

Variationer i publikations- og citationsanalyser for forskningsområder indenfor Det Naturvidenskabelige Fakultet

Bachelor opgave 7. semester
Danmarks Biblioteksskole
København

Afleveret onsdag d. 21. maj 2003

Vejleder: Ellen Bonnevie

Af Mette Colding og Nikolaj Winther Ingvorsen

Bacheloruddannelsen i Biblioteks- og Informationsvidenskab

Abstract

Opgaven undersøger variationen i publikation og citationsanalyser for forskningsområder indenfor Det Naturvidenskabelige Fakultet (DNF) på Københavns Universitet. Vores hypotese var at der ville være så stor en forskel mellem de forskellige forskningsområder at kvantitative bibliometriske undersøgelser såsom publikations- og citationsanalyser ikke kan stå alene når der skal foretages evaluering på tværs af forskningsområderne. Analyserne blev foretaget på professorerne tilknyttet institutterne ved DNF.

Citationsanalyserne blev udført i Science Citation Index (SCI), via Cited Reference Search i Web of Science, med supplerende analyser fra DialogClassic og Web of Science forfatterregister. Disse søgninger viste en systematisk forskel, afhængig af hvor godt det pågældende forskningsområde er repræsenteret i SCI. Dette resulterede i en opstilling af to modeller til forklaring af citationsanalysernes afhængighed af forskningsområdet og databasedækningen.

Publikationsanalyserne blev foretaget via DialogClassic. Publikationsanalysen gav kun et groft billede af forskellene forskningsområderne imellem. En analyse på de fem mest og fem mindst publicerende professorer viste, at publikationsraten ikke afhænger af hvilket institut de kom fra. En kombination af publikationsanalysen fra Dialog med en publikationssøgning fra Danbib for 5 udvalgte professorer, viste at de professorer som havde et lavt publikationstal i Dialog, opnåede en markant stigning i deres samlede publikationstal.

En samlet analyse af de to metoder påviste at undersøgelsesobjekternes forskningsområder skal være homogene før en sådan analyse vil give mening. Konkluderende fandt vi at citations- og publikationsanalyser som kvantitative analyser ikke kan bruges på tværs af forskningsområder.

Antal ord: 15.402

Kontributionsoversigt.....	2
1. Indledning	2
1.1 Hypotese.....	2
1.2 Problemformulering	3
1.3 Metode	3
1.3.1 Undersøgelsesobjekter - professorerne	3
1.3.2 Publikations- og citationsanalyse.....	4
1.4 Begrænsninger og forudsætninger	5
2 Bibliometriske metoder til forskningsevaluering – kritik og modkritik	6
2.1 Kan publikations- og citationsanalyser som kvantitative metoder overhovedet bruges til forskningsevaluering?	6
2.2 Problemer ved citationsanalyser	8
2.2.1 Er citationer et udtryk for kvalitet?	8
2.2.2 Selvcitationer.....	8
2.2.3 Forskel i publikations- og citationsmønstre disciplinerne imellem.	9
2.2.4 Flerforfatterskab.....	10
2.3 Problemer ved publikationsanalyser	11
2.3.1 Smalle forskningsområder	11
2.3.2 Publikationstype & Journal Impact Factor.....	11
3 Databaseværtbeskrivelse	13
3.1 De 8 scenarier i en citationsanalyse	14
3.2.1 DialogClassic	15
3.2.2 Web of Science (WoS).....	15
3.2.3 WoS General Search	15
3.2.4 WoS Cited Reference Search (CRS).....	16
3.3 Databasevalg	17
4 Citationsanalyse	21
4.1 Dataindhentning	21
4.1.1 Indsamling af data	21
4.2 Konvertering af søgeresultater	21
4.3 Diskussion af undersøgelsesobjekterne	22
4.4 Beskrivelse af de enkelte institutter	25
4.5 Analyse.....	27
4.6 Delkonklusion	30
5. Publikationsanalyse.....	32
5.1 Metode	32
5.2 Analyse.....	34
5.2.1 De samlede tal	35
5.2.2 Publikationstal kombineret med lokal database.....	36
5.3 Delkonklusion	41
6. Forhold mellem publikations- og citationstal	42
6.1 Sammenligning af de enkelte institutters forholdstal.....	42
6.2 Sammenligning af enkelte professorer.....	43
6.3 Delkonklusion	45
7. Konklusion	46
8. Kildefortegnelse	49
9. Bilagsoversigt.....	51

Kontributionsoversigt

Egentlig er hele opgaven skrevet i fællesskab, men af administrative årsager deles afsnittene op i følgende ansvarsfordeling.

Kapitlerne 1, 6, 7, 8 og 9 er skrevet i fællesskab.

Kapitlerne 2 og 5 er skrevet af Mette Colding

Kapitlerne 3 og 4 er skrevet af Nikolaj Winther Ingvorsen

1. Indledning

De bibliometriske analysemetoder, publikations- og citationsanalyser, bliver i stadig større grad benyttet til evaluering af forskning. Det kan dreje sig om evaluering og sammenligning af nationale universiteter, universiteter i forhold til udenlandske ditto, forskningsområder i forhold til tilsvarende områder på (udenlandske) forskningscentre eller sammenligning af fakulteter og institutter indenfor et universitet.

Forskningsevaluering på denne baggrund har på det seneste været debatteret (Winkel, 2002), og der har været rejst spørgsmål om disse undersøgelsers reelle anvendelighed til evaluering.

Vi finder debatten interessant, da der ifølge vores mening rent praktisk og etisk er problemer ved at basere forskningsevaluering på bibliometriske metoder. Der er mange begrænsninger og faldgruber ved udførelsen af selve analyserne. Og forskningsområderne imellem har forskellige traditioner indenfor forsknings- og publikationsmønstre, som kvantitative undersøgelser som de bibliometriske metoder ikke er egnede til at udtrykke. Disse kvalitative faktorer er nødvendige at tage under betragtning, når man udfører forskningsevaluering. Denne vinkel mangler ofte, når debatten omkring forsknings-evaluering kommer frem i offentligheden. Her benyttet de kvantitative publikations- og citationstal tilsyneladende som ”sandheden”, hvorfor debatten giver et skævvredet syn på de undersøgte universiteter, fakulteter, institutter og forskere, som jo *ikke* er homogene enheder.

1.1 Hypotese

Vores hypotese er, at der er så stor forskel på publikations- og citationsmønstre mellem de forskellige forskningsområder, at de kvantitative publikations- og citationsanalyser ikke kan stå alene, når der skal foretages forskningsevaluering. For at afprøve denne

formodning har vi valgt at lave en analyse på professorerne ansat ved Det Naturvidenskabelige Fakultet ved Københavns Universitet. Fakultetet har 15 institutter af forskellige størrelser, som dækker et bredt spektrum indenfor naturvidenskab. Denne diversitet indenfor fakultet gør det ideelt til vores undersøgelse. Vi har valgt at bruge institutternes professorer som vores undersøgelsesobjekter. Forudsætningen er, at professorer udgør forskningseliten, og har derfor et vist publikationsgrundlag bag sig. For at kunne lave en valid bibliometrisk undersøgelse er det nødvendigt at beskrive den metode man bruger minutiøst. I denne sammenhæng er det også vigtigt at beskrive de forskelle der er databaseværterne imellem. En bibliometrisk analyse kan eksempelvis få tre forskellige resultater, hvis man bruger tre forskellige databaseværter. Så for at kunne vurdere analysen, skal det være muligt at gennemskue de metoder der er brugt til at finde og behandle data.

I opgaven vil vi derfor fokusere meget på de metoder, vi bruger i forbindelse med udarbejdelse af publikations- og citationsanalyser. På den måde vil vi vise, hvor vigtige metoderne er for at få valide analyser. Og hvilke faldgruber og blindgyder man let kan komme ud for.

1.2 Problemformulering

I hvilken grad adskiller publikations- og citationsanalyser sig mellem institutterne og professorerne ved Det Naturvidenskabelige Fakultet?

Er det på baggrund af disse resultater overhovedet bruge de bibliometriske værktøjer (publikations- og citationsanalyser) til at evaluere forskning på tværs af forskningsområderne?

Hvor stor en indflydelse har værktøjet (databaser) og metoder til behandling af data på resultatet af de bibliometrisk analyser man foretager?

1.3 Metode

1.3.1 Undersøgelsesobjekter - professorerne

Vi har valgt at lave en analyse af professorerne, da flere undersøgelser (Budd, 2000; Hayes, 1983) viser, at professorer har en større publikationsrate end andre forskere. Hayes skriver bl.a. at

”...professors have presumably been represented in the data for a longer period of time than associate professors.”

Vi formoder derfor, at vi ved at undersøge professorerne fra de forskellige institutter kunne opnå et tilfredsstillende datagrundlag.

Vi startede med at samle en liste over København Universitets professorer fra Hof og Statskalenderen. Listen er på 293 personer så en udvælgelse var nødvendig for at vi skulle kunne holde os inden for rammerne af en bacheloropgave. Et datagrundlag på ca. 70 personer, fandt vi at være rimeligt datagrundlag for en analyse af denne størrelse. Valget stod mellem en fuldstændig analyse af et enkelt fakultet eller en stikprøveanalyse af et enkelt udvalgt institut fra hvert fakultet.

Vi valgte at foretage en analyse af et enkelt fakultet da dette efter vores mening ville give en mere ensartet datamængde. Vi forestillede os at forskellene ikke ville være ligeså store mellem institutter på et fakultet, som mellem institutter fra flere fakulteter. Et andet problem med at lave en undersøgelse på tværs af fakulteterne er at forskellene internt på fakulteterne og inden for institutterne trods alt også er meget store, hvorfor en udvælgelse af et enkelt institut som repræsentant for et helt fakultet eller en stikprøve af professorer inden for institutterne kan give et skævt billede. En formodning der blev begrundet da vi kom i gang med vores analyser af Det Naturvidenskabelige Fakultet. Vi har derfor valgt at se på alle institutterne og alle professorerne indenfor disse. Det vil sige vi laver en 100% undersøgelse af hver enkelt professors publikationer og modtagne citationer gennem hele dennes karriere så vidt dette er muligt.

På Det Naturvidenskabelige Fakultet er der ifølge Hof og Statskalenderen 82 professorer ansat, hvilket vi betragter som et rimeligt antal til at danne grundlag for analyserne. Vi har i vores undersøgelser undladt de professorer der ikke er omtalt på de respektive institutters hjemmeside, samt de professorer der er omtalt på hjemmesiderne men ikke er omtalt i Hof og Statskalenderen. Dette gjorde at vi endte på samlet 68 personer der blev lavet citationsanalyse på. Af disse var der to der ikke kunne lave publikationsanalyse på, fordi deres navne var for almindeligt til at kunne lave en entydig analyse på.

1.3.2 Publikations- og citationsanalyse

Metodebeskrivelserne til henholdsvis publikations- og citationsanalyserne er beskrevet i detaljer under deres respektive afsnit (henholdsvis afsnit 5 og afsnit 4). Endvidere har vi beskrevet metoden til konvertering af data i afsnit 4.2.

1.4 Begrænsninger og forudsætninger

Dataindsamling:

Vi har valgt at foretage vores publikationsanalyse i databaseværten Dialog. En publikationsanalyse burde suppleres med søgninger i lokale databaser og opslag i universitetets årbøger, for at få den så fuldstændig som muligt. På grund af tidsnød var det ikke muligt at foretage yderligere publikationsanalyse i lokale databaser for samtlige professorer, men kun for et lille udvalg. Vi afgrænsede her søgningen til Danbib, selvom yderligere supplement via Københavns Universitets Elektroniske Publikationsregister (PUF) + opslag i årbøger ville have givet et mere validt resultat.

Danbib:

En national database, som også medtager avisartikler. Dette mener vi hører med til en forskers publikationstal, da dette viser et engagement i den samfundsmæssige debat. Problemet med Danbib er, at vi kan risikere et vist antal dubletter, som det ikke maskinelt er muligt at frasortere. Samtidig indeholder Danbib kun et begrænset antal udenlandske publikationer.

Forudsætning:

I opgaven omtaler vi i forskningsområderne som henholdsvis ”stærke” og ”svage”. Vi har defineret områder således:

De stærke forskningsområder er karakteriseret ved at have et stor antal fundne poster i de internationale database.

De svage forskningsområder er karakteriseret ved et lavt antal fundne poster i de internationale databaser.

2 Bibliometriske metoder til forskningsevaluering – kritik og modkritik

I dette afsnit vil vi komme ind på de argumenter, der er kommet frem vedrørende brugen af bibliometriske metoder i forbindelse med forskningsevaluering.

I den videnskabelige normative struktur bliver publikationer betragtet som en forskers bidrag til ”kundskabs-tilvæksten”, og bliver anerkendt som forskerens intellektuelle ejendom. Dette vil sige, at forskningsresultatet bliver tilskrevet den forsker, som publicerer det. En citation givet til en publikation er en henvisning til, at dette værk har haft indflydelse i forbindelse med den undersøgelse, som ligger til grunde for den nye publikation. I princippet burde en citation til en publikation være en reel indikator for denne publikations indflydelse. Men der er stor uenighed om, hvad der ligger til grund for citationer og citationsmønstre. Debatten omkring dette har floreret gennem mange år, og en egentlig teori om citationsmønstre foreligger ikke.

Denne debat har også indflydelse på diskussionen om, hvorvidt citationsanalyser kan bruges som et kvantitativt redskab i forbindelse med forskningsevaluering.

2.1 Kan publikations- og citationsanalyser som kvantitative metoder overhovedet bruges til forskningsevaluering?

Vi har valgt at fokusere på de debatpunkter, som vi selv har haft under overvejelse i forbindelse med udarbejdelsen af denne opgave.

Birger Hjørland ridser i artiklen ”Forskningsevaluering i videnskabsteoretisk belysning” (Hjørland, 1996) 3 hovedsynspunkter op i synet på forskningsevaluering. Vi vil her give en meget simplificeret beskrivelse af disse:

- Forskning kan måles som enhver anden produktion (den bureaukratiske/positivistiske)

Det positivistiske syn betragter forskning som noget rent objektivt, som direkte kvantitativt kan tælles og måles i eksempelvis antallet af publikationer/citationer. Man forudsætter at de indikatorer, som bliver brugt, er reelle, og ikke udgør fejlkilder. Her kan nævnes antallet af publikationer, modtagne citeringer og tidsskrifter er peer-reviewed.

- Forskning kan ikke måles, men skal evalueres på grundlag af gennemgribende læsning af publikationer (pragmatisk).

Det pragmatiske syn har den holdning, at forskning ikke kan vurderes ud fra antallet af publikationer, men ud fra de reelle forskningsresultater publikationerne indeholder. Derfor skal forskning evalueres ud fra grundig gennemgang af eksperter inden for forskningsfeltet. Forskning skal ikke styres af kvantitative krav (publish or perish), da dette kan medføre et overload af publikationer, som ikke kvalitativt lever op til forskningskrav. Forskere, som udfører peer review, gør dette arbejde uden hensyntagen til subjektive synspunkter.

- Forskning skal vurderes i dialog med flere parter eksempelvis forskere, videnskabsforskere og politikere/administratorer (videnskabsteoretiske/sociologiske) Dette synspunkt anbefaler at en ren kvantitativ undersøgelse eller en kvalitativ vurdering udelukkende baseret på peers bør ikke stå alene ved evaluering af forskningsresultater. De kvantitative undersøgelser kan indeholde masser af fejlkilder. Ikke alle de publikationer, som en forsker udgiver, er repræsenteret i citationsdatabaserne. Og citationstallene er også behæftet med fejlmuligheder (som vi kommer nærmere ind på senere i afsnittet). En ren kvalitativ undersøgelse vil være påvirket af den eller de forskere, som foretager den. Faktorer som konkurrencehensyn, faglige fordomme, venskab etc. må formodes at kunne påvirke en vurdering af forskningen. Endvidere er denne type undersøgelser meget bekostelige både tidsmæssigt og økonomisk. Derfor bør forskningsevaluering foregå i samarbejde med både specialister inden for forskningsområdet og specialister inden for bibliometriske metoder, som kan dechifrere faldgruberne i publikations- og citationsanalyser (Kärki & Kortelainen, 1998).

Eugene Garfield gør selv opmærksomt på, at bibliometriske analyser ikke alene kan bruge som evaluering, men kan vise nytteværdien af publikationer :

"..The only responsible claim made for citation counts as an aid in evaluating individuals is that they provide a measure of the utility or impact of scientific work. They say nothing about the nature of the work, nothing about the reason for its utility or impact. Those factors can be dealt with only by content analysis of the cited material and the exercise of knowledgeable peer judgment. Citation analysis is not meant to replace such judgement, but to make it more objective and astute.." s.246 (Garfield, 1979)

2.2 Problemer ved citationsanalyser

2.2.1 Er citationer et udtryk for kvalitet?

Kritikere af citationsanalyser (MacRoberts & MacRoberts, 1996; Seglen 1996) argumenterer for, at citationer alene ikke er et udtryk for kvalitet, da man ikke umiddelbart ved, hvorfor en citation er blevet givet. I teorien kan en publikation, hvis forskningsresultat bliver mødt med stærk kritik, blive citeret kraftigt, da andre forskere i deres modkritik har denne publikation som reference. I en kvantitativ forskningsevaluering vil forskeren dog fremstå som succesfuld pga. et højt citationstal. Men undersøgelser viser, at forskning som udmærker sig ved at være af dårlig kvalitet oftest bliver ignoreret af andre forskere, og derfor sjældent bliver citeret. Hvis et forskningsresultat som vedrører et forskningsmæssig kerneområde og bliver publiceret i et anerkendt tidsskrift, ligger til grund for kritik vil negative citationer være et tegn på debat, og ikke nødvendigvis et dårligt tegn. (Garfield, 1979)

En anden og måske vigtigere facet af kvalitetsspørgsmålet er, at mange af de højstciterede artikler ikke er kerneartikler, men oversigts- eller metodeartikler (Hjørland, 1996; MacRoberts & MacRoberts, 1996). Det er dog et vurderingsspørgsmål om disse skal betragtes som værende af mindre kvalitet af den grund. Set fra et forskningssynspunkt er publikationer omkring nye forskningsresultater af større kvalitet. Oversigts- og metodeartikler bliver ofte brugt i situationer, hvor der ikke er mulighed for at sætte sig dybere ind i området, eller hvor man skal have et hurtigt overblik før videre tilgang til området. Dette er en vigtig funktion, som bestemt ikke skal negligeres. Men problemet er, at citeringer givet til en oversigtsartikel ofte vil ”tilhøre” en helt anden publikation, som bliver omtalt her. Dette er en skævvridning af citationstallene.

2.2.2 Selvcitationer

Er det validt at tælle selvcitationer med i det samlede citationstal? Eller giver det et bedre billede af forskeren og vedkommendes arbejde, hvis selvcitationerne sorteres fra? Debatten er her gået på, at nogle forskere citerer sig selv i overdrevent grad – dels for at ”lave reklame” for sig selv og dels for at hæve sit citationstal (bl.a. Seglen, 1992; Budd, 2000). Modargumentet er, at det er naturligt for forskere at bygge videre på egen forskning, og derfor citerer sit tidligere arbejde. (Garfield 1979; Budd, 2000). I teorien kan selvcitation være en metode for en forsker at upgrade sit citationstal. Men hvis dette er tilfældet, vil det være temmelig åbenlyst. For at opnå dette, skal forskeren have en

meget stort antal publikationer i et stort antal tidsskrifter, som i sidste ende vil være mere og mere perifere tidsskrifter. Så et abnormt højt selvcitationstal sammen med en lang publikationsliste i perifere tidsskrifter vil indikere ”snyd” med citationstal.

Derimod betyder et stort antal selvcitationer oftest, at forskeren er specialist indenfor et smalt forskningsområde. (Garfield, 1979)

Budd gør endvidere opmærksomt på, at man ikke umiddelbart kan være sikker på om en selvcitation er valid eller ej (Budd, 200). Det er nødvendigt at læse publikationerne nøje igennem, for at registrere i hvilken sammenhæng citationen er givet.

En undersøgelse af Snyder & Bonzi fra 1998 viser, at i gennemsnit 9% af alle citationer var selv-citationer (Snyder & Bonzi, 1998). Der er variationer mellem forskellige discipliner, men indenfor disciplinerne er selv-citations procenten stabil.

Debatten omkring selvcitationer foregår stadig, og basalt set er det op til en selv at vurdere, hvilken holdning man har til spørgsmålet. Vi har besluttet at bibeholde selvcitationerne i citationsanalysen, da vi er enige i argumentet om, at forskere bygger videre på egen forskning. Og derfor naturligt citerer til tidligere forskningsresultater.

2.2.3 Forskel i publikations- og citationsmønstre disciplinerne imellem.

Ifølge Kärki & Kortelainen kan man geografisk lave bibliometriske undersøgelser på 3 niveauer

- på internationalt niveau, hvor enkelte landes forskning inden for definerede områder kan sammenlignes.
- På nationalt niveau, hvor forskellige forskergrupper i et land undersøges.
- På lokalt niveau, hvor forskellige fakulteter ved et universitet undersøges på baggrund af de producerede publikationer og de citeringer disse får. (Kärki & Kortelainen, 1998)

Problemet her er, hvordan de resultater man kommer frem til, bliver fortolket. For hvis resultatet ikke bliver sammenholdt med en kvalitativ analyse eller i det mindste tager hensyn til de forskelle der er mellem forskningsdiscipliner, kan resultatet kun bruges til at påvise trends indenfor forskning i det tidsrum, man har valgt at foretage undersøgelsen.

Publikationsmønstrene mellem de forskellige discipliner er meget forskellige. Det er ikke blevet undersøgt tilbundsgående, men et eksempel er at der indenfor biokemi i gennemsnit er 30 referencer pr. publikation og i matematik kun et gennemsnit på 15 referencer. Det vil sige, at det i teorien er lettere at blive citeret indenfor biokemi end

indenfor matematik. (Garfield, 1979; Peter & Van Raan, 1994). Eller sagt med andre ord: Et bredt forskningsfelt vil tiltrække flere citationer end et forholdsvis snævert. Rosseau henviser til de forskelle der kan være i den forskningsmæssige tradition (Rosseau, 1998). Indenfor matematik (hævder Rosseau) vil man enten citerer af etiske årsager, hvor man citerer den, som første gang beviste en teori eller ”gjorde en opdagelse”, eller af bekvemmelighed, hvor citeringen er en genvej til beskrivelse af en metode, tidligere forsøg og løsningsmodeller etc. Men jo mere diffus og bred en disciplin er, jo flere referencer får forskeren brug for, for at underbygge sin egen forskning.

2.2.4 Flerforfatterskab

Indenfor disciplinerne kan der være forskel på, hvor mange forfattere/forskere der er med til at publicere artikler. Dette kan have betydning med hensyn til antallet af citationer til flerforfatter publikationer. Spørgsmålet er, om citationstallet for den enkelte forsker skal stå i forhold til det samlede antal forfattere på en publikation. Eksempel: en publikation har 4 forfattere og får 100 citationer. Dette svarer til 25 citationer pr. forfatter, som er det antal citationer, som tælles med i forskerens samlede citationstal. Dette er diskussionen om fraktioneret citationstal, som vi har valgt ikke at tage under overvejelse i analyserne.

En anden vinkel på dette spørgsmål er antallet af selvcitationer. Som nævnt er det naturligt for en forsker at basere sin forskning på tidligere forskning. Og vil derfor citere sine tidligere publikation, som omhandler tidligere forskningsresultater. Hvis man er af den holdning, at selvcitationer ikke skal indgå i det samlede citationstal, skal citationer, som kommer fra en forskers medforfattere (som jo i denne situation citerer sig selv), tælles med som selvcitation?

Som ved spørgsmålet omkring selvcitationer er disse spørgsmål stadig under debat. Hvis man i en bibliometrisk analyse undersøger enkelte personers gennemslagskraft, vil det mest korrekte være at bruge det fraktionerede citationstal. Men dette er næsten umuligt at håndtere, når den bibliometriske analyse skal omhandle et stort antal individer som eksempelvis ved undersøgelser af fakulteter/universiteter. Derfor vil der uvilkårligt være skævheder mellem disciplinerne ved bibliometriske undersøgelser.

2.3 Problemer ved publikationsanalyser

I dette afsnit kommer vi ind på nogle af de facetter i en publikationsanalyse, som kan have betydning, når man undersøger forskellige discipliners publikationsmønstre. Igen har vi valgt at fokusere på de facetter, vi selv har taget til overvejelse i forbindelse med opgaven.

2.3.1 Smalle forskningsområder

Indenfor de forskellige discipliner er der forskel på, i hvilke typer tidsskrifter det er mest relevant for forskerne at få udgivet deres forskningsresultater. Problemet er her, at der er stor risiko for, at de mere smalle forskningsområders tidsskrifter ikke bliver repræsenteret i de internationale databaser. De tidsskrifter, som disse forskningsresultater vil blive publiceret i, er simpelthen så små og nicheprægede, at de ikke er optaget i de større databaser.

Indenfor en snævert sprogområde som dansk vil der opstå tilsvarende problemer. De internationale databaser optager i sagens natur primært tidsskrifter på hovedsprogene (engelsk, tysk, fransk, russisk). Danske og nordiske tidsskrifter bliver medtaget i mindre grad, hvilket har indflydelse på publikationsanalysen.

For at få en så fuldstændig publikationsanalyse som muligt, er det nødvendigt at undersøge mere lokale databaser end dem, de store databaseværter (eksempelvis Dialog og STN) stiller til rådighed. Eller gå direkte til ”hestens mule” efter disse oplysninger. Universiteterne eller fakulteterne kan have trykte lister over de udgivelser institutionen har haft. Og nogle forskere har disse oplysninger tilgængelig på deres personlige hjemmesider. Under alle omstændigheder medfører ambitionen om fuldstændige publikationslister et større og mere besværligt arbejde, end hvis man stiller sig tilfreds med søgninger hos de store databaseværter.

2.3.2 Publikationstype & Journal Impact Factor

Kravet eller forventningen om publikationshastighed – hvor hurtigt og hvor mange publikationer en forsker forventes at levere – kan være afhængigt af forskningsområde. I områder, hvor forskning går relativt hurtigt, kan det være nødvendigt at være hurtigt ude med sine forskningsresultater. Forskningen vil blive udgivet fragmenteret for at sikre forskeren den intellektuelle ejendom til delresultater, analysemetode etc. Det kan også være nødvendigt i forskningsområder med stor konkurrence blandt forskere (konkurrence om anerkendelse, jobmuligheder, bevillinger o.s.v.) at publicere ofte (”publish or perish”). I den anden boldgade har vi de ”tunge” forskningsområder, hvor

der kan gå mange år før der foreligger forskningsresultater, og hvor publikationsraten derfor bliver lavere. Disse faktorer giver sig udslag i, hvilke typer af publikationsformer, man vælger til sine publikationer. Her menes, hvilken type publikationsform (artikler, review papers, letters, editorials, monografier, etc.) og hvilken type tidsskrift (Amin & Mabe, 2000).

Dette har betydning, hvis man laver en citations- og publikationsanalyse, som benytter sig af Journal Citation Reports (JCR) og Journal Impact Factor (JIF). Her ser man på de citationer et tidsskrifts artikler modtager over en defineret tidsperiode i forhold til det samlede antal tidsskrifter publiceret af tidsskriftet i samme periode. I forbindelse med dette kan man udregne *immediacy index*, som er et mål for, hvor hurtigt publikationerne i et tidsskrift bliver citeret efter udgivelsen. Og *cited half-life*, som er et mål for hvor længe artikler i et tidsskrift bliver citeret efter udgivelsen. Eksempelvis kan en forsker vælge at publicerer i et tidsskrift, som har et højt immediacy index men et lavt cited half-life. Dette betyder at publikationen hurtigt vil modtage citationer, men dette citationstal vil til gengæld også nå toppunktet hurtigt. Dette kan indikerer om publikationen er langtids holdbar (at forskningsresultatet er betydeligt), eller ej. JIF bliver i visse tilfælde brugt til at vurdere forskeres publikationer ud fra devisen, at et tidsskrift med et højt JIF derfor må være et høj kvalitetstidsskrift. Og det derfor er et kvalitetsstempel af forskeren at få publiceret i disse tidsskrifter. Dette er dog uholdbart, da mange af de mindre domænespecifikke tidsskrifter har et lavt JIF til trods for, at de betragtes som de bedst mulige indenfor deres område. Så hvis JIF skal bruges som indikator, skal domænerne tidsskrifter vurderes særskilt, og ikke sammenlignes op mod hinanden.

Generelt for citations- og publikationsanalyser kan siges, at de alene kan være svære at vurdere. For at kunne analysere på undersøgelserne, skal de stå i relation til andre tal, så man kan vurdere hvordan tallene forholder sig til hinanden. Og i brugen af bibliometriske analyser skal man tage hensyn til, at det er rene kvantitative resultater. Som et kvalitetsmåling kan disse metoder ikke stå alene.

3 Databaseværtbeskrivelse

Vi vil i vores citationsundersøgelser bruge databasen SCI gennem tre databaseværter, nemlig DialogClassic, WoS's General Search og WoS's CRS. Dette afsnit behandler de problemer der er i forbindelse med bibliometriske undersøgelser med netop disse værktøjer. De tre metoder vi her vil gennemgå viser at der er mange måder at lave citationsanalyser på, samt at man skal være opmærksomme på hvilke værter man bruger. Alle tre søgninger er foretaget i den samme database men alligevel er der endog meget stor forskel på resultaterne. Det hænger sammen med at selvom det er den tilsyneladende samme database er der alligevel forskelle.

I SCI's bluesheet til DialogClassic skriver de at de indekserer 4500 større (major) videnskabelige tidsskrifter fra 1974 og frem (ISI, 2001a).

I WoS indekseres der 5700 større (major) videnskabelige tidsskrifter fra 1945 og frem (ISI, 2001b).

Desuden ændres listen over tidsskrifter som SCI indekserer hele tiden. En større ændring af basens præmisser skete i 1996 da SCI gik fra at indeksere ca. 3500 tidsskrifters referencer til at indeksere ca. 5300 tidsskrifters referencer. Dette vil nødvendigvis give forhøjede citationstal og dermed et forkert billede af udviklingen (Find, 1999). Dette er ikke direkte relevant for vores undersøgelse men bør alligevel medtages for at nuancere billedet. Et andet aspekt af citationsundersøgelser der her skal omtales er den skævhed der er indbygget i SCI databasen. Visse forskningsområder er marginaliserede i forhold til SCI's indekseringspolitik, for eksempel områder hvis forskning hovedsagligt henvender sig til et "lokalt" publikum, og derfor publicerer på andre sprog end engelsk, der med omkring 90% af posterne er det fremtrædende sprog i den database.

Sprogfordelingen i SCI			
rank	language	Hits	Procent
1	ENGLISH	19102542	91%
2	GERMAN	583631	3%
3	RUSSIAN	553041	3%
4	FRENCH	440046	2%
5	JAPANESE	95256	0%

Tabel 1. De fem mest repræsenterede sprog i SCI databasen.

3.1 De 8 scenarier i en citationsanalyse

Når en forsker har publiceret artikel og den skal genfindes i forbindelse med en citationsanalyse findes der 8 forskellige scenarier den kan befinde sig i. I det følgende vil vi beskrive disse samt illustrere forholdet mellem de forskellige scenarier for forskere der publicerer indenfor et forskningsområde der er henholdsvis stærkt og svagt repræsenteret i SCI. Dette gøres ved de to modeller der er senere i dette afsnit.

1. Forskeren er **førsteforfatter** på artiklen, som er udgivet på et tidsskrift **der er indekseret af SCI** (ex. Nature), samt at citationerne på artiklen er fra tidsskrifter **der er indekseret af SCI** (ex. Nature).
2. Forskeren er **andenforfatter** på artiklen, som er udgivet på et tidsskrift **der er indekseret af SCI** (ex. Nature), samt at citationerne på artiklen er fra tidsskrifter **der er indekseret af SCI** (ex. Nature).
3. Forskeren er **førsteforfatter** på artiklen, som er udgivet på et tidsskrift **der ikke er indekseret af SCI** (ex. Ugeskrift for Læger), samt at citationerne på artiklen er fra tidsskrifter **der er indekseret af SCI** (ex. Nature).
4. Forskeren er **andenforfatter** på artiklen, som er udgivet på et tidsskrift **der ikke er indekseret af SCI** (ex. Ugeskrift for Læger), samt at citationerne på artiklen er fra tidsskrifter **der er indekseret af SCI** (ex. Nature).
5. Forskeren er **førsteforfatter** på artiklen, som er udgivet på et tidsskrift **der er indekseret af SCI** (ex. Nature), samt at citationerne på artiklen er fra tidsskrifter **der ikke er indekseret af SCI** (ex. Ugeskrift for Læger).
6. Forskeren er **andenforfatter** på artiklen, som er udgivet på et tidsskrift **der er indekseret af SCI** (ex. Nature), samt at citationerne på artiklen er fra tidsskrifter **der ikke er indekseret af SCI** (ex. Ugeskrift for Læger).
7. Forskeren er **førsteforfatter** på artiklen, som er udgivet på et tidsskrift **der ikke er indekseret af SCI** (ex. Ugeskrift for Læger), samt at citationerne på artiklen er fra tidsskrifter **der ikke er indekseret af SCI** (ex. Ugeskrift for Læger).
8. Forskeren er **andenforfatter** på artiklen, som er udgivet på et tidsskrift **der ikke er indekseret af SCI** (ex. Ugeskrift for Læger), samt at citationerne på artiklen er fra tidsskrifter **der ikke er indekseret af SCI** (ex. Ugeskrift for Læger).

De ovennævnte punkters numre svarer til områdenumrene i henholdsvis Model I og II.

En publikationsanalyse i SCI vil opfange publikationer fra grupperne 1, 2, 5 og 6. Det er fordi en publikationsundersøgelse benytter sig at bibliografiske data fra registre som for eksempel forfatterregistret.

En citationsanalyse foretaget i SCI vil aldrig opfange citationer fra grupperne 4-8, på grund af den måde databaserne og deres værter arbejder på. Vi vil her gennemgå SCI databasens tre værter og hvor stor en andel af citationsmængden de ”opfanger”.

3.2.1 DialogClassic

CR=ROSING MT, 1999, V283, P674, SCIENCE

Eksempel på post fra CR-feltet i SCI

En citationsanalyse i databaseværtten DialogClassic foretages i CR feltet.

Databaseværtten DialogClassic søger i CR feltet på de poster SCI har indekseret. Det vil sige at der kun opfanges citationer fra scenarie 1 og 3, da databasen i CR-feltet er fraseindekseret.

En post i CR feltet starter med forfatternavnet, derefter årstallet for publikationen, volumenummer, sidetal, og til sidst hvilket tidsskrift der er publiceret i.

DialogClassic giver mulighed for at søge i hele strengen, en såkaldt CR søgning; i CA-feltet, hvor man søger på forfatternavn; CY-feltet, hvor man søger på hvilket år publikationen er fra; samt CW-feltet, hvor man søger på hvor den citerede publikationen er publiceret.

For at finde ud af hvor mange gange en forfatter er citeret bruger man derfor CA-feltet, men da det kun er førsteforfatteren der er indekseret i CR feltet vil man kun finde de publikationer som den undersøgte forfatter har været førsteforfatter på. Det er heller ikke nok at kende de andre bibliografiske data, da søgningen går ind i en posts CR-felt og søger fra starten af subfieldet. Passer navnet ikke, er søgemaskinen ligeglad med om resten af posten passer.

3.2.2 Web of Science (WoS)

Web of Science (WoS) giver mulighed for at lave citationssøgning på to forskellige måder, nemlig Cited Reference Search (CRS) og det de selv kalder for ”General Search” altså en søgning i de konventionelle delfelter af en bibliografisk post. Det kan for eksempel være forfatterfeltet, Corporate Source feltet, samt hvad man ellers måtte finde interessant.

3.2.3 WoS General Search

General Search giver en mulighed for at søge i følgende index i SCI:

TS=Topic
TI=Title
AU=Author
SO=Source
AD=Address

Søgningerne resulterer i poster uden citationstal, så det er egentlig en base til publikationsanalyser, mere end til citationsanalyser.

Hvis man vil lave citationsanalyser afhjælper ”General Search” søgning i forfatterregistret problematikken om hvorvidt forfatteren er første eller andenforfatter, da databaseværten selv går ind og konverterer postens data til brugbare søgestrengene til citationssøgninger. Dermed giver denne søgning mulighed for at opfange citationer fra scenarie 1 og 2. Derudover vil den også opfange publikationer fra scenarie 5 og 6, så man derfor får mulighed for at udregne citationsfrekvensen. Denne metode giver ikke mulighed for at søge i CR-feltet hvilket i visse tilfælde kan være ganske hæmmende for en ordentlig indsamling af data. Dette er især tilfældet i forskningsområder der er marginaliserede i SCI. Dette skyldes at søgning i CR-feltet ikke er hæmmet af de begrænsninger som SCI’s indekseringspolitik indeholder. I CR-feltet kan alle blive citeret, selv folk som aldrig har publiceret noget, blot nogen har fundet hvad de har yttet sig om interessant. Derfor vil citationssøgninger i CR-feltet også give et større resultat.

3.2.4 WoS Cited Reference Search (CRS)

CRS kombinerer det bedste fra de to ovenstående metoder således at man både får den omfattende søgning fra CR-feltet med, samtidig med at man også inkorporerer de citationer som andenforfattere ikke bliver tildelt i DialogClassic. Dette er muligt da denne funktion både søger i CR-feltet samt forfatterregisteret. På den måde opfanger den scenarierne 1, 2 og 3, men ikke fra nogle af de andre ovenstående beskrevne scenarier. Dens fremvisning af søgeresultater gør derimod at den kun medtager poster der har modtaget en eller flere citationer, hvorfor den sorterer poster væk der ingen citationer har modtaget, hvilket gør det umuligt at beregne citationsfrekvensen.

3.3 Databasevalg

Vi har udvalgt en professor per institut, nogenlunde svarende til 10 procent af det samlede antal, til en testsøgning med de tre metoder vi har beskrevet ovenover.

Institut	Navn	1	2	3	4	5	6
August Krogh	Leth-Jørgensen P	7441	6543	5231	100	88	70
Botanisk Institut	Moestrup O	1946	1585	1354	100	81	70
Datalogisk Institut	Skelboe S	252	134	228	100	53	90
Geografisk Institut	Anderson NJ	1110	939	706	100	85	64
Geologisk Institut	Surlyk FC	1317	608	1032	100	46	78
Geologisk Museum	Rosing MT	744	620	293	100	83	39
Idræt	Richter EA	5404	5284	1844	100	98	34
Kemisk Institut	Mikkelsen KV	1960	1847	1159	100	94	59
Institut for Matematiske Fag	Pedersen GK	2406	1219	1712	100	51	71
Molekylærbiologisk Inst.	Egel R	2132	1943	1321	100	91	62
Niels Bohr Instituttet	Bak P	13588	12183	10840	100	90	80
Niels Bohr Instituttet	Als-Nielsen j	8228	7188	2359	100	87	29
Zoologisk Institut	Fenchel T	10662	7389	7995	100	69	75
Zoologisk Museum	Kristensen NP	852	298	831	100	35	98

Tabel 2

Tabelforklaring:

Kolonnerne 1-3 er reelle tal, mens 4-6 er indekstal sat med kolonne 1 som 100.

Kolonne 1: Antal citationer fundet med søgning med CRS i WoS

Kolonne 2: Antal citationer fundet med søgning i forfatterregistret i WoS

Kolonne 3: Antal citationer fundet med søgning i SCI, SSCI og A&HI i Dialog

Kolonne 4: Kolonne 1 sat som 100

Kolonne 5: Andelen af citationer fundet med søgning i forfatterregistret i WoS i forhold til det maksimalt fundne antal

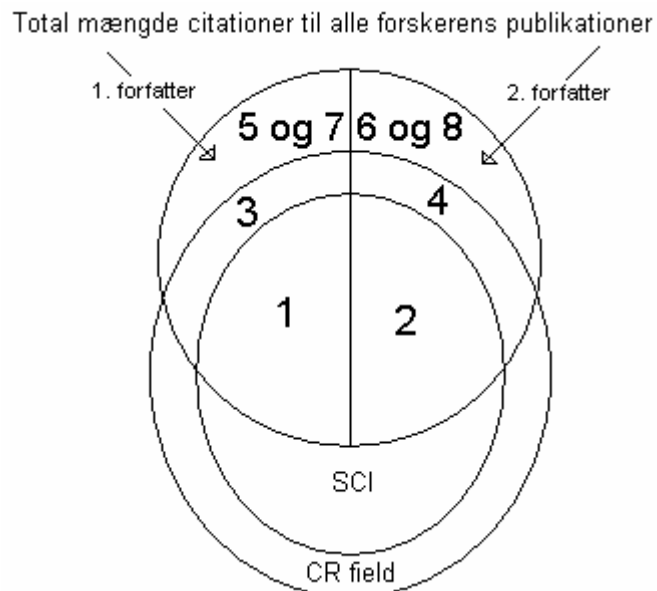
Kolonne 6: Andelen af citationer fundet med søgning i SCI, SSCI og A&HI i Dialog forhold til det maksimalt fundne antal

Note til kolonne 4. Udfra de scenarier vi har beskrevet ovenfor er vi godt klar over at det ”rigtige” antal citationer er praktisk umuligt at finde. Derfor har vi måtte nøjes med det næstbedste, nemlig at sætte den metode der giver det største recall som benchmark.

Som det ses varierer citationstallet voldsomt alt efter hvilken vært man bruger, men det afhænger også af hvem man søger på. Vores teori var at jo færre citationer man havde modtaget, jo mindre ville andelen af citationer fra WoS forfatterregister være i forhold til resultatet fra CRS. Samtidig ville andelen fra DialogClassic ville være stor. Det vil sige at jo bedre en forfatter er repræsenteret i WoS forfatterregister, jo færre citationer fundet i CRS vil blive genfundet ved hjælp af DialogClassic.

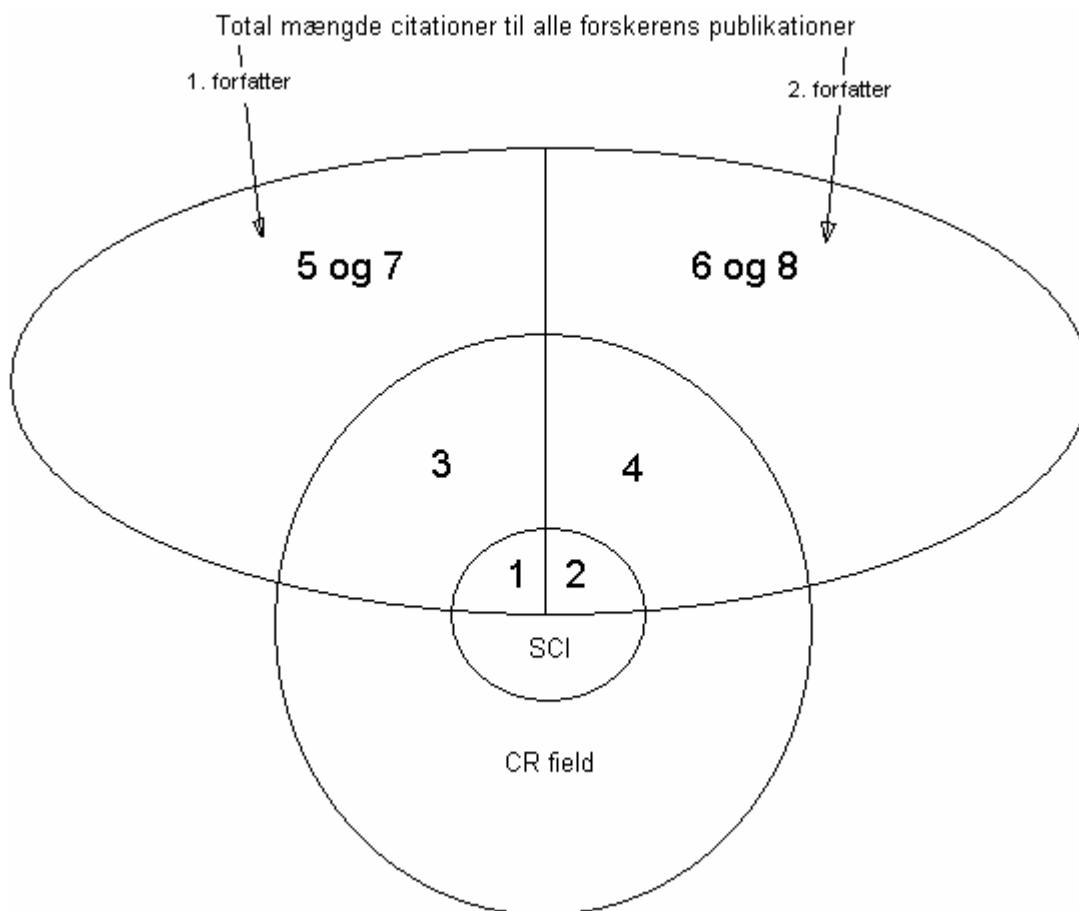
Disse resultater er faktisk meget interessante. Grunden hertil skal findes i den måde SCI er indekseret på, samt hvorledes den måde passer til den enkelte forskers publikationspraksis. De ”stærke” (i SCI sammenhæng) forskningsområder er godt

repræsenteret i SCI forfatterregister/hovedregister da mange publikationer indenfor disse områder automatisk indekseret af SCI. Samtidig er andelen af genfundne citationer i DialogClassic lavere da andelen af andenforfatterskaber er større. Dette kan illustreres af følgende model.



Model I. Forskningsområde der har stærk repræsentation i SCI. Tallene henviser til de 8 citationsscenarioer beskrevet tidligere. Som det kan ses er område 3 forholdsvis mindre end områderne 1 og 2 tilsammen, derfor er der ikke ligeså stor forskel på søgningerne fra CRS og WoS forfatterregister/hovedregister.

Derfor er det også logisk at de (i SCI sammenhæng) marginaliserede forskningsområders andel af citationer indhentet i WoS forfatterregister/hovedregister er lille i forhold til det samlede indhentede citationstal, fordi den pågældende forskers publikationer sandsynligvis ikke er indekseret i SCI. Derfor vil der samtidig også være en forholdsvis stor andel af citationerne der vil blive genfundet i DialogClassic søgningerne på trods af at disse ikke indeholder citationer fra andenforfatterskaber. Andenforfatterskabernes betydning formindskes når hovedvægten af forskerens citationsmængde kommer fra forskningsområder der ikke er indekseret i SCI. Dette kan illustreres af følgende model.



Model II. Forskningsområde der har svag repræsentation i SCI. Tallene henviser til de 8 citationsscenarioer beskrevet tidligere. Som det kan ses er område 2 forholdsvis mindre end områderne 1 og 3 tilsammen, derfor er der ikke ligeså stor forskel på søgningerne fra CRS og DialogClassic.

Disse scenarier har en selvforstærkende effekt da man oftest får citationer fra områder der er nær ved ens forskningsområde. Er dette ikke indekseret vil det være sandsynligt at en stor del af citationerne heller ikke bliver det. Denne metode kan bruges til at indikere hvor stor en del af forskningsområdet der er indekseret udenfor SCI databasen ved at undersøge hvor stor en andel af citationerne der kan genfindes i henholdsvis DialogClassic og WoS forfatterregister/hovedregister.

Da CRS dækker flest områder, nemlig områderne 1,2 og 3 bruges denne som benchmark.

DialogClassic dækker områderne 1 og 3.

WoS forfatterregister/hovedregister dækker områderne 1 og 2.

Er andelen af genfundne citationer i DialogClassic (nr. 1 og 3) stor indikerer der en svag dækning i SCI databasen.

Er andelen af genfundne citationer i WoS forfatterregister/hovedregister (nr. 1 og 2) stor indikerer der en stærk dækning i SCI databasen.

På baggrund af ovenstående overvejelser valgte vi derfor at foretage vores undersøgelse i CRS da denne metode giver det største og efter vores mening bedst egnede datasæt.

4 Citationsanalyse

4.1 Dataindhentning

Citationsanalyserne er lavet i SCI databaseværten WoS via søgemuligheden CRS. Vi har tidligere beskrevet forskellene mellem de tre databaseværter vi har haft adgang til, så her vil vi nu komme en detaljeret beskrivelse af hvordan man forholdsvis hurtigt og effektivt kan lave citationsanalyser indenfor denne vært.

Problemet med CRS i WoS er at den er beregnet til slutbrugeren, hvis formål med databasen slet ikke er at lave citationsanalyser, men derimod at lave citationssøgninger. Men med lidt efterbearbejdning af de frembragte datasæt er det muligt at lave endog ganske seriøse analyser.

Fordelen ved CRS er som sagt at den ligesom søgning i DialogClassic i SCI i CA-feltet hurtigt kan give et overblik over hvor mange citationer en forfatter har modtaget, samtidig med at den i modsætning til DialogClassic også medtager de publikationer hvor forfatteren er andenforfatter. Dette kan vise sig at skabe ganske store forskelle på det endelige resultat.

4.1.1 Indsamling af data

CRS er som sagt beregnet til slutbrugeren hvorfor det er meget let at finde rundt i søgeformularen. Det kan selvfølgelig også have noget at gøre med at det kun er muligt at søge i tre indeks, nemlig citeret forfatter (CA), citeret værk (CW), og citeret år (CY). Samtidig er det kun muligt at indhente 500 poster ad gangen. Dette besværliggør især analyser af forfattere der har udgivet mange publikationer, da de let kan komme op på over 500 poster. Dette kan så afhjælpes med at søge indenfor bestemte år, hvilket dog besværliggør processen da man så skal lave mange flere søgninger for at indhente et samlet forfatterskabs citationer.

4.2 Konvertering af søgeresultater

Konvertering af søgeresultater fra CRS i WoS til resultater der kan præsenteres i dataredigeringsværktøjer som for eksempel Excel.

Man starter med at foretage sin søgning i "CRS" i WoS. I dette eksempel tages udgangspunkt i en søgning af NOVIKOV ID i "CA" feltet. Dette giver et søgeresultat på 443 poster (26.03.03), hvilket er uoverskueligt i WoS visningsformat, da dette kun viser 20 poster ad gangen.

Man begynder med at kopiere resultaterne over i Notesblok. Dette gøres ved at markere og kopiere resultaterne. Dette medfører at koderne fra Internettet fjernes hvilket gør efterbehandlingen af data lettere.

Når man har samlet alle sidernes resultater, i dette tilfælde var de fordelt på 23 sider, kopieres de over i Word hvor de skal manipuleres så de passer ind i Excel.

Her afspiller man så den til lejligheden udformede makro ”CITATIONSANALYSE” som ændrer teksten så den kan konverteres til en tabel. En kopi af den makro vi brugte kan findes på http://ix.db.dk/k00niwi/Noter/makro_citations.htm hvorfra den kan kopieres ind i makro redigeringsprogrammet visual basic.

Derefter afspilles makroen hvorefter teksten gerne skulle være næsten klar til konvertering. Der mangler oftest at ændre _ (underscore) mellem forfatternavn og tidsskriftnavn til ; (semikolon). I dette tilfælde gøres dette ved hjælp af søg og erstat funktionen, hvor man dog skal sørge for at have en markant tekststreng så det ikke er den forkerte tekst der erstattes. I dette tilfælde kan det være [novikov id_] der erstattes med [novikov id;].

Når teksten er markeret er den parat til konvertering. Dette gøres ved hjælp af funktionen Konverter Tekst til tabel der findes i Tabel. Der sørger man for at vælge semikolon som kolonne adskiller, hvorefter teksten bliver konverteret til en tabel. Det er derefter muligt at kopiere tabellen over i Excel. I Excel er der så mulighed for at udnytte de funktioner det program stiller til rådighed, i dette tilfælde især sorteringsmuligheder og sumberegning. Det er derfor muligt lynhurtigt at finde ud af hvor mange citationer pågældende forfatter har modtaget (3399). Da der er forskel på hvorledes navnet skrives alt efter om forskeren har været første eller andenforfatter er det også muligt at udregne hvor mange citationer han har modtaget på publikationer han har været andenforfatter på (1757).

4.3 Diskussion af undersøgelsesobjekterne

En af de oprindelige objektiver med denne opgave var at se på hvorledes forskellene mellem institutterne udmøntede sig i citationstallet.

Vores mål var så at sige at vende bevisførelsen. I stedet for at bruge en bibliometrisk undersøgelse til at måle kvalitet, ville vi måle en bibliometrisk undersøgelse ved hjælp af kvalitet.

Som udgangspunkt valgte vi en population der var uangribelig på deres faglige ekspertise, nemlig professorer ansat på Det Naturvidenskabelige Fakultet på Københavns Universitet. Det er kun forskningseliten forundt at blive professor. En udnævnelse kræver at man kan dokumentere vigtig forskning indenfor det område professoratet dækker, og udvælgelsen er benhård. Derfor mente vi at dette benchmark ville være så bestandigt at det kunne bruges som objektivt undersøgelsesobjekt. Vi vil i følgende afsnit vise de resultater vi kom frem til og derefter diskutere hvorledes de forskelle der er kan forklares.

Vores søgninger er foretaget i CRS, og der var i alt 5 forskellige problemer vi stødte ind i undervejs. Her er en beskrivelse af de forskellige problematikker ved søgning i CRS som vi kom ud for.

1. Søgning på en professor hvis navn kun kan staves på en måde og hvor alle de fundne poster tilhører ham. Et eksempel er professor Øjvind Moestrup fra Botanisk Institut: Søgestrengen i CA feltet er: *moestrup o*. Der blev fundet i alt 165 poster, og alle blev vurderet til at tilhøre ham.
2. Søgning på en professor hvis navn kan staves på flere forskellige måder og hvor alle de fundne poster tilhører ham. Et eksempel er professor Peter Leth-Jørgensen fra August Krogh Instituttet. Søgestrengen i CA feltet er: *lethjorgensen p or jorgensen pl*. Der blev fundet i alt 300 poster, og alle blev vurderet til at tilhøre ham.
3. Søgning på professor hvor der var navnesammenfald med andre forskere. I et sådant tilfælde skal søgesættet renses manuelt for ”falske” poster. Der var tre hovedtilgange til vurdering af poster. Først og fremmest om dokumentets alder passer med forskerens eksempelvis: Forskeren er født i 1949 mens dokumentet er fra 1961, forskeren ville da have været 12 år. Dokumenter der var publiceret indenfor 20 år efter forskerens fødsel blev konsekvent kasseret hvorefter vi gik videre til næste stadie. Og det var emneområdet, hvis dette ligger helt udenfor forskerens eget område blev poster derfra også kasseret. Eksempelvis, forskeren arbejder indenfor geologi mens dokumentet omhandler cancerbehandling. Det var tidsskriftets navn vi her kiggede på. Disse to typer blev straks fjernet. Dokumenter som vi var i tvivl om blev undersøgt nøjere. Her er et screendump fra en søgning i CRS som viser hvorledes søgeresultaterne blev præsenteret.

	Hits	Cited Author	Cited Work	Volume	Page	Year
■	3	HAMANN O	BIODIVERS CONSERV	2	138	1993
■	19	...Hamann O	BIOL CONSERV	44	9	1988
■	12	HAMANN O	BIOL CONSERV	15	215	1979
■	8	HAMANN O	BIOL CONSERV	7	37	1975
■	1	HAMANN O	BIOL DTSCH DICHTER D			1923
■	1	HAMANN O	BIOL GALAPAGOS	8	14	1995
■	2	HAMANN O	BIOL Z	10	440	1890
■	26	...Hamann O	BIOSCIENCE	38	272	1988
■	2	HAMANN O	BIOTROPICA			
■	14	HAMANN O	BIOTROPICA	11	101	1979
■	7	HAMANN O	BOT J LINN SOC	78	67	1979
■	1	HAMANN O	BOT J LINNEAN SOC			
■	3	HAMANN O	BOT NOTISER	132	435	1979
■	6	HAMANN O	BOT NOTISER	127	245	1974
■	5	HAMANN O	BOT NOTISER	127	252	1974
■	7	HAMANN O	BOT NOTISER	127	309	1974
■	1	HAMANN O	BONNS KL ORD TIERRE	2		1904
■	1	HAMANN O	BONNS KLASS ORDNUNG 3	2		1904
■	1	HAMANN O	BONNS KLASS ORDNUNG	2		1904
■	1	HAMANN O	CONSERVATION 21 CENT		137	1997

Screendump fra CRS i WoS

Posterne kan her opdeles i to kategorier. Dem der er markeret som hyperlinks og dem der ikke er. Når posten har hyperlink kan man klikke på den og komme direkte til dens "hoveddata" hvor man kan se alle de almindelige bibliografiske data, samt for eksempelvis (med)forfattere, forfatterens affiliation og abstract. De poster der har hyperlinks er dem der er registreret i SCI databasen. Se i øvrigt afsnit 3.2.3 om problematikken om SCI indekseringspolitik.

De poster der ikke er markeret med hyperlinks kan man til gengæld checke ved at markere posten og via "Search"knappen søge på de artikler der har citeret denne. Hvis de artikler også er indenfor det samme forskningsområde er det sandsynligt at posten tilhører "vores" forfatter.

Der er fire farver for de fire måder vi har vurderet en post.

Rød: Posten kasseres uden nærmere eftersyn, fordi det er umiddelbart synligt at posten ikke har noget med "vores" forsker at gøre. I dette eksempel fordi artiklerne blev skrevet før han blev født!

Orange: Posten skal evt. checkes yderligere da der mangler bibliografiske data.

Turkis og mørkeblå: Posten er rigtig! Derudover har den blå hyperlink og kan derfor checkes.

Et eksempel på en søgning der krævede sortering var professor Ole Jørgen Hamann fra Botanisk Have: søgestrengen i CA feltet: *hamann o*. Der blev fundet i alt 132 poster, og

52 blev vurderet til at tilhøre ham. De resterende 54 blev valgt fra fordi de enten var for gamle, eller fordi de var udgivet indenfor forskningsområder han ikke arbejder i.

4. Søgning på er professor hvor søgesættet overstiger de 500 hits der er adgang til i WoS. Da er det nødvendigt at differentiere søgningen på år, således at man søger på navn kombineret med for eksempel to år (2000 eller 2001) ad gangen. På denne måde er det muligt at indhente alle postene selvom det oprindelige søgesæt overstiger det maksimale antal. Grunden til at der kun medtages to år er fordi ellers vil svartiden blive for lang. Et eksempel er professor Per Bak: En søgning i CA feltet på *bak p* giver 774 hits.

5. Søgning på en professor hvis navn er så almindeligt at det er praktisk umuligt at indhente dennes citationsdata. I teorien kan man sikkert godt men det kræver et kendskab til dennes komplette publikationsliste. Et eksempel er professor Neil Jones: En søgning i CA feltet på *jones n* giver 1218 hits.

4.4 Beskrivelse af de enkelte institutter

Det Naturvidenskabelige Fakultet er delt op i 15 administrative dele: 12 institutter, 2 museer og Botanisk Have. Her kommer en kort beskrivelse af de enkelte institutter hvor mange professorer der er ansat samt hvilke forskningsområder de beskæftiger sig med. Beskrivelser har været nødvendige i forbindelse med vurderingen af posterne i CRS.

August Krogh Instituttet

August Krogh Instituttet har to professorer ansat. Instituttet foretager forskning indenfor en bred vifte af områder fra molekylær biokemi over cellefysiologi til studier om muskelfysiologi.

Instituttet er administrativt delt op i tre afdelinger, biokemisk afdeling, laboratorie for cellulær og molekylær fysiologi og laboratorie for zoologisk fysiologi.

Botanisk Have

Botanisk Have har en professor – Ole Jørgen Hamann – hvis forskningsinteresser ligger indenfor vegetationsdynamik, økologi, naturbevarelse og systematik på Galápagosøerne. Instituttet beskæftiger sig endvidere med eksempelvis taxonomi, fylogeni, tropisk flora og vegetation og udvikling af bevaringsmetoder for genetisk materiale.

Botanisk Institut

Botanisk Institut har 3 professorer. Instituttet har 5 afdelinger : Afdeling for økologi, evolution og biologisk mangfoldighed; Afdeling for evolutionær biologi; Afdeling for Mycologi; Afdeling for Fysiologi samt Afdeling for Fysiologisk Økologi.

Botanisk Museum og Centralbibliotek

Botanisk Museum og Centralbibliotek har en professor – Ib Friis. Instituttet beskæftiger sig med forskellige områder indenfor botanik eks. plantegeografi og diversitet i Afrika og botanisk nomenclatur.

Datalogisk Institut

Datalogisk Institut har 6 professorer. Der er ikke underafdelinger i instituttet. Instituttet forsker bl.a. i semantikbaseret programbehandling, programanalyser og programtransformation, kompleksitetsteori og map theory.

Geologisk Institut

Geologisk Institut har 5 professorer. Der er ingen underafdelinger i instituttet. Der forskes blandt andet i pladetektonik, jordbunde, vandressourcer og klimamodeller.

Geologisk Museum

Geologisk Museum har 2 professorer tilknyttet. Disse beskæftiger sig bl.a. med termodynamik og massetransportprocesser og palæontologi. Der er ingen forskningsmæssige underafdelinger tilknyttet museet.

Geografisk Institut

Geografisk Institut har 8 professorer tilknyttet. Der er ikke tilknyttet underafdelinger til instituttet. Hoved research områder er bl.a. miljømæssige problemer i forskellige sammenhænge (ex. Antarktisk) og agro-industrielt rekonstruktion, migration og urbanisering i udviklingslande.

Kemisk Institut

Kemisk Institut har 5 professorer tilknyttet. Der er 7 centre og 6 laboratorier tilknyttet instituttet.

Institut for Idræt

Der er 2 professorer tilknyttet . Der er 3 afdelinger i instituttet (afdeling for humanfysiologi; afdeling for historie og samfundsvidenskab, afdeling for praktik, pædagogik og psykologi).

Institut for Matematiske Fag

Det er lidt diffust at finde ud af, hvor mange professorer der reelt er på dette institut. Den Kongelige Hof og Stat kalender har listet 5 professorer, men institutter hjemmeside har en del flere. Vi har valgt at undersøge de fem, som vi fandt i Hof & Stat. Institutet har 3 underafdelinger (matematisk afdeling, afdeling for statistik og operationsanalyse og forsikringsmatematisk laboratorium).

Molekylærbiologisk Institut

Molekylærbiologisk Institut har 7 professorer tilknyttet. Der er 8 centre tilknyttet instituttet (generel mikrobiologi, genetik, molekylær celle biologi, plante fysiologi, protein kemi, immunologi og bioinformatik).

Niels Bohr Instituttet for Astronomi, Fysik og Geofysik

Niels Bohr Instituttet for Astronomi, Fysik og Geofysik har 18 professorer tilknyttet. Der er 4 underafdelinger i instituttet (Astronomisk Laboratorium; Geofysisk Afdeling; Niels Bohr Instituttet og Ørsted Laboratoriet)

Zoologisk Institut

Zoologisk Institut har 8 professorer tilknyttet. Der er 6 afdelinger tilknyttet instituttet (Cellebiologi; zoomorfologi; adfærdsbiologi; terrestrisk økologi; populationsøkologi; evolutionsbiologi)

Zoologisk Museum

Zoologisk Museum har 5 professorer tilknyttet. Der er 3 afdelinger under museet.

4.5 Analyse

Det var vores mål at indsamle så nøjagtigt et citationstal som muligt med de redskaber vi havde til rådighed. Som tidligere beskrevet var CRS det bedst egnede søgeredskab til dette formål, da denne vært havde det største recall. Efter søgning på hver enkelt professor hvor vi havde rensset resultatet bedst muligt, har vi nu tabeller som viser citationstallene for hver enkel professor samt for institutterne. Disse er fundet ved at addere de hvert instituts professorers citationstal. Tabellerne kan vise noget om hvor mange citationer hvert enkelt institut modtager, men ikke baggrunden for det. Man kan ikke vurdere om det samlede antal citationer bliver givet fordi instituttet har et stort antal publikationer bag sig, eller om det er fordi der er enkeltpersoner der trækker gennemsnittet op.

Efter at have indsamlet alle citationstallene (der kan ses i bilag I-IV) har vi lavet følgende tabeller:

Ansatte	Institut	Citationer	Cit/pers
16	Niels Bohr Instituttet	71188	4.449
6	Zoologisk Institut	19029	3.172
6	Molekylærbiologisk Inst.	17112	2.852
2	August Krogh	10438	5.219
4	Kemisk Institut	6756	1.689
5	Geologisk Institut	4703	941
2	Idræt	4629	2.315
3	Botanisk Institut*	3926	1.309
5	Institut for Matematiske Fag	3905	781
5	Zoologisk Museum	3020	604
5	Geografisk Institut	1701	340
5	Datalogisk Institut*	1469	294
2	Geologisk Museum	1341	671
1	Botanisk Museum	201	201
1	Botanisk Have	191	191

Tabel 3

Tabellen ovenfor viser Det samlede antal modtagne citationer per institut, samt det gennemsnitlige antal modtagne citationer per person per institut. I toppen ligger Niels Bohr Instituttet og Zoologisk Institut med henholdsvis 71188 og 19029 citationer. For Niels Bohr Instituttets vedkommende er det dog ikke videre overraskende da det også er klart det største institut med 16 professorer. Også Zoologisk Institut er ganske stort, med 6 professorer ansat. Kigger man derimod på det gennemsnitlige antal citationer per professor er det til gengæld August Krogh Instituttet der ligger i front med 5219 citationer mod Niels Bohr Instituttets 4449 og Zoologisk Instituts 3172. I bunden er det Botanisk have og Botanisk Museum der er placeret, med henholdsvis 191 og 201 citationer. Dette har dog også noget at gøre med at der kun er en professor ansat. Kigger man derimod på gennemsnittet begynder Geografisk og Datalogisk Institutter at blande sig med henholdsvis 294 og 340 citationer. Disse tendenser er endnu mere tydelige når man isolerer de 5 mindst og mest citerede forskere.

Institut	Navn	Citationer fra CRS	Andel fra WoS forf. register	Andel fra DialogClassic
Niels Bohr Instituttet	Bak P	13588	12183	10840
Zoologisk Institut	Fenchel T	10662	7389	7995
Niels Bohr Instituttet	Als-Nielsen J	8228	7976	2390
August Krogh	Leth-Jørgensen P	7441	6572	5248
Niels Bohr Instituttet	Hansen JD	7167	5814	706
Botanisk Have	Hamann O	191	112	137
Datalogisk Institut	Pisinger D	126	81	117
Geografisk Institut	Matthiessen C	81	- (6)	- (12)
Geografisk Institut	Reenberg A	75	30	71
Institut for Matematiske Fag	Kiming I	34	28	34

Tabel 4

Noter til tabellen: Søgning på JD Hansen i WoS forfatterregister viste at der var homonymer. Dette blev afhjulpet ved at begrænse søgesættet til poster fra Niels Bohr Instituttet. Interessant var det også at mange af hans publikationer havde endog meget store forskergrupper som forfattere. Dette kan muligvis forklare det lave citationstal fra DialogClassic, da denne vært ikke medtager andenforfatteres citationer.

Professor C Matthiessens citationstal er meget lavt, dette kan blandt andet være fordi han både har udgivet under navnet Matthiessen C og Matthiessen CW. Blandt citationerne til Matthiessen C var det ikke klart hvilke der tilhørte "vores" forsker hvorfor de er udeladt. Vi har derfor valgt at sætte hans resultat i parentes og udeladt at kommentere på det.

Tabellen viser de fem professorer med flest og færrest citationer samt hvilket institut de kommer fra. Desuden er der kombineret med citationsandele fra de to andre metoder til citationsindhentning; data fra DialogClassic og WoS forfatterregister.

Niels Bohr Instituttet er repræsenteret med 3 professorer blandt top 5, mens Geografisk Institut har 2 professorer blandt de mindst citerede. Spekulationerne kan her gå på hvorvidt det er institutternes disciplinære forskelligheder der kommer til udtryk i forskellene på antallet af modtagne citationer. Er det de smalle forskningsområder der får de færreste citationer fra SCI, eller er det blot fordi der er forskel på kvaliteten af den forskning der udøves?

Vi kan egentlig ikke rigtigt gå videre med analyse på disse tal, men hvis man derimod konverterer tallene til indekstal hvor 100 er Citationer fra CRS fås denne tabel:

Institut	Navn	Indeks CRS	Indeks WoS forf. register	Indeks DialogClassic
Niels Bohr Instituttet	Bak P	100	90	80
Zoologisk Institut	Fenchel T	100	69	75
Niels Bohr Instituttet	Als-Nielsen J	100	97	29
August Krogh	Leth-Jørgensen P	100	88	71
Niels Bohr Instituttet	Hansen JD	100	81	10
Botanisk Have	Hamann O	100	59	72
Datalogisk Institut	Pisinger D	100	64	93
Geografisk Institut	Matthiessen C	100	- (7)	- (15)
Geografisk Institut	Reenberg A	100	40	95
Institut for Matematiske Fag	Kiming I	100	82	100

Tabel 5

Hvis vi forudsætter at de mest citerede forskeres forskningsområde er blandt de stærke forskningsområder vil indekstallene fra WoS forfatterregister være større end DialogClassic.

Hvis vi samtidig forudsætter at der blandt de færrest citerede forskeres forskningsområder er svag repræsentation i SCI vil indekstallene fra DialogClassic være større end tallene fra WoS's forfatterregister.

Disse forudsætninger holder stik på alle undtagen professor T Fenchel. Hvis man kigger på de tre professorer fra Niels Bohr Instituttet er det til gengæld tydeligt at se at det er et forskningsområde der er stærkt repræsenteret i SCI. Alle tre forskeres indekstal fra forfatterregisteret i WoS er højere end 80, en kommer helt op på 97. Dette betyder at hvis man søger i forfatterregisteret genfindes over 80% af de samlede mulige antal citationer. Blandt de svagt citerede forskere er der også en tydelig tendens, blot med modsat fortegn. Hvis man ser bort fra Professor Matthiessen hvis tal muligvis ikke valide, er der en meget stor genfindingsrate fra DialogClassic, mens WoS forfatterregister ikke finder særligt mange af citationerne.

4.6 Delkonklusion

Citationsanalyserne på forskerne viste at der er en verden til forskel på hvor mange citationer man kan finde i SCI alt efter hvilket forskningsområde man har. Kigger vi på de rene citationstal var forskellen mellem de fem mindst og de fem mest citerede faktor 92, hvilket betyder at de mest citerede modtog i gennemsnit 92 gange så mange citationer som de mindst citerede gjorde. Dette regnes ud ved at tage summen af de mest citeredes tal og dele med summen af de mindst citeredes tal.

Det er interessant at der ikke er nogen institutter der optræder i både bund og top. Dette kan tyde på at der ikke er kvalitetsforskel men derimod forskel på citationspraksis disciplinerne imellem. Desuden er der en indikation på at de lavt citerede forskningsområder falder uden for SCI's indekseringsområde. Når smalle forskningsområder i den grad bliver forbigået i en citationsanalyse vil den derfor i sig selv være et dårligt evalueringsredskab.

5. Publikationsanalyse

5.1 Metode

Publikationsanalyserne er lavet i DialogClassic. Dialog indeholder som databasevært omkring 300 databaser, som dækker over et stor akademisk område. Da det er muligt at lave søgninger i flere databaser ad gangen (via Onesearch), er Dialog ideel til vores publikationsanalyse. Da vi ikke altid vidste, hvilke databaser, som ville være relevante til de enkelte publikationsanalyser, besluttede vi i første omgang at definere databaser via søgninger i Dialindex.

Vores første strategi til at definere databaser var, at udlede emneord via de enkelte institutters hjemmesider. Denne plan viste sig, at være alt for usikker. På grund af størrelsen valgte vi at dele Niels Bohr Institutet, så vi undersøgte de 3 underafdelinger hver for sig. Samtidig søgte vi samlet på hele instituttet. Søgninger viste, at der var meget stor forskel på, hvilke databaser som kom blandt de 10 mest relevante (se eksempel i bilag VII). Dette indikerer, at der er stor forskel på, hvor professorerne ved de enkelte institutter publicerer og derfor forskel på, hvilke databaser der vil være relevante for os at bruge. Vi fandt det alt for usikkert, at basere publikationsanalysen på grundlag af ordabstraction fra hjemmesiderne. Risikoen for, at vi ville få en skævt billede af forskere, som publicerer i smalle tidsskrifter optaget i fagspecifikke databaser, var for stor. Analysen ville ikke blive valid. Så vi valgte at lave søgninger på hver enkelt professor som forfatter i Dialindex.

Et af problemerne ved at bruge en stor databasevært som Dialog er, at der er problemer med navneforme. Et er problemet med homonymer, hvor flere forskellige forfattere hedder det samme. Men de forskellige databaser bruger også forskellig metode at anføre forfatternavnene på (med eller uden komma mellem efternavn/fornavn, initialer eller fuldt navn, punktum efter navn etc.). Før søgningen i Dialindex konstruerede vi en søgestreng for den enkelte professor med de navneformer vi kunne risikere at finde med ham. Et ekstra check var den citationsliste, som vi allerede have til rådighed fra citationsanalyserne, hvor forskellige stavemåder fremgik. Variationer med mellemrum, punktummer og kommaer blev tilføjet. Vi vil gennem metodebeskrivelsen benytte Professor Niels P Kristensen fra Zoologisk Museum som eksempel på vores søgestrategi:

S AU=((KRISTENSEN NIELS P) OR (KRISTENSEN, NIELS P) OR
(KRISTENSEN, NP) OR (KRISTENSEN NP) OR (KRISTENSEN, NP.))¹

Opslaget i Dialindex på forfatternavn viser det antal databaser, hvor forfatternavnet optræder i publikationer, og hvor mange poster i databasen navnet forekommer. På den måde finder man forholdsvis let ud af, hvilke databaser er mest relevante for søgninger efter professorens publikationer. Man kan ikke direkte søge i Dialindex, men skal fastsætte hvilken superkategori som vil passe til ens søgning. Da vi undersøger forskellige områder indenfor det naturvidenskabelige felt, valgte vi at bruge kategorien AllScience. Allscience er en kategori der dækker 227 databaser, hvis emneområde er indenfor naturvidenskab.

Vi har udvalgt 5 baser, hvori den undersøgte professor havde det største antal poster, med den undtagelse, at hvis der kun var ganske få databaser til rådighed, blev de alle valgt til videre søgning. Til illustration vises Dialindex søgningen på Professor Niels P Kristensen:

```
S AU=((KRISTENSEN NIELS P) OR (KRISTENSEN, NIELS P) OR (KRISTENSEN, NP) OR  
(KRISTENSEN NP) OR (KRISTENSEN, NP.))
```

```
      Items      File  
      -----  
          13      5: Biosis Previews(R)_1969-2003/Apr W3  
          10      34: SciSearch(R) Cited Ref Sci_1990-2003/Apr W3  
Examined 50 files  
          5      144: Pascal_1973-2003/Apr W2  
          1      155: MEDLINE(R)_1966-2003/Apr W3  
          22      185: Zoological Record Online(R)_1978-2003/Apr  
Examined 100 files  
          19      434: SciSearch(R) Cited Ref Sci_1974-1989/Dec  
          20      440: Current Contents Search(R)_1990-2003/Apr 24  
          7 files have one or more items; file list includes 227 files.
```

One or more terms were invalid in 111 files.

B 185,440,434,5,34,144,155

(de udvalgte baser)

Efter databasevalg blev søgestrengen på forfatternavn genbrugt til at søge efter publikationer i de valgte databaser. Remove Duplicats funktionen blev brugt til at fjerne dubletter fra søgesættet. Til sidst rankede vi på RANK AU CONT ALL :

```
DIALOG RANK Results  
RANK: S2/1-56 Field: AU= File(s): 185,440,434,5,34,144,155  
(Rank fields found in 56 records -- 35 unique terms)  
RANK No.      Items      Term  
1              29      KRISTENSEN NP  
2              27      KRISTENSEN NIELS P  
3              9       NIELSEN ES
```

¹ Da vi startede søgninger valgte vi hurtigt at droppe stavemåderne som indeholdt komma, da antallet af hits var enten ikke-eksisterende eller meget lille.

4	9	REPRINT)
5	7	NIELSEN EBBE S
6	2	DAVIS DONALD RAY
7	2	DUGDALE JOHN S
8	2	HEATH J
9	2	KARSHOLT OLE
10	2	KLASS KD

(kun første 10 ud af 35 terms er vist)

Dette havde to formål. For det første var det et check på søgningens validitet. Vi har brugt forskellige stavemåder på den undersøgte professor, og den første på rank-listen *skal* være en af dem. De øvrige stavemåder *skal* også være blandt de første på rank-listen. Tilsammen skal de forskellige stavemåder sammenlagt svare til antal fundne records I det viste tilfælde blev der kun fundet to forskellige stavemåder af Niels P Kristensens navn i de valgte databaser, som stemmer overens med antallet af records. Det andet formål var, at vi havde en ide om at antallet af forfattere (terms) kunne bruges til at udregne en gennemsnitlig forfatterfrekvens på artiklerne. Dette ville kunne bruges til at analysere på citationstallene. Hvis en professor generelt publicerer med flere forfattere, kan dette have betydning for det samlede citationstal. Dels vil citationsandelen, hvis man vælger at lave fraktionerede citationstal, være lavere da professorens andel af citationstallet ville være citationstal/gennemsnitlig antal forfattere. Dels kan flerforfatterskab have en betydning for selvcitationstal (se afsnit 2.2.2 og 2.2.4 vedr. dette). Desværre viste databaserne at være for vilkårlige med henhold til konsistens i forbindelse med anførelse af forfatterskaberne, at dette ikke kunne lade sig gøre. Nogle databaser skriver navnene på forfatterne flere gange (eks. både som Kristensen NP og som Kristensen Niels P). I sådan et tilfælde kan vi risikere at et søgesæt med 100 poster indeholder den enkelte forfatters navn 170 gange. At bruge disse data til udregning af forfatterfrekvens vil være alt for usikkert. Alle data blev herefter overført til et text-dokument (se eksempel på publikationsanalyse i bilag VII). De relevante tal er ført ind i et Excel regneark.

5.2 Analyse

Institutterne har professorer ansat, som har vidt forskellige forskningsområder, og herudover er der forskel på hvor længe den enkelte professor har været aktiv (simpel aldersforskel), hvilket kan have indflydelse på, hvor mange publikationer den enkelte professor har. Vi var nervøse for, at hvis vi begyndte at lave tilfældige stikprøver, kunne dette have stor betydning for tallene specielt for de små institutter med kun en eller to

professorer. Dette er et generelt problem med undersøgelsesområder med en lav population. Faren for at det bliver statistisk forvredet er større når undersøgelsesgrundlaget er lille end når det er stort. Hvis vi skulle have lavet en stikprøve, som svarer til antal professorer pr. institut i forhold til det samlede antal professorer, ville de små institutter blive overrepræsenteret, fordi de forholdsmæssigt ikke engang skulle stille med 1 professor.

5.2.1 De samlede tal

En publikationsanalyse, som kun indeholder rådata uden at differentiere tallene, kan kun give en overfladisk vurdering af institutterne, som ikke direkte kan bruges til evaluering. Tallene viser intet om institutternes og professorernes disciplinære forskellighed, som har indflydelse på publikationstallene. Blandt andet er der forskel på, hvilke tidsskrifter de enkelte forskningsområder publicerer i.

Publikationsanalysen i rene tal er repræsenteret således:

Ansatte	Institut	Publ	Publ/pers
16	Niels Bohr Instituttet	2666	167
6	Zoologisk Institut	899	150
6	Molekylærbiologisk Inst.	822	137
5	Geologisk Institut	813	163
4	Kemisk Institut	674	169
5	Zoologisk Museum	565	113
2	Idræt	552	276
5	Geografisk Institut	319	64
5	Institut for Matematiske Fag	314	63
2	August Krogh	288	144
3	Botanisk Institut*	278	139
5	Datalogisk Institut*	265	66
2	Geologisk Museum	253	127
1	Botanisk Have	47	47
1	Botanisk Museum	128	128

Tabel 6: Antal publikationer pr. institut samt gennemsnitlig antal publikationer pr. professor

Botanisk Institut og Datalogisk Institut har hver en professor, som vi af tidshensyn valgte ikke at undersøge via Dialog, da navnene var homonymer med andre forfattere.

Ikke overraskende er Niels Bohr Instituttet det institut med det største antal publikationer, da det har det største antal professorer tilknyttet. Gennemsnitsværdien for publikationer ligger på 167 pr. professor. Dette gennemsnitstal bliver dog overgået af Kemisk Institut (169 publikation pr. professor) og Institut for Idræt (276 publikationer pr. professor)

Botanisk Have er det institut med det laveste antal publikationer (der er kun 1 professor, som har udgivet 47 publikationer if. Dialog) efterfulgt af Institut for Matematiske Fag (63 publikationer pr. professor) og Geografisk Institut (64 publikationer pr. institut). At gå ind og analysere videre på disse tal, vil være rene spekulationer. Man kunne begynde at spekulere i forskellene i disciplinerne, gennemsnitlig alder for professorerne osv. Følgende vil vi vise de fem professorer med henholdsvis flest og færrest publikationer, hvilket institut de kommer fra samt deres alder, for at se, om man ud fra dette kan støtte de ovennævnte spekulationer:

Institut	Navn	Alder	Publikationer
Niels Bohr Instituttet	Hansen JD	62	382
Kemisk Institut	Nielsen OJ	48	373
Niels Bohr Instituttet	Als-Nielsen j	65	345
Idræt	Richter EA	50	330
Zoologisk Institut	Fenichel T	62	277
Geografisk Institut	Matthiessen C	57	11
Institut for Matematiske Fag	Kiming I	40	19
Niels Bohr Instituttet	Tscherning CC	60	34
Institut for Matematiske Fag	Jensen CU	66	35
Niels Bohr Instituttet	Quadfasel D	52	38

²Tabel 7

Niels Bohr Instituttet har 2 professorer blandt de mest publicerende – men også 2 blandt de mindst publicerende. Publikationsraten har åbenbart ikke noget med instituttet at gøre. Aldersmæssigt ligger disse 4 tæt på hinanden (mellem 52-65 år), så det er ikke denne faktor, som har spillet nogen betydning. Så selv indenfor det samme institut kan der ikke gives et entydigt svar medmindre man minutiøst gennemgår hver enkelt professors forskningsområde og traditionerne indenfor dette specifikke felt.

5.2.2 Publikationstal kombineret med lokal database

For at opnå en mere fuldstændig publikationsliste, er det nødvendigt at undersøge lokale databaser for publikationer. Vi vil give et eksempel på Professor Jon Fjeldså fra Zoologisk Museum. Ved at checke op på den listen over citeringer til Jon Fjeldså, som vi tidligere har lavet, kan vi se, at han ofte publicerer i lokale tidsskrifter (eksempelvis Steenstrupia, som udgives af Zoologisk Museum. Og som ikke er blandt de indekserede tidsskrifter hos ISI). Tallene fra publikationsanalysen foretaget hos databaseværtten Dialog viser, at der for Zoologisk Museum i alt er 565 publikationer, som svarer til 113 publikationer pr. professor i gennemsnit. Gennemsnitstallet for publikationer pr. institut er på 625 publikationer, så Zoologisk Museum ligger lige omkring gennemsnittet. Og

² bilag III viser publikationstallene for de 66 professorer, som vi har lavet publikationsanalyse for.

med 5 professorer ansat er instituttet i mellemstørrelsen. Jon Fjeldså er noteret for 231 publikationer (se bilag III)

Når man undersøger Jon Fjeldsås publikationer i DanBib, som er en national database, får vi følgende resultat:

Fund	Base	Søgning
149	Danbib	lfo=fjeldså jon
10	Danbib	lfo=fjeldså jon og ti=fuglene
98	Danbib	lfo=fjeldså jon og sp=dan
20	Danbib	lfo=fjeldså jon og sp=eng
78	Danbib	Lfo=fjeldså jon og ma=bå (materiale=alle typer bøger)
88	Danbib	Lfo=fjeldså jon og kat=(mono eller peri) (monografier el. periodika)
60	Danbib	Lfo=fjeldså jon og kat=(ana) (kun artikler)

Tabel 8: Søgning på Jon Fjeldså's publikationer i Danbib

Søgningen viser, at der er 149 publikationer af Jon Fjeldså i Danbib. 10 af dem var fra "Fuglene i Danmark", som er et opslagsværk, hvor Jon Fjeldså er redaktør. De bør dog tælles med alligevel, da redaktør (editor) bliver kategoriseret som ophav. 98 af publikationerne er danske og 20 er engelsksprogede. Af disse var 17 monografier, og generelt har Jon Fjeldså mange publikationer i form af monografier (78 i alt i Danbib). Der er generel usikkerhed omkring registreringen af monografier i de større internationale databaser. Eksempelvis er monografier stort set ikke-eksisterende i citationsdatabaserne. Søgningerne på henholdsvis monografier, periodica og tidsskriftsartikler giver tilsammen et resultat på 148.

Søgning i PUF (det elektroniske publikationsregister ved Københavns Universitet) på Jon Fjeldså giver et resultat på 132 publikationer. Basen går tilbage til 1991, hvorefter man opfordres til at checke universitetets årbøger for oplysninger. Vi gentog vores søgning på Jon Fjeldså, og rankede søgesættet på udgivelsesår. Dette gav et resultat på 128 publikationer for 1991 og frem. Så umiddelbart er der ikke den store forskel på antallet i henholdsvis PUF og databaserne i Dialog. Ideelt burde man checke publikationstitlerne fra de forskellige søgninger for at opnå et mere validt resultat. Dette er dog meget tidskrævende, så denne arbejdsproces har vi været nødt til at fravælge. Vi gentog Dialogsøgningen for Jon Fjeldså for at undersøge sprogkoder, for at definere hvor mange af publikationer fra Danbib kunne være registreret i Dialog :

```
RANK: S2/1-231   Field: LA=   File(s): 185,5,440,34,27
(Rank fields found in 219 records -- 7 unique terms) Page 1 of 1
RANK No.  Items  Term
1         179    ENGLISH
2         24     DANISH
3         9      NORWEGIAN
4         7      NONENGLISH
5         6      SPANISH
6         2      GERMAN
```

49 ikke-engelske publikationer var medtaget i Dialog. Der blev fundet 129 ikke-engelsksprogede publikationer i DanBib, hvorfor der kan formodes at der fra søgningen i DanBib er 80 dokumenter der ikke kan findes i Dialog. Det korrekte publikationstal kan derfor komme op på 311, da publikationsanalysen i Dialog gav et resultat på 231. Det vil sige, at ca. 26% af Jon Fjeldsås publikationer ikke findes ved publikationsanalyser i databaseværten Dialog. Hvis dette er en generel tendens hos professorerne i instituttet (og andre smalle discipliner), har dette betydning for det samlede antal publikationer. Og for forholdstallet mellem citationer og publikationer, men mere om dette senere.

For at se, om vi kunne spore en tendens, lavede vi søgninger på fire udvalgte professorer:

Erik Richter og Tom Fenchel, som begge er repræsenteret blandt de fem mest publicerende professorer i Dialogsøgning.

Carl Christian Tscherning og Christian W Matthiesen, som er blandt de fem mindst publicerende professorer i Dialogsøgning.

Søgning i DanBib på Erik Richter:

Fund	Base	Søgning
22	Danbib	Lfo=richter e
34	Danbib	fo=richter e
2	Danbib	fo=richter e og sp=dan

Tabel 9: Erik Richter/Institut for Idræt

Der var ingen poster i Danbib Erik Richter. De poster, som kom frem ved at søge på "Richter E" havde ingen tilknytning til Professor Erik Richter.

Dialogsøgning på Erik Richters publikationer, hvor vi rankede på sprog, gav følgende resultat:

```
DIALOG RANK Results
RANK: S2/1-330 Field: LA= File(s): 5,155,34,399,144,434
(Rank fields found in 318 records -- 4 unique terms)
RANK No. Items Term
1 314 ENGLISH
2 4 NONENGLISH
3 3 DANISH
4 1 HEBREW
```

Kun 8 af de 318 poster, som havde sprogkoder tilknyttet (dvs. at 12 af de i alt 330 poster ikke havde denne kode), var ikke-engelske. Heraf var 3 på dansk, som er publikationer vi ikke kunne genfindes i Danbib.

Søgning på Tom Michael Fenchel i DanBib

Fund	Base	Søgning
56	Danbib	lfo=fenchel tom
36	Danbib	lfo=fenchel tom og sp=dan
19	Danbib	lfo=fenchel tom og sp=eng
33	Danbib	lfo=fenchel tom og kat=(mono eller peri)
23	Danbib	lfo=fenchel tom og kat=(ana)

Tabel 10: Tom Michael Fenchel/Zoologisk Institut

Dialogsøgning på Tom Fenchels publikationer, hvor vi rankede på sprog, gav følgende resultat:

```
RANK: S2/1-277  Field: LA=  File(s): 440,5,185,144,34,434
(Rank fields found in 248 records -- 4 unique terms) Page 1 of 1
RANK No.  Items  Term
      1      243  ENGLISH
      2         4  DANISH
      3         1  NON-ENGLISH
      4         1  NORWEGIAN
```

Kun 6 ud af de 243 publikationer, som havde sprogkode tilknyttet (29 af publikationerne havde ikke sprogkode), var ikke-engelsk sprogede.

Publikationstallet incl. de publikationer, vi har fundet i Danbib fratrukket de engelsksprogede og de 6 ikke engelsksprogede fundet i Dialog (56-19-6=31) har han 308 publikationer. Det vil sige, at ca. 10% af hans publikationer ikke blev fundet ved publikationssøgningen i Dialogs databaser.

Søgning på CC Tscherning i DanBib

Fund	Base	Søgning
58	Danbib	Fo=Tscherning C
28	Danbib	Fo=Tscherning C og sp=dan
22	Danbib	Fo=Tscherning C og sp=eng
53	Danbib	Fo=Tscherning C og kat=(mono eller peri)
5	Danbib	Fo=Tscherning C og kat=(ana)

Tabel 11: Carl Christian Tscherning/Niels Bohr Instituttet

Dialogsøgning på Carl Christian Tscherning publikationer, hvor vi rankede på sprog, gav følgende resultat:

```
RANK: S2/1-34  Field: LA=  File(s): 440,34,434,144,89
(Rank fields found in 34 records -- 1 unique terms) Page 1 of 1
RANK No.  Items  Term
      1      34  ENGLISH
```

Ifølge Dialog har han 34 publikationer, hvoraf ingen er på andet sprog end engelsk. Publikationstallet incl. Publikationstallet inklusiv de ikke engelsk-sprogede publikationer fra Danbib er på 70 publikationer. Det vil svare til at ca. 51%

Søgning på CW Matthiessen i DanBib

Fund	Base	Søgning
87	Danbib	lfo=Matthiessen Christian W?
65	Danbib	lfo=Matthiessen Christian W? og sp=dan
10	Danbib	lfo=Matthiessen Christian W? og sp=eng
40	Danbib	lfo=Matthiessen Christian W? og kat=(mono eller peri)
47	Danbib	lfo=Matthiessen Christian W? og kat=(ana)

Tabel 12 : Christian W Matthiessen/Geografisk Institut

Dialogsøgning på Christian W Matthiessen publikationer, hvor vi rankede på sprog, gav følgende resultat:

```
RANK: S2/1-11  Field: LA=  File(s): 440,34,434,118
(Rank fields found in 11 records -- 2 unique terms) Page 1 of 1
RANK No.  Items  Term
      1      10  ENGLISH
      2       1  GERMAN
```

Ifølge Dialog har han 11 publikationer, hvoraf en enkelt er på andet sprog end engelsk. Publikationstallet inklusiv de 76 ikke engelsk-sprogede poster i Danbib er på 87 publikationer. Det vil sige at ca. 87% af hans publikationer ikke blev fundet via Dialogsøgningerne.

Når vi sammenholder resultaterne fra publikationsanalysen i Dialog databaserne med de ikke engelsksprogede publikationerne i Danbib, viser det at 10% af Tom Fenchel/Zoologisk Institut publikationer, 45% af Carl Christian Tscherning/Niels Bohr Instituttet publikationer og 87% af Christian W Matthiessen/Geografisk Institut ikke blev fundet ved publikationsanalysen i Dialog. Erik Richter/Institut for Idræt var ikke repræsenteret i DanBib hvilket tyder på, at han primært publicerer i udenlandske tidsskrifter.

Tom Fenchel og Erik Richter var blandt de top 5 mest publicerende professorer ifølge Dialogsøgningerne. Carl Christian Tscherning og Christian W Matthiesen var blandt de 5 mindst publicerende ifølge Dialogsøgningerne.

Dette kan indikere, at professorer som tilhører et smalt forskningsområde, publicerer meget i lokale tidsskrifter eller andet som medfører et lavt publikationstal i de internationale databaser, vil kunne opnå et signifikant højere publikationstal, hvis man inkluderede lokale databaser i en publikationsanalyse.

5.3 Delkonklusion

Publikationsanalyser som sådan kan kun give et meget groft overslag over, hvor meget der bliver publiceret fra de enkelte institutter. At forsøge at udlede noget fra disse tal, vil være rene spekulationer. Hvis man undersøger de fem mest og de fem mindst publicerende professorer fra de forskellige institutter under Det Naturfaglige Fakultet, viser det sig, at Niels Bohr Institutet er repræsenteret i begge grupper. Dette indikerer, at man ikke umiddelbart kan hævde, at professorerne indenfor et enkelt institut har en højere publikationsrate end professorer fra andre institutter. Det er nødvendig at se på den enkelte professors publikationsmønster, hvilket forskningsområde han beskæftiger sig med og hvilke traditioner der er indenfor dette område når det kommer til forskning, publicering etc., hvis man skal foretage en evaluering af instituttet. For selv indenfor samme institut, som formodes at beskæftige sig med de samme discipliner, vil der være forskel på publikationsmønster.

For at få et mere korrekt publikationstal, kan man med fordel supplere en publikationsanalyse foretaget i internationale databaser med lokale databaser. Vi benyttede Danbib til at supplere publikationsanalysen foretaget via Dialog. Vi udvalgte 5 professorer, som repræsenterede 3 vinkler:

Jon Fjeldså, som vi via citationsliste kunne se publicerede en del i lokale tidsskrifter.

Erik Richter og Tom Fenchel, som begge er repræsenteret blandt de fem mest publicerende professorer i publikationsanalysen fra Dialog.

Carl Christian Tscherning og Christian W Matthiessen, som er blandt de fem mindst publicerende professorer i publikationsanalysen fra Dialog.

Jon Fjeldså, Carl Christian Tscherning og Christian W Matthiessen opnåede signifikant højere publikationstal, når man kombinerer internationale og en lokal database i en publikationsanalyse. Tallet steg ikke væsentligt for Tom Fenchel og Erik Richter (som ikke var repræsenteret i Danbib overhovedet). Dette *indikerer* nødvendigheden af at lave kombinerede publikationsanalyser for forskere, som pga. smalt forskningsområde, publicerer i lokale tidsskrifter o.lign., får lave publikationstal i større internationale databaser.

Da tallene fra en publikationsanalyse lavet hos en stor databasevært som Dialog, ikke markant siger noget om de enkelte institutter eller den enkelte forsker, skal disse analyser laves og anvendes med stor varsomhed.

6. Forhold mellem publikations- og citationstal

I de sidste to afsnit har vi set på udformning og fortolkning af en citations- og publikationsanalyse hver for sig. Umiddelbart kan man ikke bruge en af analyserne selvstændigt, da resultaterne vil give et skævt billede af institutterne.

Forskningsområderne mellem institutterne adskiller sig så meget fra hinanden – og også indenfor det enkelte institut – at det vil være uansvarligt at bygge forskningsevaluering, især tværfaglig forskningsevaluering, på baggrund en af disse analyser alene.

I det følgende vil vi se på resultaterne af en sammenlægning af de to analyser.

Vi har foretaget en sammenligning af citationstal i forhold til publikationstal.

Formodningen er at man vil have to kvantitative undersøgelser at evaluere ud fra, og dette bør kunne sige mere om institutternes end citations- eller publikationsanalysen alene. Desuden vil forholdet mellem publikationer udgivet og antal modtagne citationer give et mere reelt billede af publikationernes betydning. Kvantitet er ikke lig med kvalitet, så det er ikke et kvalitetstegn alene at publicere meget. Det er heller ikke udelukkende et kvalitetstegn at modtage mange citationer, da citationstallene hænger sammen med det forskningsområde man tilhører. Men et forholdstal mellem antal publicerede artikler og antal modtagne citationer kan indikere, i hvilken grad den publicerede forskning bliver ”brugt” af andre forskere.

Vi vil i det følgende se på, hvordan forholdstal mellem publikationer og citeringer viser sig, når vi ser på de enkelte institutter i forhold til hinanden og på enkelte forskere i forhold til hinanden.

6.1 Sammenligning af de enkelte institutters forholdstal

I afsnit 4 og 5 har vi hver for sig analyseret citations- og publikationstal for professorerne fra de 15 institutter under Det Naturvidenskabelige Fakultet. Disse tal har vi samlet på institutniveau i en tabel, som viser citationstallene, publikationstallene, forholdstallene (cit/pub) samt de rank-numre institutterne har i de 3 analyser.

Ansatte	Institut	Cit	Pub	Cit/publ	cit/pub rank	cit rank	pub rank
2	August Krogh	10438	288	36,24	1	4	10
16	Niels Bohr Institutet	71188	2666	26,70	2	1	1
6	Zoologisk Institut	19029	899	21,17	3	2	2
6	Molekylærbiologisk Inst.	17112	822	20,82	4	3	3
5	Institut for Matematiske Fag	3905	314	12,44	5	9	9
4	Kemisk Institut	6756	674	10,02	6	5	5
3	Botanisk Institut*	3926	278	8,99	7	8	11
2	Idræt	4629	552	8,39	8	7	7
5	Geologisk Institut	4703	813	5,78	9	6	4
5	Zoologisk Museum	3020	565	5,35	10	10	6
5	Geografisk Institut	1701	319	5,33	11	11	8
2	Geologisk Museum	1341	253	5,30	12	13	13
1	Botanisk Have	191	47	4,06	13	15	15
5	Datalogisk Institut*	1469	265	3,94	14	12	12
1	Botanisk Museum	201	128	1,57	15	14	14

Tabel 13: Forholdstal mellem publikationer & Citationer for institutterne.

*Botanisk Institut og Datalogisk Institut har hver en professor, som vi ikke kunne lave publikationsanalyse på pga. homonym.

Det er interessant at se, at hvor cit/pub-rank og cit-rank ikke har de store udsving, er der til gengæld forskel på cit/pub-rank og pub-rank. August Krogh Institutet, som ligger som rank nr. 1 i både cit/pub-rank og cit-rank, var nr. 10 på pub-ranklisten. Og Institut for Matematiske Fag, som har ligget som rank nr. 9 i både citations- og publikationsanalysen, ligger som nr. 5, når man sætter tallene i forhold til hinanden. Der er meget stor forskel på faktortallene for henholdsvis publikationer (faktor 12) og citationer (faktor 92), dette influerer på forholdstallene så cit/pub-rank og cit-rank ligger tæt opad hinanden. Derfor følger institutternes cit/pub-rank nummer stort set cit-rank numrene. I afsnittet omkring citationer var vi inde på, at dette kan skyldes hvordan de forskellige forskningsområder er repræsenteret i SCI. Denne skævvridning følger med, selvom vi arbejder med forholdstal, og tallene er derfor misvisende i henhold til forskningsevaluering.

6.2 Sammenligning af enkelte professorer

Forholdstal viste sig ikke være valide at bruge, når man skal sammenligne institutter. Vi vil nu se på, hvilken indflydelse forholdstal har på de enkelte professorer. Vi sammenligner de professorer med de 5 højeste og fem laveste forholdstal. Samtidig viser vi deres publikations- og citationstal, samt hvilke ranknumre professorerne har haft i de forskellige analyser.

Institut	Navn	Cit	Pub	Cit/pub	cit/pub rank	cit rank	pub rank
Niels Bohr Institutet	Bak P	13588	248	54,79	1	1	7
August Krogh	Leth-Jørgensen P	7441	138	53,92	2	4	28
Zoologisk Institut	Fenichel T	10662	277	38,49	3	2	5
Niels Bohr Institutet	Jackson AD	5132	155	33,11	4	11	25
Niels Bohr Institutet	Jensen MH	5155	158	32,63	5	10	24
Zoologisk Museum	Enghoff H	368	178	2,07	62	54	19
Datalogisk Institut	Pisinger D	126	63	2,00	63	65	50
Institut for Matematiske Fag	Kiming I	34	19	1,79	64	68	65
Botanisk Museum	Friis I	201	128	1,57	65	63	29
Geografisk Institut	Reenberg A	75	71	1,06	66	67	48

Tabel 14: Forholdstal mellem citationer og publikationer for professorer med de fem højeste og laveste forholdstal.

Når vi sammenligner de enkelte professorer er det tydeligt at se, at variationen mellem cit/pub ranknumrene varierer langt mere, end ved sammenligningen af de enkelte institutter. AD Jackson og MH Jensen fra Niels Bohr Institutet, som ligger som henholdsvis nr. 4 og nr.5 på rank-listen over forholdstal, figurerede ikke på top-5 listen i hverken citationsanalysen (hvor de lå som nr.11 og nr.10) eller publikationsanalysen (hvor de lå som nr. 25 og nr. 24). Ved professorer med lavt forholdstal er forholdet mellem cit/pub-rank og cit-rank nummer nogenlunde statistisk. Men til gengæld er der stor forskel mellem cit/pub-rank nummer og pub-rank nummer. H Enghoff har nr. 19 på pub-rank listen og I Friis nr. 29, og de ligger som henholdsvis nr. 62 og 65 på cit/pub-listen.

Igen kommer faktoren om forskningsområde, publikationer i SCI og generelle publikations- og citationsmønstre indenfor forskningsområdet ind. Hvis en forsker publicerer indenfor et forskningsområde, som er bredt repræsenteret i de internationale databaser, vil man kunne bruge forholdstallene, *hvis man sammenligner indenfor det samme forskningsområde*. Det stiller sig anderledes an for forskere, som publicerer indenfor forskningsområder, som ikke er bredt repræsenteret i internationale databaser, men derimod publicerer i nationale tidsskrifter. Deres forholdstal vil være meget ustabil, da det er vanskeligt at finde frem til et publikationstal, som kan bruges til udregning af forholdstal. Og forholdstallet vil blive en del mindre, end det man umiddelbart kan regne sig frem til via citations- og publikationsanalyser lavet i store databaser. Som et eksempel kan vi bruge C Matthiessen fra Geografisk Institut. I afsnittet om publikationsanalyser, fandt vi ved at kombinerer publikationstallet fra Dialogsøgningerne (11 pub. fundet) med publikationstal fra Danbib (77 pub. fundet) et publikationstal på 88. Hans forholdstal vil ændre sig fra at være på 7,36 analyserne fra de store databaser, til at være på 0,92 når man tager forholdet mellem det kombinerede

publikationstal og antal modtagne citationer fra SCI. Her skal man også tage i betragtning, at når han publicerer i lokale tidsskrifter, vil han sandsynligvis også modtage flest citeringer fra publikationer i samme tidsskrifter. Og disse vil ikke figurere i citationsdatabaserne, hvorfor tallet under alle omstændigheder bliver skævvredet.

6.3 Delkonklusion

Vi brugte to måder at samle de to analysemetoder på. Den ene gik ud på at sammenholde de to datasæt med hinanden på forholdstalsniveau (citationer delt med publikationer). Den anden metode vi udførte var at lave en korrelationsanalyse som vil vise hvor stærk en statistisk sammenhæng der er mellem citationer og publikationer. Vi vil i dette afsnit konkludere på de to metoder og vurdere deres anvendelighed. Udregninger af forholdstal virker umiddelbart som en god metode til sammenligning af resultaterne. Denne metode har dog det problem, at man skal have to homogene datasæt at sammenligne før man kan få et resultat der er interessant. I denne forbindelse tænker vi her på problemerne med den store forskel i faktorerne mellem citationer og publikationer. Hvor forskellen på de lavest citerede og de højest citerede var faktor 92, var den ved publikationerne kun faktor 12. Dette betyder at det vil være citationernes betydning der vil overskygge publikationens betydning, hvorfor resultatet vil blive skævvredet i retning mod citationerne. De samme skævheder, som vi tidligere har beskrevet under citations- og publikationsanalysen, vil slå igennem. Under alle omstændigheder vil det ikke kunne bruges til sammenligning af institutter fra forskellige forskningsområder, men vil eventuelt kunne bruges individuel analyse af forskere indenfor det samme forskningsområde som er stærkt repræsenterede i de store databaser som for eksempel SCI. Denne metode kan ikke bruges ved forskningsområder som er svagt repræsenteret i de store databaser. Når forskere er repræsenteret bredt i lokale eller nationale tidsskrifter, vil dette ikke figurere i de store internationale databaser. En stor del af de citeringer til deres publikationer vil komme fra de samme tidsskrifter, hvorfor en stor del af deres forskning ikke vil blive fundet ved bibliometriske undersøgelser i databaser som SCI. Når andelen af den fundne totalmængde er mindre, vil usikkerheden angående validiteten også stige, hvorfor det især er blandt de svagt repræsenterede forskningsområder at denne problematik findes.

7. Konklusion

Formålet med opgaven var at undersøge forskellen mellem publikation og citationsmønstrene på de forskellige institutter på Det Naturvidenskabelige Fakultet på Københavns Universitet. Vores hypotese var at der ville være så stor en forskel mellem de forskellige forskningsområder at de kvantitative publikations- og citationsanalyser ikke kan stå alene når der skal foretages analyse på tværs af forskningsområderne. De værktøjer og metoder man kan bruge i de bibliometriske analyser har stor indflydelse på de resultater man får.

Vi foretog en citations- og publikationsanalyse på professorerne på institutterne på Det Naturvidenskabelige Fakultet på Københavns Universitet da professorerne formodes at have den største publikationsrate. Samtidig må denne population på grund af deres stilling som professor betegnes som tilhørende forskningseliten, hvorfor det må formodes at deres forskning er af høj kvalitet.

Vores gennemgang af de forskellige søgemuligheder samt resultaterne fra disse viser at der er stor forskel på hvilken databasevært der bruges til citationssøgningerne. Der var tendenser i den måde forskellene udmøntede sig, alt efter hvilken databasevært man brugte. De svagt repræsenterede forskningsområder havde en lav genfindingsrate fra hovedregisteret i forhold til det totale antal fundne citationer fra CRS. Samtidig var genfindingsraten fra DialogClassic højere end fra WoS forfatterregister. De stærkt repræsenterede forskningsområder havde til gengæld en høj genfindingsrate fra WoS forfatterregister samt en lav genfindingsrate fra DialogClassic. Vi fandt at tendenserne i disse forskelle var systematiske og at de derfor kunne bruges til at vise hvor stærkt forskningsområdet var repræsenteret i SCI. For at illustrere dette konstruerede vi to modeller som visualiserede henholdsvis stærk og svag repræsentation.

Citationsanalyse: Vores formål var at lave en så fuldstændig citationsanalyse som muligt med de redskaber vi har til rådighed. Metoden til dette er nøje gennemgået.

Konklusionen på citationsanalysen var, at der rent faktisk er så stor forskel på de forskellige forskningsområder, så en citationsanalyse på tværs af domænerne ikke kan stå alene. Blandt de fem mest citerede og fem mindst citerede var forskellen en faktor 92. Det vil sige, at de mest citerede i gennemsnit fik 92 gange så mange citationer end de mindst citerede. Dette mener vi ikke kun kan tilskrives en forskel i kvalitet, der skal inddrages faktorer som databasedækning og citationsmønstre indenfor de enkelte forskningsområder.

Publikationsanalyse: Analysen blev lavet i relevante databaser via DialogClassic. Vi definerede databaserne via Dialindex for hver enkelt professor, hvor vi tog højde for de forskellige stave- og indekseringsmåder et navn kunne optræde på i de forskellige databaser. Metoden er nøje beskrevet i afsnittet. Analysen viste, at en publikationsanalyse kun kan give et meget groft overslag over, hvor meget der bliver publiceret fra de enkelte institutter. Ved at se på de fem mest og fem mindst publicerende professorer, var forskellen på de fem mest og de fem mindst publicerende professorer på faktor 12. Det vil sige at de mest publicerende i gennemsnit publicerede 12 gange så meget som de fem mindst publicerende.

Niels Bohr Instituttet repræsenteret begge steder med to professorer. Vi undersøgte om professorernes alder kunne have en indflydelse, men aldersmæssigt lå professorerne tæt på hinanden. Dette kan indikere, at selv indenfor det samme institut vil der være forskel på publikationsrate, alt efter hvilket forskningsområde forskeren beskæftiger sig med. For at lave en evaluering af et instituts forskere er det derfor nødvendigt at se på den enkelte professors forskningsområde og traditioner indenfor dette med hensyn til forskning, publicering etc., for at vurdere kvaliteten af instituttets forskning.

For at få et mere korrekt publikationstal, er det nødvendigt at inddrage lokale databaser. Vi lavede søgninger på publikationer i DanBib for 5 udvalgte professorer:

J Fjeldså, som vi via hans citationsliste kunne se publicerede i lokale tidsskrifter.

E Richter og T Fenchel, som begge er repræsenteret blandt de fem mest publicerende professorer i publikationsanalysen fra Dialog.

CC Tscherning og CW Matthiessen, som begge er repræsenteret blandt de fem mindst publicerende professorer i publikationsanalysen fra Dialog.

Resultatet viste at J Fjeldså, CC Tscherning og CW Matthiessen opnåede et markant højere publikationstal, når man kombinerer internationale og nationale databaser ved en publikationsanalyse. T Fenchels publikationstal steg ikke i væsentlig grad og E Richter var ikke repræsenteret i DanBib. Dette kan indikere nødvendigheden af at lave kombinerede publikationsanalyser når man skal undersøge forskningsområder, som er svagt repræsenteret i de internationale databaser.

Publikationsanalyser som er foretaget i internationale databaser, og som udelukkende baserer sig på optælling af publikationer siger ikke noget specifikt om et institut.

Sådanne analyser kan ikke stå alene i forskningsevaluering, men skal suppleres af mere dybdeborende kvantitative og kvalitative undersøgelser.

Ideen med sammenlægningen af de to analysemetoder var at se om forholdet mellem citationer og publikationer kunne sige noget om den reelle forskel mellem institutterne og deres forskningsområder. Denne metode viste sig dog at være uhensigtsmæssig da forskellene i citationsanalysen fuldstændigt dominerede de forholdstal vi prøvede at udregne. Derfor var de skævheder vi fandt i citationsanalysen stadig til stede i den samlede analyse. Vi vil dog påpege at forholdstallene ser ud til at være velegnede som supplement til forskningsevaluering indenfor et stærkt repræsenteret forskningsområde hvis man sammenligner indenfor det samme forskningsområde.

Den endelige konklusion er at citations- og publikationsanalyser i sig selv ikke er egnede til at evaluere forskning på tværs af domænerne. Hvis man derimod anvender dem som supplement til kvalitativ evaluering og med de iboende begrænsninger for øje kan de være anvendelige, hvis man holder sig inden for et forskningsområde ad gangen. Debatten omkring brugen af de kvantitative bibliometriske metoder til forskningsevaluering har vekslet mellem det positivitistiske syn, som mener at forskning og forskningskvalitet kan ”måles og vejes”, og det pragmatistiske syn, som holder fast på at forskning skal vurderes rent kvalitativt. Det videnskabsteoretiske/sociologiske synspunkt anbefaler, at man kombinerer de kvantitative og kvalitative metoder. Dette synspunkt vil vi tilslutte os, da vi mener at de kvantitative metoder ikke afdækker de årsager der kan ligge til grund for forskernes publikations- og citationsmønstre. Men som et supplerende redskab i forskningsevaluering for et homogent forskningsområde er bibliometriske metoder velegnede.

8. Kildefortegnelse

Amin M & Mabe M (2000). Impact Factors: Use and abuse. Perspective in publishing no.1. Lokaliseret d. 21. maj 2003 på World Wide Web:
<http://www.ece.rochester.edu/users/elstat/perspectives1.pdf>

Budd JM (2000); Scholarly production of U.S. Lis Faculty: An Update. *Library Quarterly*, 70(2): 230-245.

Find, Søren. (1999). *Medicinsk sundhedsforsknings vilkår og fremtid - med særligt henblik på sygehusområdet*. København: Forskningsministeriet og Sundhedsministeriet.

Garfield E; (1979) Perspective on citation analysis of sciences (Kap.10). In: *Citation indexing – its Theory and Application in Science, Technology and Humanities*. Lokaliseret d. 5. maj 2003 på World Wide Web:
<http://www.garfield.library.upenn.edu/ci/title.pdf>

Hayes RM. (1983). Citation Statistics as a Measure of Faculty Research Productivity. *Journal of Education for Librarianship* 23: 151-72.

Hjørland, Birger. (1996). Forskningsevaluering i videnskabsteoretisk belysning. *Biblioteksarbejde*, 48: 35-57.

Institute for Scientific Information. (2001a). *SCISearch Bluesheet*. Lokaliseret den 5. januar 2003 på World Wide Web: <http://library.dialog.com/bluesheets/pdf/bl0034.pdf>

Institute for Scientific Information. (2001b). *Web of Science Indexes: Science Citation Index Expanded*. Lokaliseret den 5. januar 2003 på World Wide Web:
<http://isi2.isiknowledge.com/portal.cgi?DestApp=WOS&Func=Frame>

Kärki R & Kortelainen T (1998) *Introduktion till bibliometri*. Nordinfo 41. Helsingfors.

MacRoberts MH & MacRoberts B. (1996). Problems of Citation Analysis. *Scientometrics*, 36(3): 435-444

Peters HPF & Van Raan FJ. (1994). A Bibliometric Profile of Top-scientists - A Case Study in Chemical Engineering. *Scientometrics* 29(1): 115-136

Rosseau R. (1998) Citation analysis as a theory of friction or polluted air. *Scientometrics* 43(1): 63-67

Seglen, Per O. (1996). Bruk av siteringer og tidsskriftimpaktfaktor til forskningsevaluering. *Biblioteksarbejde*, 48: 27-34.

Seglen PO. (1992). The Skewness of Science. *Journal of the American Society for Information Science*. 43(9):628-638

Snyder H & Bonzi S. (1998) Patterns of Self-Citation Across Disciplines (1980-1989). *Journal of Information Science*, 24(6): 431-435

Winkel, Jasper Steen. (2002) Dansk forskning delt i A- og B-hold. *Mandag morgen*, 43: 22-27.

Supplerende litteratur

Folly G. (1981). Some Methodological Problems in Ranking Scientists by Citation Analysis. *Scientometrics*, 3(2): 135-147

Frolich C & Resler L. (2001). Analysis of Publications and Citations from a Geophysics Research Institute. *Journal of the American Society for Information and Science Technology*, 52(9): 701-713

Hurd JM. (1992). Interdisciplinary Research in the Sciences - Implications for Library Organization. *Coll Res Libr*, 53(4): 283-297

Lange LL. (2001). Citation Counts of Multi-Authored Papers - First-Named Authors and Further Authors. *Scientometrics*, 52(3): 457-470

Royle P & Over R. (1994). The Use of Bibliometric Indicators to Measure the Research Productivity of Australian Academics. *AARL*, June: 77-88

Schwartz CA. (1991). Research Productivity and Publication Output - An Interdisciplinary Analysis. *Coll Res Libr*, 52(5): 414-424

Schwartz CA. (1997). The Rise and Fall of Uncitedness. *Coll Res Libr*, 58(1): 19-29

Van Raan AFJ. (1998). In Matters of Quantative Studies of Science. The Fault of Theorists is Offering too Little and Asking too Much. *Scientometrics*, 43(1): 129-139

9. Bilagsoversigt

Bilag I-IV :Oversigt over alle citations- og publikationsanalyser:

Bilag I: Listen er sorteret efter institutnavn

Bilag II: Listen samlet efter citationer

Bilag III: Listen er samlet efter publikationer

Bilag IV: Listen er samlet efter citationer per publikation (forholdstal)

Bilag V: Eksempel på citationssøgning fra CRS (Novikov ID)

Bilag VI: Eksempel på citationssøgning fra CRS (Novikov ID) indsat i Excel regneark

Bilag VII: Eksempel på publikationssøgning fra DialogClassic (Kristensen NP)

En komplet liste over alle søgninger, både citations- og publikationssøgninger er tilgængelig via Internet på: <http://ix.db.dk/k00niwi/bachelor> .