

MonAquabasen; Trender 2000-2005

Eystein Skjerve

Norges veterinærhøgskole, Senter for Epidemiologi og biostatistikk

Boks 8146 dep., 0033 Oslo

Eystein Skjerve@veths.no

Innledning

Norges veterinærhøgskole, Senter for epidemiologi og biostatistikk (SEB) har i flere år arbeidet sammen med MonAqua omkring en bruk av data og epidemiologisk tolking av data fra MonAqua-basen. I dette spesifikke prosjektet tok SEB på seg ansvaret for (i samarbeid med MonAqua) å gå gjennom den siste versjon av MonAqua-basen som den forelå ved sommeren 2007. En lignende, foreløpig analyse ble også gjennomført i 2005. Fra SEB var Eystein Skjerve ansvarlig for prosjektet

Prosjektet ble gjennomført i følgende faser:

1. Orienterende gjennomgang av databasen, juni 2007 (møte i Trondheim med Skjerve/ Aunsmo fra SEB og Gunnes/ Kosmo fra MonAqua)
2. Revidering og oppdatering av basen, sommeren 2007 (Kosmo)
3. Gjennomgang av revidert base august/ september 2007 (Skjerve)
4. Analyseseminar i Åre 20-21.9 der hele den reviderte databasen ble gjennomgått (Skjerve, Larssen, Valle, Aunsmo fra SEB, Gunnes og Kosmo fra MonAqua)
5. Sluttrevisjon og litt utfylling av basen etter møtet i Åre (Kosmo)
6. Analyse av den sluttreviderte basen i oktober 2007 (Skjerve)
7. Presentasjon av resultatene under Produktivitetskonferansen i Kristiansund 31.10.2007.

Prosjektet har vært drevet interaktivt med jevn kommunikasjon mellom SEB og MonAqua, og med en sterk fokus på biologisk relevante forklaringsmodeller.

Data brukt i analysene

Arbeidet tok utgangspunkt i hele MonAqua-basen fra 1998. På grunn av utvikling av basen og mange situasjoner med manglende verdier på relevante variable i den tidlige del av basen ble det tidlig i prosjektet bestemt å arbeide kun med et datasett som representerer 2000-2005 årgangene i basen. MonAqua har brukt mye tid på oppdatering og revidering av denne basen, og for enkelte grupper variable oppnådde vi en bortimot komplett base med få (<10 %) manglende verdier. For visse grupper av variable var det fremdeles mange manglende verdier,

som gjorde det vanskelig å ta med disse variable i en analyse. Dette gjelder særlig variable som har med føring og vaksinerings å gjøre.

Datasettet illustrerer både styrker og svakheter med MonAquabasen:

1. Den er fleksibel og tilpasses nye fagspørsmål og ønsker fra medlemmene og er derfor godt egnet til en detaljert oppfølging av enkelte anlegg.
2. Basen er omfattende og prøver å fange flest mulig av de variable som kan bidra til det økonomiske resultatet i oppdrett
3. En endring av basen gjør at det kan være vanskelig å etablere trender, da visse viktige variable ikke er tilstede i hele basen.
4. Basens struktur med varierende rekruttering geografisk i forskjellige årganger gjør det faglig vanskelig å analysere datasettet hvis en tillater mange manglende verdier på nøkkelvariable. En kan da få skjevhet inn i materialet som i realiteten kun er en eksponent for rekrutteringsgruppene og ikke biologiske forklaringsvariable.

Databasen utvikles kontinuerlig videre, og våre analyser kan forhåpentligvis bidra litt til utvikling av basens elementer for både å fungere i forhold til benchmarking for det enkelte anlegg og samtidig gi et rimelig korrekt bilde av populasjonen av oppdrettsanlegg i Norge og de faktorer som dominerer når det gjelder økonomisk utbytte.

Analysemetode

Det er mange tilnærminger til statistisk analyse av en slik database. I utgangspunktet var vi interessert i å finne mulige viktige faktorer **på populasjonsnivå** som bestemmer nivået av de viktigste økonomiske indikatorene innen oppdrett. Følgende utfallsvariable ble benyttet: **Biologisk forfaktor (BFF)**, **økonomisk forfaktor (ØFF)**, **Tilvekst**, **Svinn %** og **Superior %**.

Som statistisk metode ble valgt å bygge **multivariate regresjonsmodeller**, der utfallsvariabel testes i forhold til forskjellige kombinasjoner av forklaringsvariable. Slike modeller er nært koblet til vurdering av årsaksfaktorer og gir også informasjon om forklaringsgrad. Metodisk ble det brukt den teknikken som er anbefalt innen moderne veterinær epidemiologi (Referanse: Dohoo, Martin, Stryhn, *Veterinary Epidemiological Research*; <http://www.upei.ca/ver/>). Alle analysene ble gjennomført på data fra Excel-basen, overført til statistikkprogrammet Stata/ SE for Windows versjon 10.0. På grunn av mange variable og delvis manglende verdier for visse variable ble følgende prosedyre brukt:

1. Eliminering av variable med stor andel manglende verdier

2. Beskrivende analyse av utfallsvariable og antatt viktige forklaringsvariable
3. Modellbygging med utgangspunkt i de antatt viktigste biologiske forklaringsvariable for hver enkelt utfallsvariabel. Modeller ble utvidet med en og en variabel inntil forklaringsgraden ikke økte vesentlig ved inklusjon av nye variable.
4. Diskusjon av modell/ konkurrerende modeller opp mot biologisk forklaringsverdi
5. Justering av modell med bruk av avanserte metoder for vurdering av grad av avhengighet på lokaltitetsnivå.
6. Tradisjonell analyse av modellens robusthet og forklaringsgrad, inklusive residualanalyse.

Resultater fra analysene

Vedlagte presentasjon viser de viktigste resultatene som presentert ved Produktivitetskonferansen 31.10 2007. Punktvis kan resultatene oppsummeres som følger:

1. En fikk etablert stabile modeller for de to viktigste økonomiske indikatorene (BFF/ ØFF). BFF og ØFF er beslekta, derfor også beslekta forklaringsvariable
 - a. Forklaringsgrad gikk opp mot 25 % av variasjon i data; >75 % er uforklart
 - b. Forklaringsvariable som går igjen
 - i. Breddegrad/ region/ dager produksjon/ lys benyttet
 - ii. 1-åring vs. 0-åring
 - iii. Sortering
 - c. Faktorer kobla til foring/ vaksine mer ustabil i modellene
2. For **Svinn** fikk vi i mindre grad etablert stabile modeller. Svinn preges av episoder, og til tross for delvis samme risikofaktorer som BFF og ØFF er store svinn forklart av sjukdom eller uhell. Statistiske modeller med vekt på populasjonsgjennomsnitt fungerer dårligere når en har slike fenomener.
3. For Tilvekst ble det valgt å ikke slutføre analysene Materialet skal bearbeides videre i forbindelse med Arnfinn Aunsmos doktorgrad. Ett av problemene er koblet til modellering av tilvekst gjennom forskjellige livsfaser og at tilvekst kan måles på mange forskjellige måter, % tilvekst; gram/ dag; matematisk modellert etter temperatur, lys med mer.
4. For % **Superior** viste analysene markerte trender over tid. Til en viss grad kan dette forklares ved at graderingen er delvis avhengig av kunnskaper om marked og pris. Sånn sett vil den endres over tid og vil ikke være egnet til en statistisk behandling.

Hovedkonklusjoner og anbefalinger

Prosjektet har fra SEB's side vært en svært interessant øvelse i samarbeid med MonAqua. Analysen av slike databaser kan ikke gjennomføres på en fornuftig måte uten et svært tett samarbeid mellom spesialister innen epidemiologi/ statistikk og de som kjenner basen. Samtidig må alle modellene diskuteres grundig fra et biologisk ståsted, og vi mener å ha oppnådd det ved bruk av den felles erfaringsbakgrunn fra oppdrettsnæringa fra deltaker i prosjektet fra SEB og MonAqua.

En hovedverdi av prosjektet har vært å gå forbi de ofte misvisende resultatene en kan få ved enkel grafisk analyse av trender eller enkle statistiske assosiasjoner mellom en utfallsvariabel og en forklaringsvariabel. Verdien av en multivariabel analyse er at en i stor grad kan luke ut slike, antatt falske, sammenhenger og slik bedre kunne forkusere på de som antas å være de viktigste i populasjonen som sådan.

Hva har vi så lært?

1. Vi fikk relativt stabile forklaringsmodeller for de viktigste økonomiske parametre, men i stor grad klimatiske/ geografiske faktorer der det er begrenset hva en enkelt oppdretter kan påvirke dette i sin drift.
2. Viktig for næringa at en har data som kan si noe om hele kysten og driftsforhold, og slike resultater kan brukes strategisk av næringa for å bedre driftskvaliteten og justere forventningene til økonomisk resultat ut fra de rådende klimatiske/ geografiske forhold.
3. Slike databaseanalyser er gyldige for hele populasjonen i basen, men er ikke nødvendigvis relevant for å gi råd til den enkelte oppdretter
4. Direkte råd til enkeltoppdrettere har vært basert på detaljert gjennomgang av data fra hvert enkelt anlegg. Slike råd ikke erstattes av avanserte statistiske modeller, men disse kan være et godt supplement.
5. Resultatene illustrerer en komplementær bruk av slike baser, der en kan si noe overordna om hele populasjonen med bruk av de teknikker som er beskrevet her, og ved en annen tilnærming med detaljert gjennomgang av data fra hvert enkelt anlegg hvis råd skal gis til enkeltanlegg.
6. Et tett samarbeid mellom forskere og næring er absolutt nødvendig for å tolke og utvikle MonAqua-basen og lignende databasesystemer.

Videre arbeid

Som nevnt skal deler av databasen analyseres videre via Arnfinn Aunsmos doktorgradsprosjekt og publiseres i internasjonale tidsskrifter. Det skal videre skrives en popularisert artikkel til Norsk Fiskeoppdrett med vekt på å sette resultatene fra resultatene inn i en sammenheng for næringa og foreslå utviklingsmodeller for databaser innen oppdrettsnæringa.