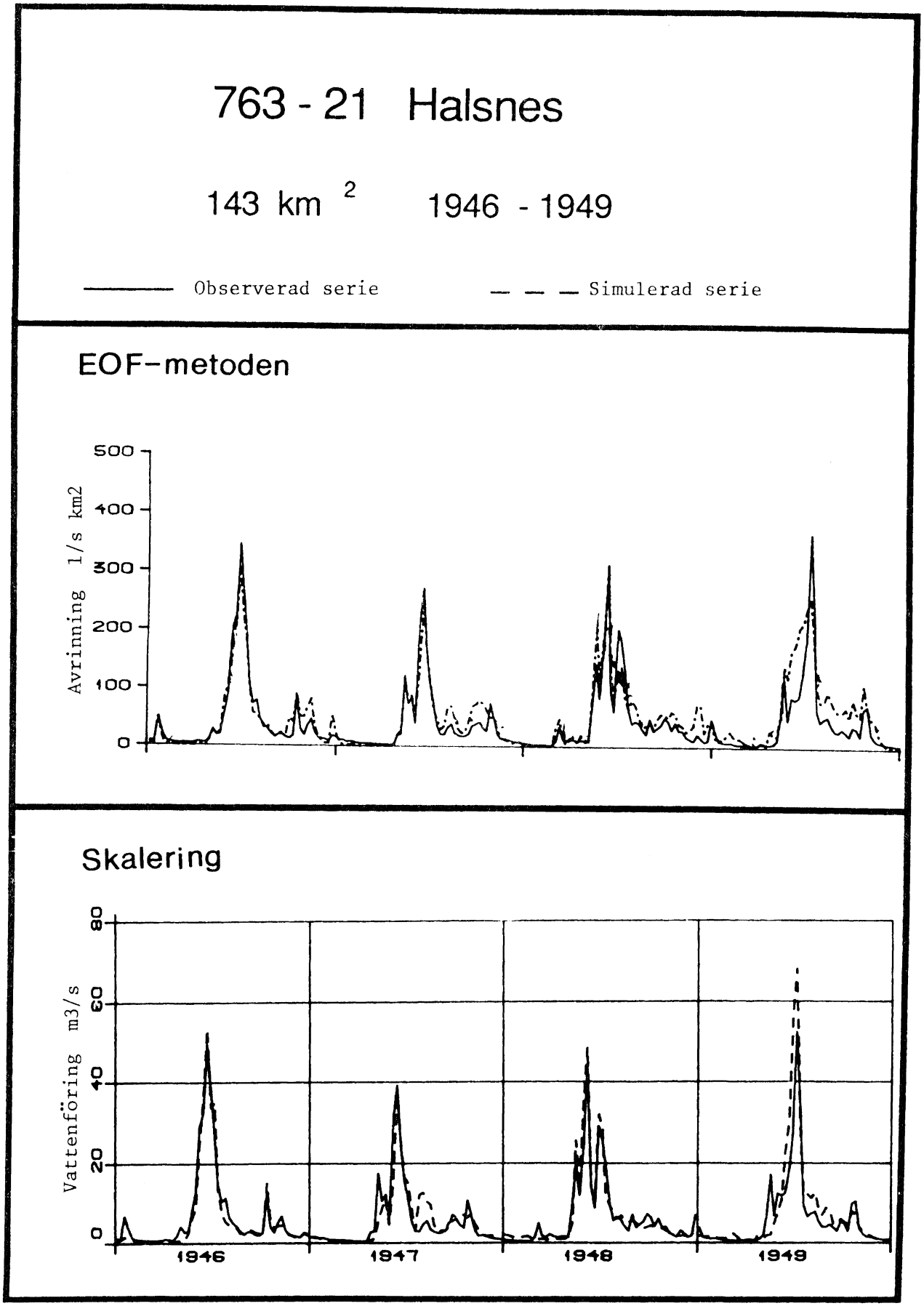
Det föreligger däremot större skillnader mellan metoderna om man ser på handhavandet av rutinerna än de som man kan finna vid jämförelse av prestationsförmåga.
EOF-metoden ger klart enklare rutiner som kräver mindre datorkapacitet. Metoden ger också till skillnad från regression och skalering konsistenta resultat. Resultatet blir ej beroende av användaren.

6. VILKA ÅTGÄRDER KRÄVS FÖR ATT GÖRA EOF-METODEN OPERATIV ?

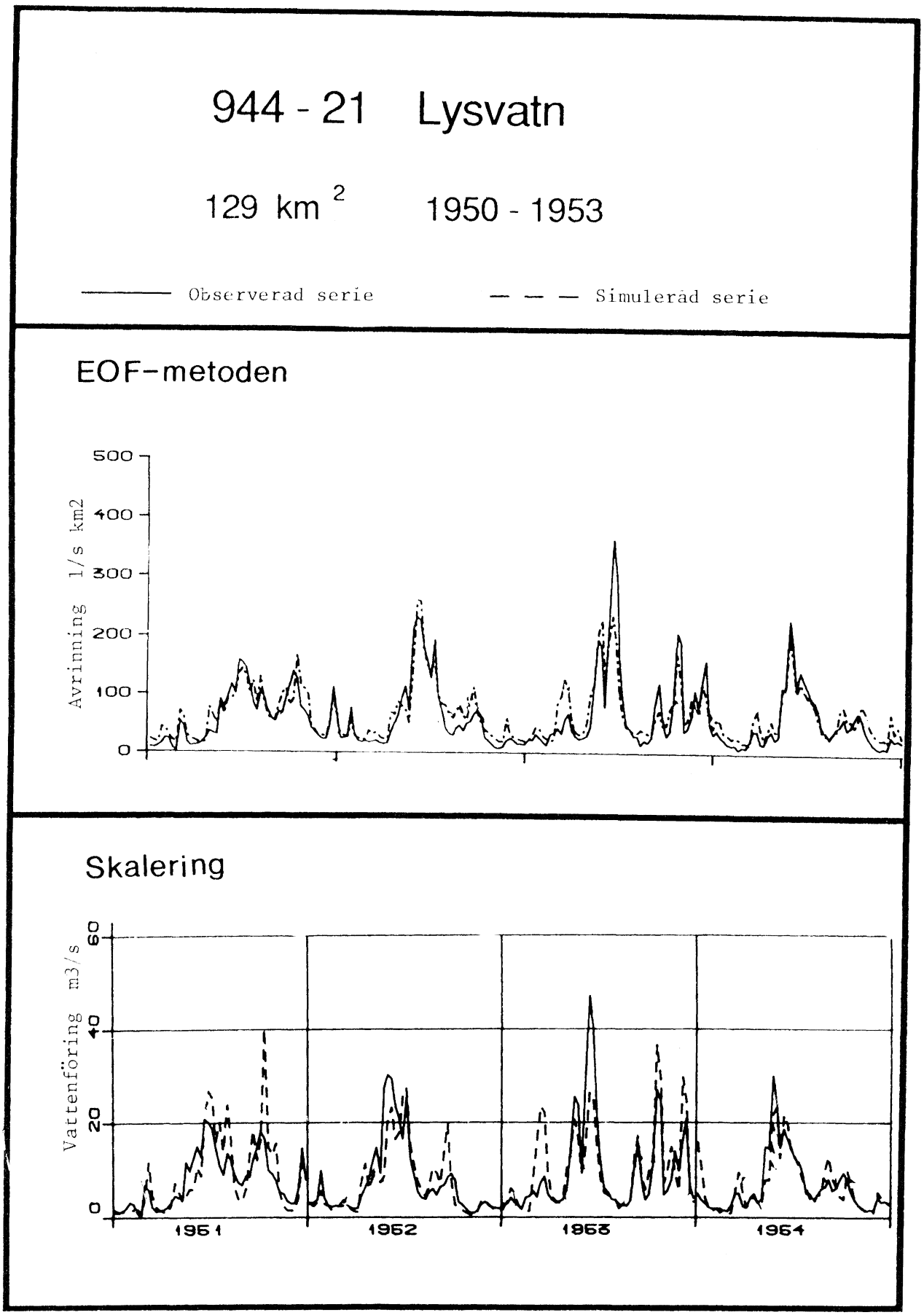
Oavsett om metoden ska användas för extrapolation och/eller interpolation, måste en EOF-analys genomföras för alla de av landets stationer som är aktuella. Analysen kan ses på som en engångsföreteelse så när som på att man behöver genomföra uppdateringar på grund av tillkommande observationer.



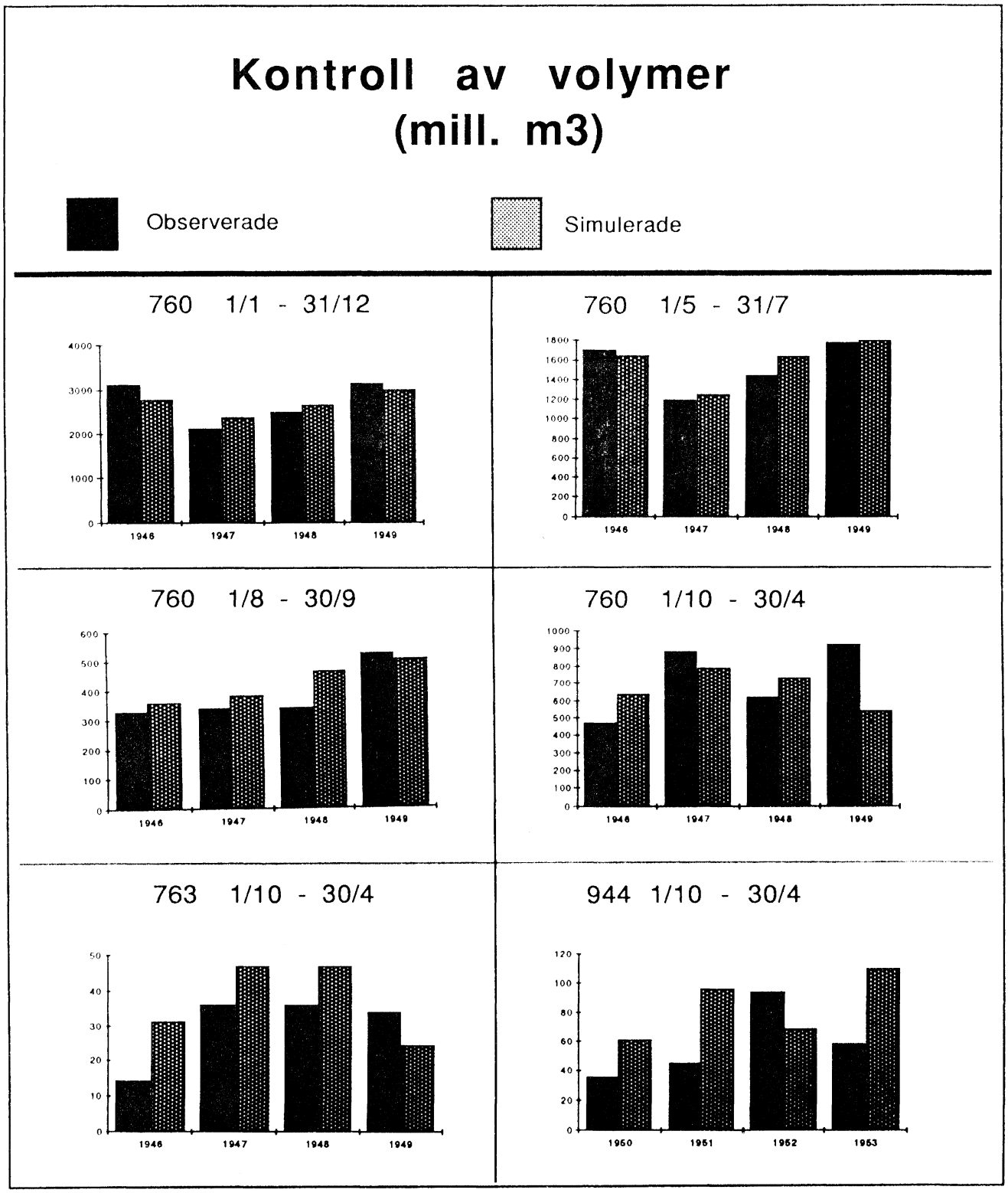
Figur 5a. Jämförelse av metoder vid interpolation. Station 760.



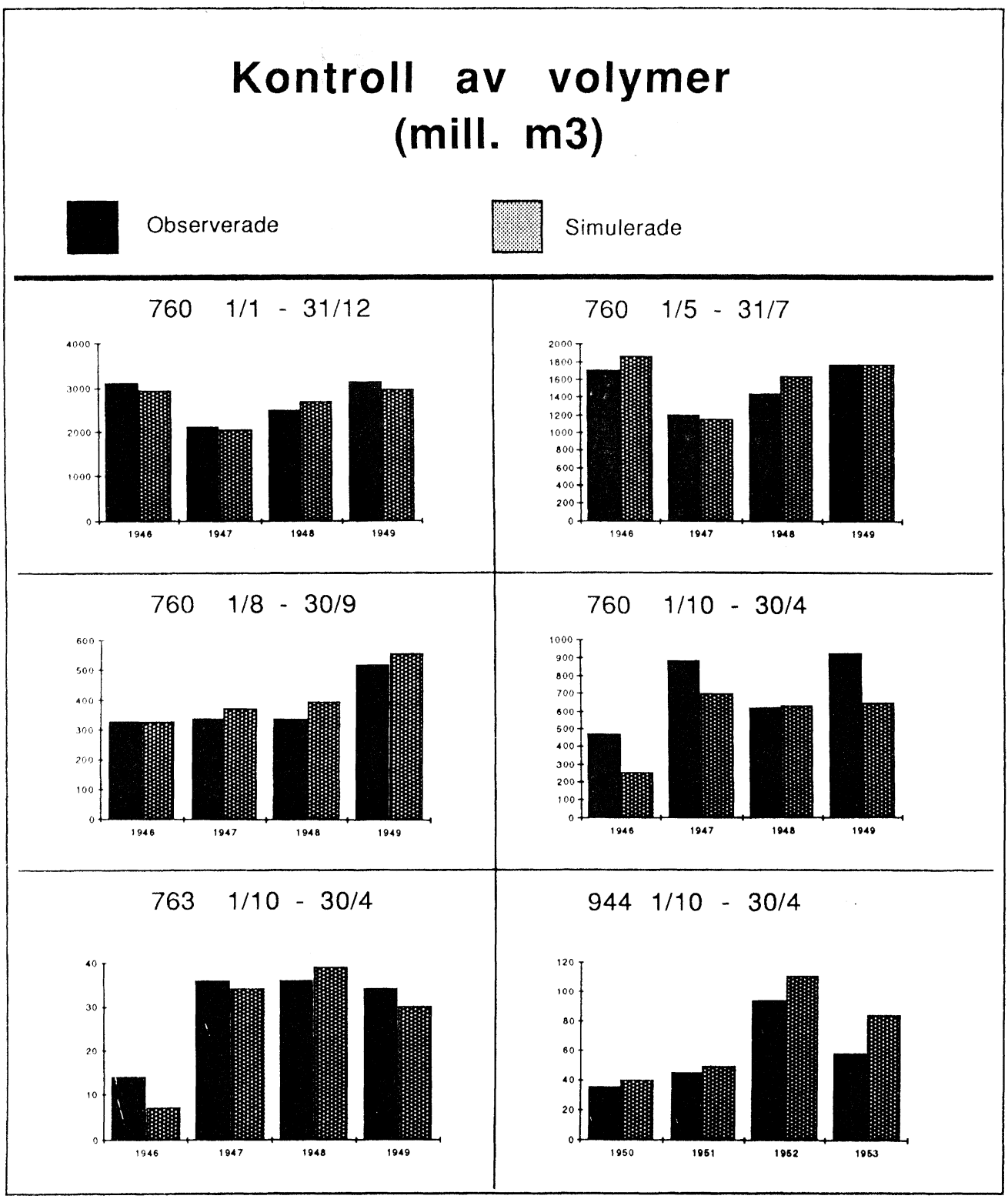
Figur 5b. Jämförelse av metoder vid interpolation. Station 763.



Figur 5c. Jämförelse av metoder vid interpolation. Station 944.



Figur 6a. Kontroll av volymer vid simulering via EOF.



Figur 6b. Kontroll av volymer vid simulering via skalering.

Det kan vara lämpligt att använda en indelning av landet i regimregioner som grundval för en gruppering av stationerna och analysera dessa grupper var för sig. Man får på så vis en effektivare beskrivning genom att undvika en alltför stark blandning av avrinningsmönster.
Det är naturligtvis möjligt att analysera landets alla stationer samtidigt, men det alternativet har ej analyserats här eftersom det ger relativt små fördelar rent praktiskt.

Det kan vidare vara nödvändigt att dela upp analysmaterialet i olika tidsperioder.

Man kan i EOF-analysen enbart inkludera de stationer som täcker hela den avsedda tidsperioden och antalet möjliga stationer kan då bli lågt vid analys av en lång tidsperiod.
En uppdelning i tidsperioder är därför lämplig för att få med så många stationer som möjligt. När alla analyser genomförts kan sedan de erhållna amplitudfunktionerna sammankopplas.
Även begränsningar i datorns kapacitet kan framtvinga en uppdelning av analysmaterialet.

EOF-analysen görs på värden med en, i förhållande till metoden, högsta rimliga tidsupplösning. Här rekommenderas veckovärden som högsta upplösning.

När EOF-analysen är genomförd för samtliga regioner och tidsperioder är i stort sett det mesta klart för extrapolation av serier.
Extrapolation av serier kräver då endast ett administrerande program som hämtar tillgänglig serie och motsvarande delar av amplitudfunktionerna, beräknar korrelationerna dem emellan och simulerar serien.

Interpolation av serier kräver en skattning av amplitudfunktionernas viktkoefficienter som här rekommenderas bli utförd via regression mot egenskaper hos avrinningsområdena.
Detta kan kräva information, som saknas eller endast till viss del finns i existerande arkiv. Det här gäller exempelvis uppgifter om effektiv sjöprocent som har dålig täckning i stationsregistret, men som i den här undersökningen ofta figurerade i regressionssambanden som oberoende variabel.

Idag har vi ej komplett tillgång till digital kartinformation men det dröjer säkerligen inte länge innan vi via sådan information relativt enkelt och snabbt kan få goda skattningar på parametrar som idag kräver stora insatser.
En annan möjlighet att skatta viktkoefficienterna vid interpolation är att använda exempelvis trendytor för rumsbeskrivningen.

7. REFERENSER

Bartlein, P.J.
1982 : Streamflow anomaly patterns in the USA and southern Canada 1951 - 1970. Journal of Hyd. Vol. 57, No 1/2

Holmström, I.
1969: Extrapolation of meteorological data. SMHI Notiser och preliminära rapporter, Serie Met. nr 22, Stockholm

Jutman, T.
1985: Extension of runoff series and interpolation using empirical orthogonal functions. Beiträge zur Hydrologie, Sonderheft 5,1 pp 387 - 398. Kirchzarten