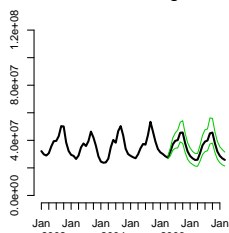
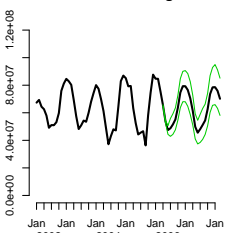


Prediksjon av biomasse for norsk oppdrettslaks – fase 1

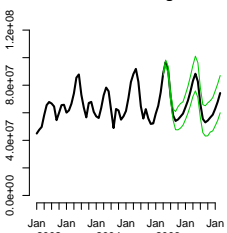
Vektklasse 0 beholdning biomasse



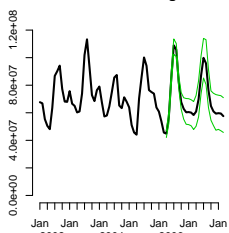
Vektklasse 1 beholdning biomasse



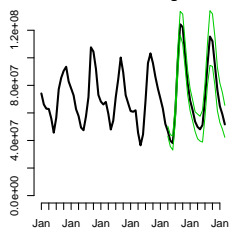
Vektklasse 2 beholdning biomasse



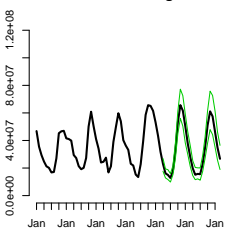
Vektklasse 3 beholdning biomasse



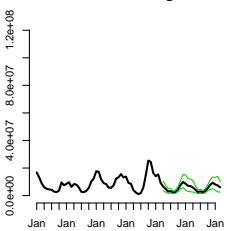
Vektklasse 4 beholdning biomasse



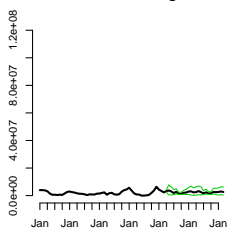
Vektklasse 5 beholdning biomasse



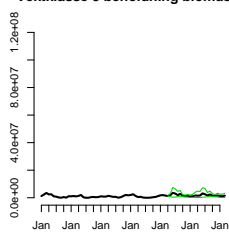
Vektklasse 6 beholdning biomasse



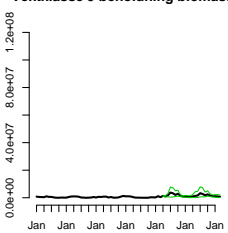
Vektklasse 7 beholdning biomasse



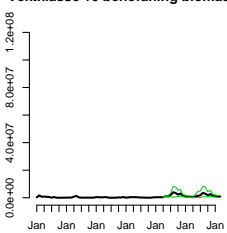
Vektklasse 8 beholdning biomasse



Vektklasse 9 beholdning biomasse



Vektklasse 10 beholdning biomasse



Notatnr
Forfattere

SAMBA/04/07
Anders Løland
Ragnar Bang Huseby
Magne Aldrin

Dato

14. mars 2007

Norsk Regnesentral

Norsk Regnesentral (NR) er en privat, uavhengig stiftelse som utfører oppdragsforskning for bedrifter og det offentlige i det norske og internasjonale markedet. NR ble etablert i 1952 og har kontorer i Informatikkbygningen ved Universitetet i Oslo. NR er et av Europas største miljøer innen anvendt statistikk. Det jobbes med svært mange forskjellige problemstillinger slik som estimering av torskbestandene, finansiell risiko, beskrivelse av geologien i petroleumsreservoarer og overvåking av klimaendringer. NR er også ledende i Norge innen utvalgte deler av informasjons- og kommunikasjonsteknologi. Problemstillinger kan være overvåke inntrengning i datasystemer, e-læring i skole og næringsliv, bruk av datateknologi i markedsanalyser samt anvendelser av multimedia på forskjellige plattformer. NRs visjon er forskningsresultater som brukes og synes.

Tittel **Prediksjon av biomasse for norsk oppdrettslaks – fase 1**

Forfattere **Anders Løland** <anders.loland@nr.no>
Ragnar Bang Huseby <ragnar.huseby@nr.no>
Magne Aldrin <magne.aldrin@nr.no>

Dato 14. mars 2007

Publikasjonsnummer SAMBA/04/07

Sammendrag

Dette notatet oppsummerer arbeidet med prosjektet "Prediction of biomass in Norwegian fish farms – fase 1", finansiert av Fiskeri- og havbruksnæringens forskningsfond.

Notatet tar for seg notasjon, modell for antall og vekst, sjøtemperatur, utsetting, slakt og ukontrollert frafall (død, rømning, tap og annet), samt resultater, det vil si prediksjoner med usikkerhet. Vi diskuterer på hvilke punkter modellen kan forbedres i en eventuell fase 2.

Emneord Prediksjon, laks, biomasse, slakt, død, vekstfaktor

Målgruppe

Tilgjengelighet Åpen

Prosjekt Prognoser for oppdrettslaks

Prosjektnummer 220269

Satsningsområde Miljø og marine ressurser

Antall sider 40

Copyright © 2007 Norsk Regnesentral

Innhold

1	Innledning	9
2	Modellsammendrag	11
3	Resultater	13
3.1	Prediksjoner med usikkerhet	13
3.1.1	Innsetting	13
3.1.2	Stående beholdning	14
3.1.3	Ukontrollert frafall	14
3.1.4	Slakt og utkast	14
3.2	Scenarier	21
4	Konklusjon	23
4.1	Videre arbeid	23
	Referanser	25
A	Notasjon	27
B	Modell – detaljer	29
B.1	Innsetting	29
B.2	Stående biomasse	29
B.2.1	Vekstfunksjon	32
B.3	Ukontrollert frafall: død, rømning, tap og annet	33
B.3.1	Bevaring av forventningen	34
B.4	Slakt og utkast	34
B.4.1	Utkast	35
B.5	Eksterne variable	35
B.5.1	Sjøtemperatur	35
B.5.2	Modell for sjøtemperatur	36
B.5.3	Daglengde	37
B.6	Estimering av parametre	38
B.6.1	Praktiske detaljer i estimeringen	38
B.6.2	Eksempel på modelltilpasning	38
B.6.3	Estimering av modellene for innsett, ukontrollert frafall, slakt og utkast	39
B.7	Prediksjon	39

Figurer

3.1	Predikert antall innsatte fisk med 90% konfidensintervall (i grønt).	13
3.2	Predikert antall fisk i beholdning med 90% konfidensintervall (i grønt). Øverst: Sum over alle vektklasser. Nederst: Per vektklasse.	15
3.3	Predikert gjennomsnittsvekt for fisk i beholdning med 90% konfidensintervall (i grønt).	16
3.4	Predikert stående biomasse av fisk med 90% konfidensintervall (i grønt). Øverst: Sum over alle vektklasser. Nederst: Per vektklasse.	17
3.5	Predikert antall ukontrollert frafall med 90% konfidensintervall (i grønt). Øverst: Sum over alle vektklasser. Nederst: Per vektklasse.	18
3.6	Predikert slaktet biomasse med 90% konfidensintervall (i grønt). Øverst: Sum over alle vektklasser. Nederst: Per vektklasse.	19
3.7	Predikert antall til utkast per vektklasse med 90% konfidensintervall (i grønt).	20
3.8	Predikert antall fisk i beholdning (tykk svart linje) med 90% konfidensintervall (i grønt). Stiplede linjer i svart tilsvare de fem scenariene der innsettet var høyest. Stiplede linjer i rødt tilsvare de fem scenariene der innsettet var lavest.	21
B.1	Fordeling av fisk i en vektklasse.	30
B.2	Sjøtemperatur per fylke.	36
B.3	Daglengde i timer per fylke per måned (1–12).	37
B.4	Modelltilpasning for vektklasse 1.	40

1 Innledning

Dette notatet oppsummerer arbeidet med prosjektet "Prediction of biomass in Norwegian fish farms – fase 1", finansiert av Fiskeri- og havbruksnæringens forskningsfond. Målet med prosjektet (fase 1 og fase 2) er å utvikle et modellrammeverk for prediksjon av biomasse og slakt for norsk oppdrettslaks, mer spesifikt:

1. Undersøke usikkerheten i data.
2. Lage en foreløpig statistisk modell.
3. Lage en avansert statistisk modell. (Ta hensyn til det vi har lært etter 1. og 2., og lage en avansert, mer realistisk modell, estimere parametrene automatisk og med usikkerhet, og produsere prediksjoner.)
4. Validere og kommunisere resultatene. Publisere resultatene i anerkjente tidsskrifter og populærvitenskapelige tidsskrifter.

Fase 1 består av punkt 1. og 2. En preliminær dataanalyse (Løland et al., 2006), gjort på data fra Havbruksdata¹, ligger til grunn for arbeidet. Data fra Havbruksdata er konfidensielle, og derfor er ikke dette arbeidet offentlig tilgjengelig.

Dette notatet omhandler modellering av oppdrettslaksdata fra Havbruksdata (punkt 2.). Vi har fokusert på laks, siden vi har mest data for laks. Notatet tar for seg notasjon, modell for antall og vekst, sjøtemperatur, utsetting, slakt og ukontrollert frafall (død, rømning, tap og annet).

Notatet er organisert som følger: kapittel 2 beskriver modellen kort. Resultater (prediksjoner med usikkerhet) er gitt i kapittel 3. Et sammendrag og diskusjon om videre arbeid er gitt i kapittel 4. Appendiks A gir en oversikt over notasjon og definisjoner. En detaljert gjennomgang av modellen er gitt i appendiks B.

1. <http://www.havbruksdata.no/>

2 Modellsammendrag

I dette kapittelet beskriver vi modellene vi bruker for å beskrive og framskrive (predikere) bestanden av norsk oppdrettslaks, måned for måned. Detaljene i modellen er beskrevet i appendiks B.

Vi har modeller for fire hovedtyper av data:

Innsetting (appendiks B.1). Vi modellerer antallet fisk som blir satt inn som en funksjon av tid på året samt uforklart tilfeldig variasjon.

Stående biomasse (appendiks B.2). Vi modellerer gjennomsnittsvekt og antall i hver vekt-klasse. Vi antar at fisken vokser som en funksjon av gjennomsnittsvekt/vektklasse, sjøtemperatur, daglengde og tid på året. Vekstfunksjonen er mer komplisert for den første vektclassen enn for de andre. Vi lar innsetningsvekten på fisk variere over året. Basert på prinsippet om antallbalanse og gitt gjennomsnittsvekt beregner vi hvor mange fisk som blir i sin vektklasse og hvor mange som vokser over til neste vekt-klasse, samt hvor mye disse veier etter å ha vokst. Dersom fisken har gjennomsnittsvekt nær øvre grense for vektclassen, vil mange vokse seg over til neste vekt-klasse. Dersom fisken derimot har gjennomsnittsvekt nær nedre grense for vektclassen, vil de fleste bli værende igjen i vektclassen, til tross for at de vokser.

Ukontrollert frafall (død, rømning, tap og annet, appendiks B.3). Vi antar at summen av død, rømning, tap og annet kan modelleres som funksjon av vektklasse, tid på året og andel av stående beholdning, samt noe uforklart tilfeldig variasjon.

Slakt og utkast (appendiks B.4). Tilsvarende som for ukontrollert frafall modellerer vi slakt som funksjon av vektklasse, tid på året og andel av stående beholdning i forrige måned. Vi kan dermed ikke slakte flere fisk enn vi har.

Utkast modelleres tilsvarende som slakt, men som andel av slakt, ikke beholdning. Implisitt antar vi da at antall i utkast ikke kan være større enn antall som slaktes, noe som stort sett er rimelig.

Vi modellerer hver av disse spesielt og knytter dem sammen.

I tillegg har vi mer eksterne data, som sjøtemperatur, fôr og daglengde, som er med på å forklare vekst (appendiks B.5). Sjøtemperatur påvirker veksten (høyere temperatur gir vanligvis høyere vekst). Høyere fôrforbruk ser ut til å være koblet med høyere vekst, men her kan det være vel så mye muligheten for høy vekst som gir høyere fôrforbruk. Fôrforbruk har vi derfor valgt å se bort fra i denne omgang. Dagledd påvirker også veksten (lysere dager gir generelt høyere vekst). Dagledd har vi deterministiske formler for.

Modellene beskrevet over inneholder mange parametre som bestemmer vekst og annet. For at modellene skal bli gode, må parametrene bestemmes slik at modellen gjenska-per historien så godt som mulig. Parametrene i modellene estimeres basert på historiske

data, ved å minimere avviket mellom predikerte verdier og faktiske verdier (se appendiks B.6). Vi estimerer parametrene på en måte som skal sørge for at prediksjonene er gode både på kort og lang sikt.

Basert på de estimerte parametrene og dagens bestand, kan framtidig biomasse, slakt, utkast og ukontrollert frafall predikeres ved å kjøre modellene framover (appendiks B.7).

3 Resultater

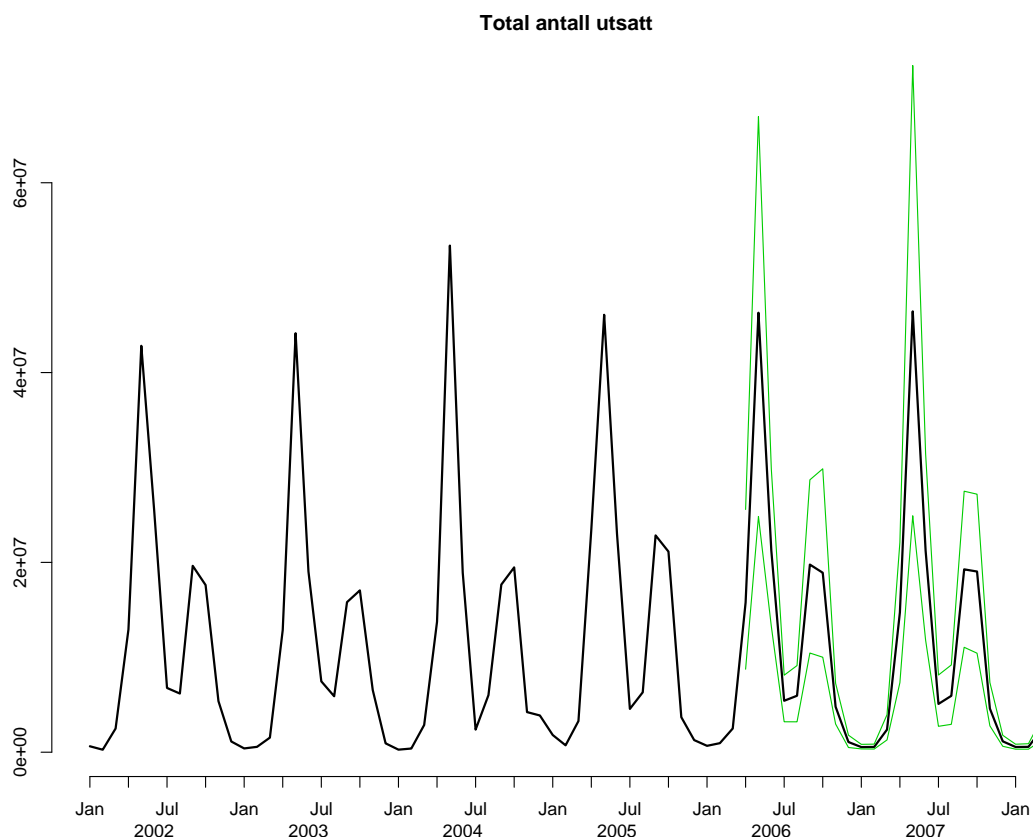
I dette kapitlet viser vi eksempler på prediksjoner, og kommenterer kort hvor gode de er. Vi henviser til appendiks B.6.2 for tilpasning av modellparametre.

3.1 Prediksjoner med usikkerhet

Prediksjonene er basert på hundre stokastiske simuleringer. Det vil si at vi har simulert hundre temperaturscenarier. For hver av disse har vi beregnet vekst og simulert innsetting, frafall og slakt. Alle prediksjonene er basert på at sist kjente observasjon er fra mars 2006 og historiske data fra og med januar 2002 og fram til denne datoen er vist sammen med prediksjonene.

3.1.1 Innsetting

I figur 3.1 ser vi predikert antall innsatte med 90% konfidensintervall. Modellen vi har valgt gjenspeiler det historiske nivået på antall som settes inn. Usikkerheten ser imidlertid ut til å være urealistisk høy.



Figur 3.1. Predikert antall innsatte fisk med 90% konfidensintervall (i grønt).

3.1.2 Stående beholdning

Figur 3.2 viser predikert antall med 90% konfidensintervall, totalt og per vektklasse. Prediksjonene av antall per vektklasse ser rimelig ut, det samme gjør stort sett usikkerheten også, selv om den synes å være noe høy i vektklasse 0 og noe lav fra og med vektklasse 4. At usikkerheten er noe høy i vektklasse 0, kommer av at usikkerheten på innsetting fra figur 3.1 forplanter seg inn i antall i vektklasse 0 (siden det er der de blir satt inn). For vekt (figur 3.3) ser bildet nokså likt ut; prediksjonene ser rimelige ut, men usikkerheten er kanskje noe lav i de nedre vektclassene. Merk at vi for vektklasse 0 ikke får helt med den lille nedgangen i gjennomsnittsvekt fram mot desember. Vi ser også at usikkerheten er relativt stor i de øvre vektclassene (fra og med vektklasse 6), noe som ikke er uventet, siden gjennomsnittsvekten der varierer relativt mye.

Ved å kombinere antall og vekt, får vi stående biomasse som vist i figur 3.4. Stort sett ser det ut til at prediksjonene ligner på historien. Usikkerheten ser heller ikke urimelig ut.

3.1.3 Ukontrollert frafall

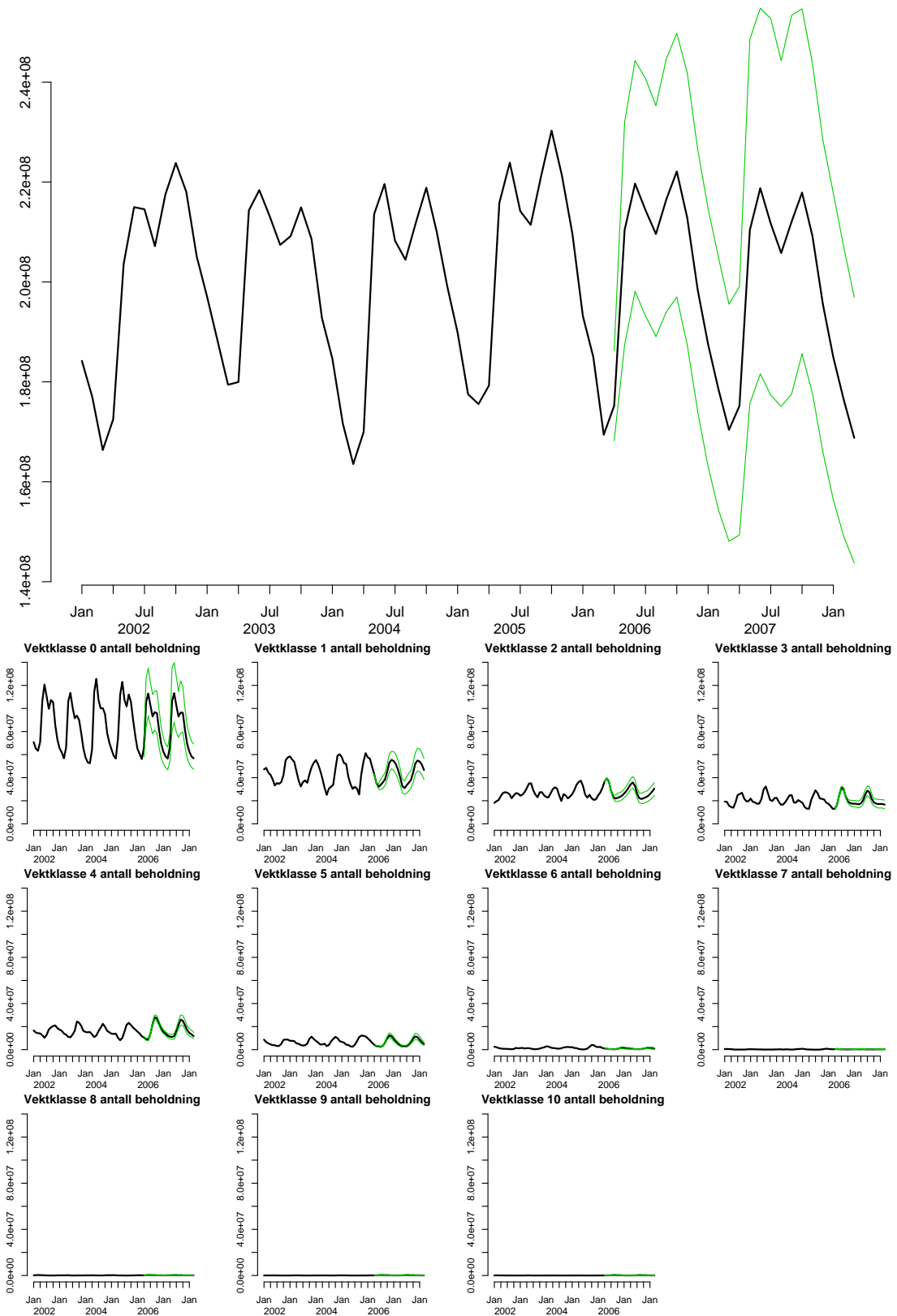
Prediksjoner av ukontrollerte frafallet (død rømning, tap og annet) er vist i figur 3.5. Vi fanger opp det sesongvarierende frafallet og at frafallet er størst i de første vektclassene (særlig vektklasse 0). Usikkerheten ser rimelig ut.

3.1.4 Slakt og utkast

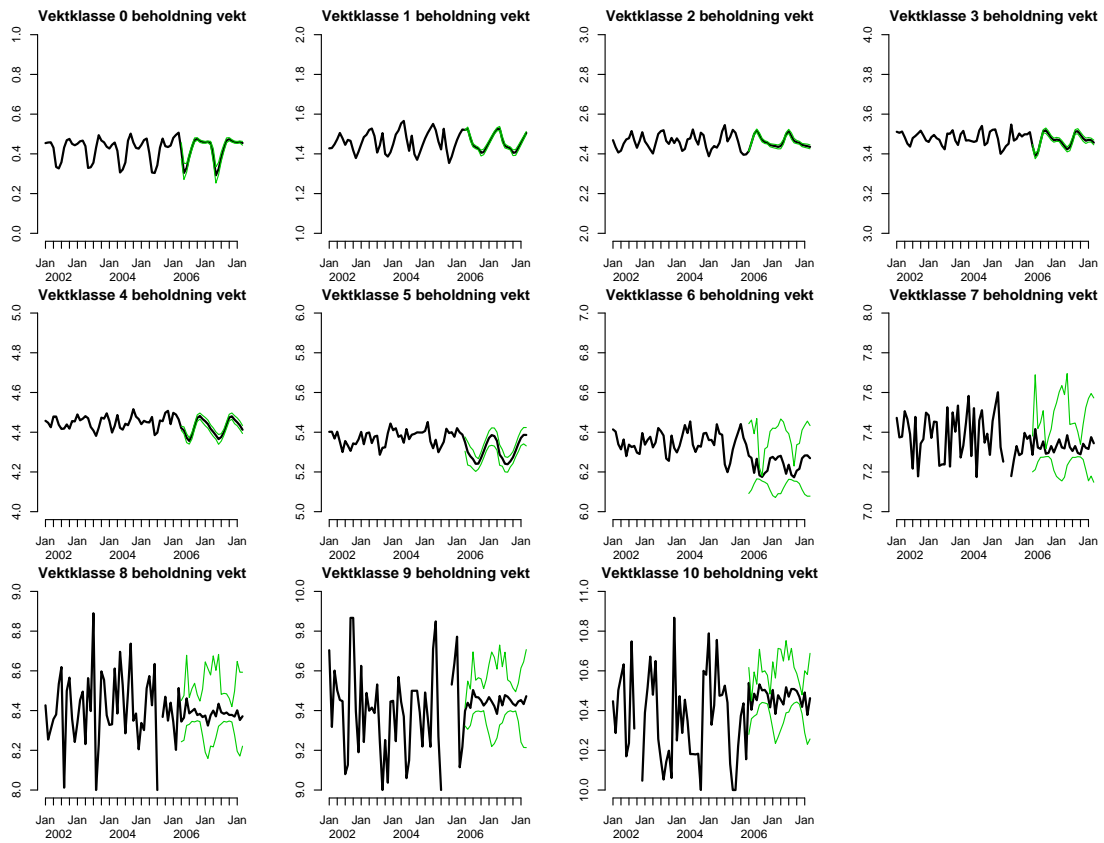
Predikert slaktet biomasse er vist i figur 3.6. Historisk har det vært slaktet mer i 2004/2005 enn i 2002/2003. Vi har modellert slakt som en andel av stående beholdning i forrige måned. Derfor er den predikerte slaktebiomassen noe lavere enn i 2004/2005. Usikkerheten ser ikke helt urimelig ut med tanke på endrede slaktingsstrategier i løpet av årene. I tillegg fanger vi opp at mesteparten av biomassen slaktes i vektklasse 4 og 5.

Antall til utkast er vist i figur 3.7. Selv om vi ikke slakter så mange i vektklasse 0–2 går en stor andel av de som slaktes til utkast.

Totalt antall beholdning

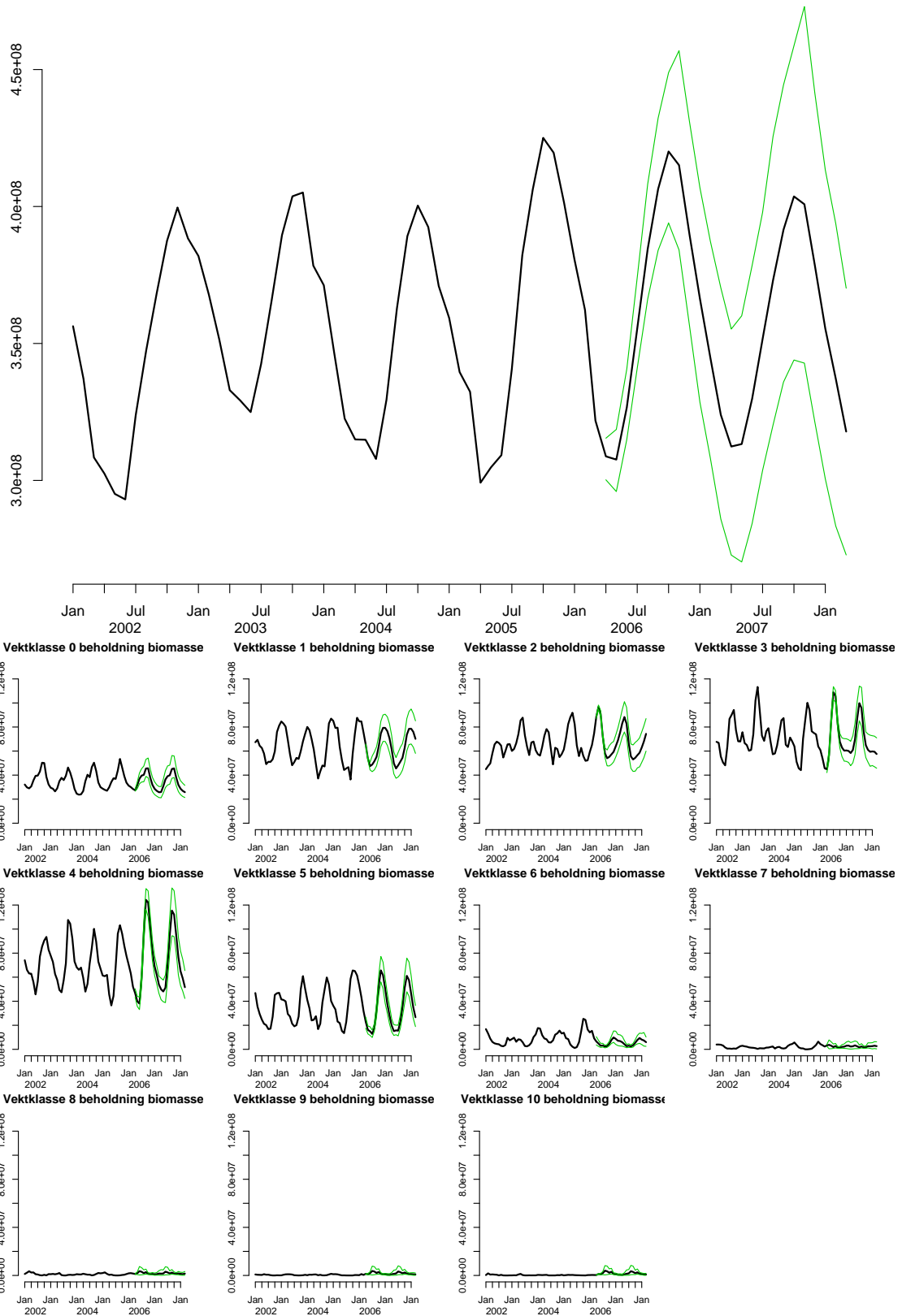


Figur 3.2. Predikert antall fisk i beholdning med 90% konfidensintervall (i grønt). Øverst: Sum over alle vektclasser. Nederst: Per vektklasse.



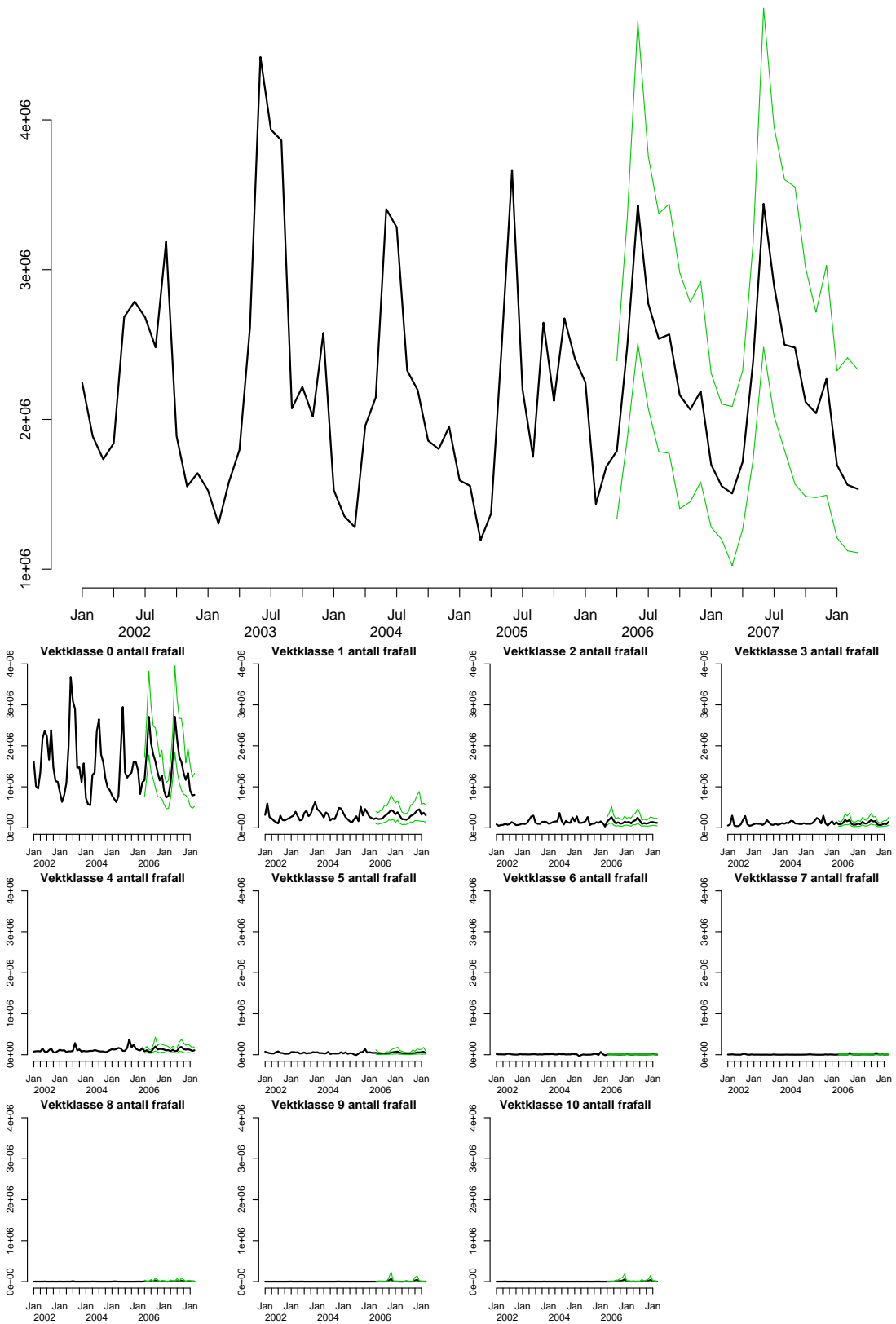
Figur 3.3. Predikert gjennomsnittsvekt for fisk i beholdning med 90% konfidensintervall (i grønt).

Total beholdning biomasse

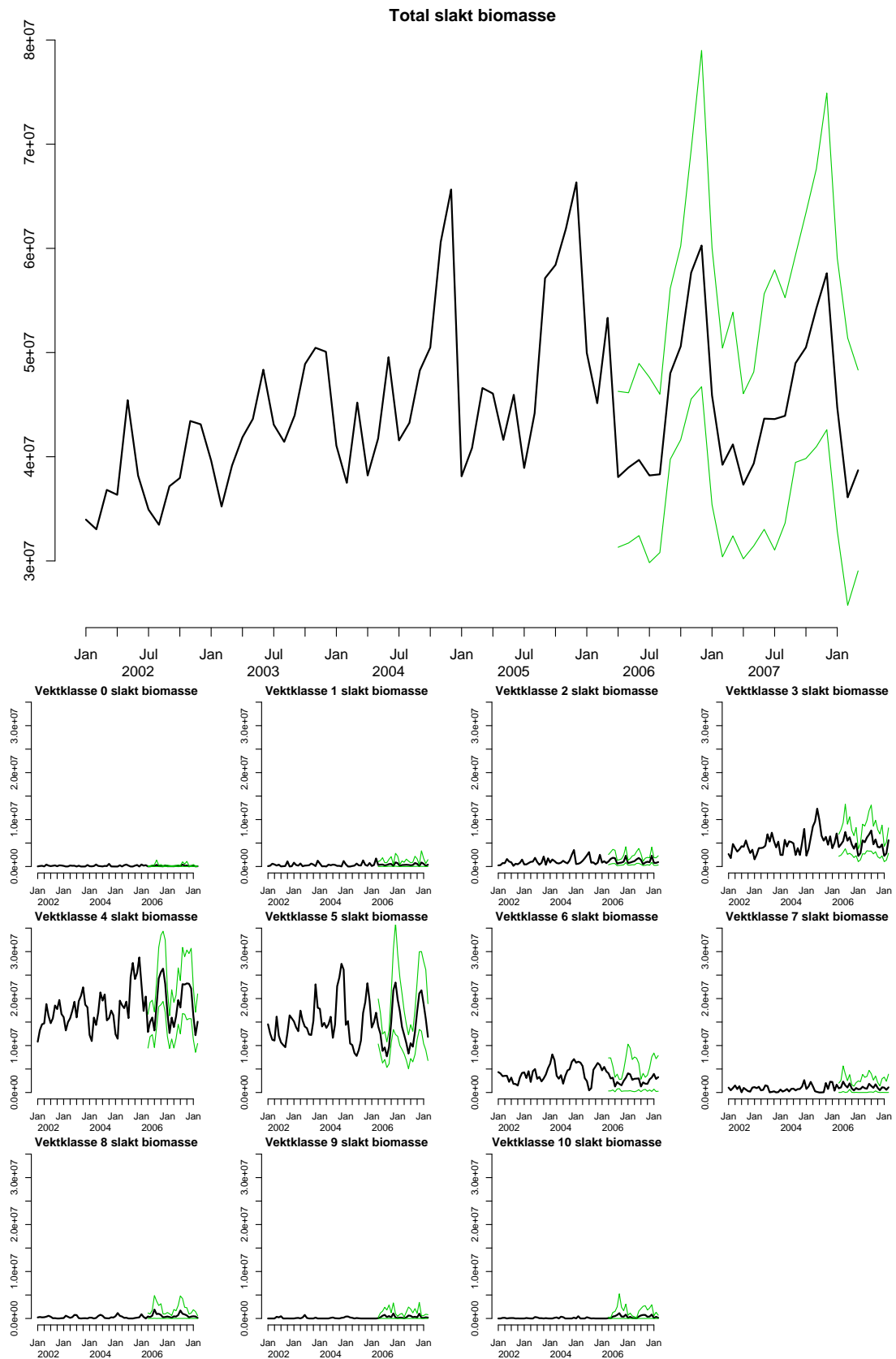


Figur 3.4. Predikert stående biomasse av fisk med 90% konfidensintervall (i grønt). Øverst: Sum over alle vektklasser. Nederst: Per vektklasse.

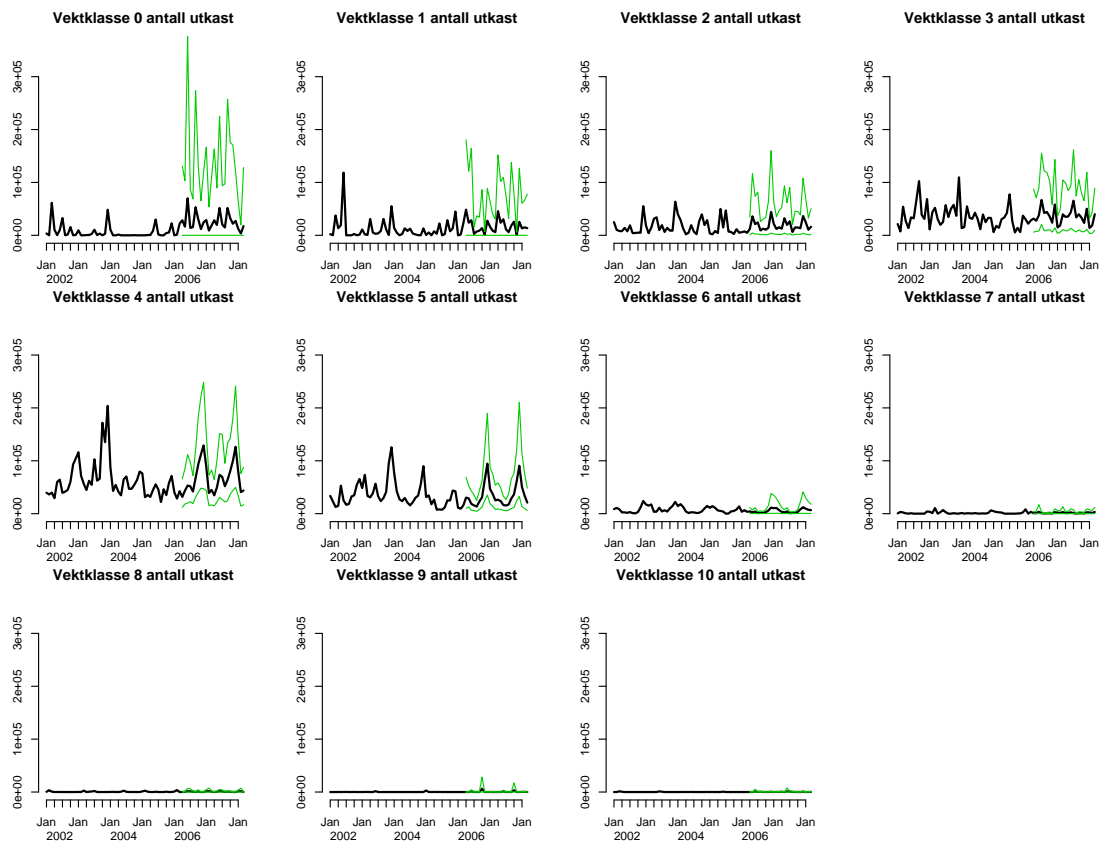
Totalt antall frafall



Figur 3.5. Predikert antall ukontrollert frafall med 90% konfidensintervall (i grønt). Øverst: Sum over alle vektclasser. Nederst: Per vektklasse.



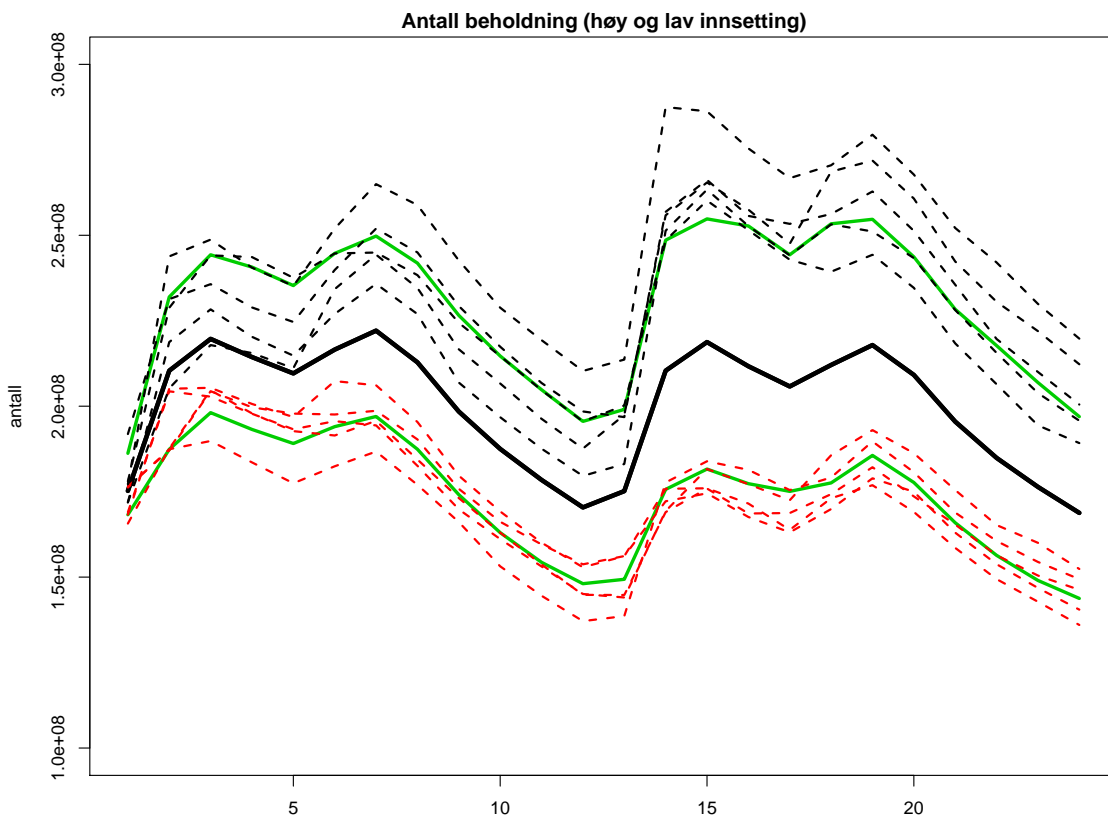
Figur 3.6. Predikert slaktet biomasse med 90% konfidensintervall (i grønt). Øverst: Sum over alle vektclasser. Nederst: Per vektklasse.



Figur 3.7. Predikert antall til utkast per vektklasse med 90% konfidensintervall (i grønt).

3.2 Scenarier

En måte å bruke resultatene på, er å se på enkelte scenarier, for eksempel hva som skjer om vi setter inn mange fisk eller få fisk. Siden prediksjonene våre er basert på stokastiske simuleringer, har vi svært stor frihet til å undersøke slike scenarier. I figur 3.8 har vi sett på scenarier der gjennomsnittlig innsetting er blant de fem høyeste og fem laveste (fortsatt basert på hundre simuleringer). Beholdningen blir dermed relativt lav eller høy. Det er betryggende at prediksjonene reagerer riktig på høyt og lavt innsett.



Figur 3.8. Predikert antall fisk i beholdning (tykk svart linje) med 90% konfidensintervall (i grønt). Stiplede linjer i svart tilsvarer de fem scenariene der innsettet var høyest. Stiplede linjer i rødt tilsvarer de fem scenariene der innsettet var lavest.

Framtidig innsetting kan være vanskelig å predikere, blant annet fordi næringen kan ha bestemt seg for å øke eller minke innsettet, men dette trenger ikke å gjenspeile seg i de historiske dataene. Det kan derfor være kjent a priori at det vil bli satt inn relativt mye framover. Derfor kan det være fornuftig å produsere prediksjoner betinget på scenarier for innsetting. Dessuten kan man teste ut strategier for innsetting og slakt ved å prøve ut ulike scenarier. Merk at dersom vi betinger på ett innsettingsscenario, vil usikkerheten i prediksjonene bli lavere. Vi har da fjernet muligheten for at innsettet kan bli lavere eller høyere enn i innsettingsscenarioet.

4 Konklusjon

Vi har presentert en statistisk modell for å beskrive og framskrive bestanden av norsk oppdrettslaks med usikkerhet. Modellen tar utgangspunkt i hvor mange fisk som befinner seg i en vektklasse og hvor mye de veier. Basert på dette og en vekstfunksjon, beregner modellen hvor mange fisk som blir værende i sin vektklasse, og hvor mange som vokser seg inn i neste vektklasse. I tillegg tar modellen hensyn til innsetting, ukontrollert frafall, slakt og utkast. Modellen har mange parametre. Disse estimerer vi, basert på historiske data, ved å minimere avviket mellom predikerte verdier og faktiske verdier. Resultatene er gode og gir grunn til å arbeide videre med fase 2. Noen arbeidsoppgaver under er ikke nevnt i den opprinnelige søknaden, men er kanskje vel så viktige.

4.1 Videre arbeid

Resultatene vist her er tall for Norge som helhet. Neste skritt er å estimere fylkesvise parametre. I tillegg kan nok modellene, særlig innsettingsmodellen og modellen for slaktevekt, forbedres noe, spesielt når det gjelder beskrivelse av usikkerhet. I resultatkapittelet tok vi startverdiene fra mars 2006 for gitt. Disse er imidlertid estimater med usikkerhet/skjevhet på grunn av manglende rapportering/feilrapportering. I det videre arbeidet vil vi ta hensyn til denne usikkerhetskilden i større grad. Interaksjonen mellom slakt, innsetting og stående beholdning kan også modelleres bedre. Dessuten bør modellene valideres grundigere enn vi har gjort her.

Vi har sett på laks i første omgang. Det gjenstår å teste modellen på flere arter, i første omgang ørret. En muligens større problemstilling da er manglende data: Da bestanden av oppdrettsørret er mindre enn den for laks, vil det trolig for flere kombinasjoner av vektklasse, fylke og måned være avvik fra antallsbalansen. Ufullstendig/feilaktig innrapportering kan isåfall få større konsekvenser for ørret enn for laks.

Et tilleggsaspekt er hvordan resultatene skal presenteres. Som vist i kapittel 3 er det mulig å presentere svært mange resultater. Dersom vi har tilgang til tilsvarende prediksjoner per fylke, kan omfanget bli overveldende stort. Dessuten bør en bruke a priori-informasjon, om for eksempel innsetting, så sant den er tilgjengelig eller relevant. Da kan scenarietankegangen være relevant, også hvis en vil teste ut ulike scenarier for slakt og innsetting. Hva blir prediksjonene gitt at det blir satt inn ekstra mange fisk? Eller hva om det blir uvanlig varmt eller kaldt? Gitt at det slaktes en viss biomasse framover, hva står vi da igjen med av stående biomasse? Slike spørsmål kan besvares ved hjelp av en realistisk modell og stokastisk simulering.

Når modellen er god nok, vil vi publisere metodikken i et anerkjent internasjonalt tidsskrift.

Referanser

Løland, A., Huseby, R. B. og Aldrin, M. (2006). Ny statistisk analyse av Havbruksdata. Technical Report SAMBA/18/06, Norsk Regnesentral.

McCullagh, P. og Nelder, J. A. (1989). *Generalized Linear Models*. Chapman & Hall, second edition.

A Notasjon

Vi benytter følgende notasjon:

- t er tid på månedlig oppløsning. (*) viser at vi observerer data. Alle data er observert
 - per fylke,
 - per årsklasse (det året de ble satt ut),
 - per vektklasse (0-1 kg, 1-2 kg, ..., 10 + kg).

- Følgende notasjon er mengde i kg:

(*) B_t : total biomasse (i levende vekt) ved tid t

(*) F_t : fôrforbruk ved tid t

U_t : utkast ved tid t

T_t : tellefeil ved tid t

R_t : rømt ved tid t

D_t : død ved tid t

A_t : tap av annen årsak t

(*) S_t : slakt (uttak) ved tid t

Er oppgitt i rundt vekt, med mindre annet er oppgitt.

Isåfall antas sammenhengen rund vekt = $0,93 \times$ levende vekt.

I_t : utsatt ved tid t

- Følgende notasjon er antall fisk:

(*) $N_t = N_t^B$: totalt antall fisk ved tid t

(*) N_t^U : antall fisk i utkast ved tid t

(*) N_t^T : antall fisk med tellefeil ved tid t

(*) N_t^R : antall fisk rømt ved tid t

(*) N_t^D : antall fisk døde ved tid t

(*) N_t^A : antall fisk tapt av annen årsak ved tid t

(*) N_t^S : antall fisk slaktet ved tid t

(*) N_t^I : antall fisk utsatt ved tid t

- I tillegg bruker vi følgende notasjon for gjennomsnittsvekt:

(*) $V_t = V_t^B = \frac{B_t}{N_t}$: gjennomsnittsvekt på stående biomasse (i levende vekt) ved tid t

(*) $V_t^S = \frac{S_t}{N_t^S}$: gjennomsnittlig slaktevekt (rund vekt) ved tid t .

- Vi definerer totalt tap ved tid t som summen av tellefeil, rømt, død, annet, slakt og utkast:

$$(*) N_t^L = N_t^T + N_t^R + N_t^D + N_t^A + N_t^S + N_t^U.$$

Utenom slakt og utkast har vi ukontrollert frafall,

$$(*) N_t^X = N_t^T + N_t^R + N_t^D + N_t^A.$$

· I tillegg har vi følgende notasjon for vektklasse v :

- $N_t(v)$: totalt antall fisk ved tid t i vektklasse v
- (*) $N_{t+1}(v, v)$: antall fisk (utenom utsetting, men inkludert tap) ved tid $t + 1$ i vektklasse v som var i vektklasse v ved tidspunkt t
- (*) $N_{t+1}(v + 1, v)$: antall fisk (utenom utsetting, men inkludert tap) ved tid $t + 1$ i vektklasse $v + 1$ som var i vektklasse v ved tidspunkt t
- $V_t(v)$: gjennomsnittsvekt på fisk ved tid t i vektklasse v
- (*) $V_{t+1}(v, v)$: gjennomsnittsvekt på fisk (utenom utsetting, men inkludert tap) ved tid $t + 1$ i vektklasse v som var i vektklasse v ved tidspunkt t
- (*) $V_{t+1}(v + 1, v)$: gjennomsnittsvekt på fisk (utenom utsetting, men inkludert tap) ved tid $t + 1$ i vektklasse $v + 1$ som var i vektklasse v ved tidspunkt t
- (*) $V_t^*(v, v)$: gjennomsnittsvekt på fisk (utenom utsetting, men inkludert tap) ved tid t i vektklasse v for de som fortsatt er i vektklasse v ved tidspunkt $t + 1$
- (*) $V_t^*(v + 1, v)$: gjennomsnittsvekt på fisk (utenom utsetting, men inkludert tap) ved tid t i vektklasse v for de som vokser inn i vektklasse $v + 1$ til tidspunkt $t + 1$

Vektklasse $v = 0$ tilsvarer 0-1 kg, vektklasse $v = 1$ tilsvarer 1-2 kg og så videre opp til vektklasse $v = 10$, som tilsvarer 10 kg og mer.

Stående biomasse er ved registrert ved utgangen av måneden. Slakt/utkast/død/rømt/utsett er i løpet av måneden.

Størrelsene over er teoretiske størrelser. Observerte størrelser merkes som N_t^{obs} . N_t er modellens representasjon av totalt antall, mens N_t^{obs} er observert antall.

Videre har vi

- ST_t : Sjøtemperatur på tidspunkt t . Sjøtemperatur er gjennomsnittlig sjøtemperatur i et fylke i en måned.
- D_t : Daglengden på tidspunkt t .
-

$$h(x; p, q) = g(x), \quad 0 \leq x \leq 1,$$

er tettheten i betafordelingen. p og q er parametrene i fordelingen.

B Modell – detaljer

I dette appendikset beskriver vi modellene vi bruker for å beskrive og framskrive (predikere) bestanden av norsk oppdrettslaks som beskrevet i appendiks 2 i mer detalj.

B.1 Innsetting

Innsetting skjer bare i vektklasse 0, selv om det i dataene ser ut til å være noe innsetting i enkelte høyere vektklasser. Dette dreier seg som regel om flytting, og vi har sett bort fra innsetting i høyere vektklasser.

Vi modellerer innsetting som funksjon av tid på året.

La

$$N_t^I \sim \text{Gamma}(\mu_t, d)$$

angi antallet som blir satt inn. Her er d en variansparameter og μ_t forventningen;

$$\log(\mu_t) = \beta_{\text{måned}} \cdot$$

Typisk vil det være slik at det sette inn mye i enkelte måneder og (nesten) ingen ting i andre.

Gjennomsnittsvekt på innsatt fisk er lik v_0 (definert i ligning (B.7)).

B.2 Stående biomasse

Vi antar at vekt på fiskene i vektklasse v ved tidspunkt t er betafordelt med parametre (p, q) . (p, q) bestemmes fra følgende sammenhenger:

$$\text{Forventning} = \mu = (V_t^{\text{obs}} - v) = \frac{p}{p + q}$$

er observert.

$$\begin{aligned} \text{Varians} &= \frac{pq}{(p + q)^2(p + q + 1)} = \gamma^2(\mu(1 - \mu)) \\ &\Downarrow \\ \gamma^2 &= \frac{1}{p + q + 1}, \end{aligned} \tag{B.1}$$

der γ er en global parameter som estimeres basert på data. Ligning (B.1) sier at standardavviket stiger med forventningen til midt i vektklassen, hvor det når sin maksimale verdi. For vektklasse 0 er γ er en funksjon av tid på året;

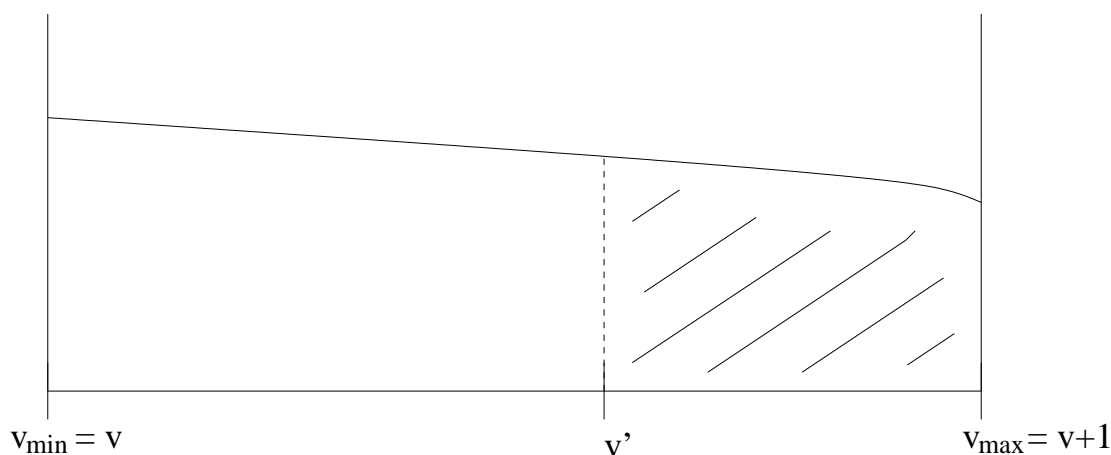
$$\gamma_t^2 = g \left(\beta_\gamma^0 + \beta_\gamma^1 \sin \left(\frac{2\pi t}{12} \right) + \beta_\gamma^2 \cos \left(\frac{2\pi t}{12} \right) \right)$$

der

$$g(x) = \frac{1}{2} + \frac{\arctan(x)}{\pi} \tag{B.2}$$

er en funksjon som ligger mellom 0 og 1.

Vi antar at alle fiskene i samme vektklasse har samme vekstfaktor $f = f_t(V_t(v), \dots)$ fra t til $t + 1$. f defineres i appendiks B.2.1.



Figur B.1. Fordeling av fisk i en vektklasse.

Dermed vil alle fisk tyngre enn v' bytte vektklasse (se figur B.1), der v' er gitt ved $f \cdot v' = v_{\max}$, det vil si

$$v' = \frac{v_{\max}}{f}. \quad (\text{B.3})$$

Hvor mange som bytter vektklasse: Antall fisk som beholder og bytter vektklasse er gitt ved

$$N_t(v) \rightarrow \begin{cases} N_{t+1}(v, v) = aN_t^{\text{obs}}(v) \\ N_{t+1}(v+1, v) = (1-a)N_t^{\text{obs}}(v). \end{cases}$$

a angir andel som ikke hopper over i neste vektklasse. Jo nærmere gjennomsnittsvekten i en vektklasse er lik starten på neste vektklasse, jo større sannsynlighet er det for at flere fisk kommer seg over i neste vektklasse.

a er gitt ved

$$\int_0^{v'-v} h(x)dx = P(X < v' - v),$$

der h er tettheten i betafordelingen og P er den kumulative betafordelingen.

Totalt antall i vektklasse: Dermed blir antallet i vektklasse v i neste måned, $t + 1$, en sum av antallet i vektklasse $v - 1$ i forrige måned som vokste seg inn i vektklasse v og antallet som ble igjen i vektklasse v , minus tap pluss innsetting;

$$N_{t+1}(v) = N_{t+1}(v, v-1) + N_{t+1}(v, v) - N_{t+1}^{L'}(v) + N_{t+1}^I(v). \quad (\text{B.4})$$

Dette blir predikert verdi som sammenliknes med tilsvarende observert verdi ved estimering av parametrene. I ligning (B.4) er¹

$$N_{t+1}^{L'}(v) = \min\{N_{t+1}^L(v), N_{t+1}(v)\}$$

for å være sikret positive prediksjoner (se appendiks B.6.1 for detaljer).

For tap antar vi at fisken i gjennomsnitt tapes i midten av måneden. Den rekker derfor ikke å vokse en hel måned. Imidlertid er det vanskeligere å predikere antall tap enn stående biomasse, siden vi ikke kjenner fordelingen på de som kommer til å gå til tap i neste måned. Siden det bare er cirka halvparten (i forhold til stående biomasse) som skifter vektklasse på en halv måned, er antall til tap dekomponert som

$$\begin{aligned} N_{t+1}^L(v, v-1) &= a^L \cdot N_{t+1}^{L'}(v) \\ N_{t+1}^L(v, v) &= (1 - a^L) \cdot N_{t+1}^{L'}(v), \end{aligned}$$

hvor

$$a^L = \frac{N_{t+1}(v, v-1)}{N_{t+1}(v, v-1) + N_{t+1}(v, v)}.$$

Gjennomsnittsvekt i vektklasse: Siden alle fisk i vektklasse v vokser med samme vekstfaktor f , har vi at

$$V_t(v) \rightarrow \begin{cases} V_{t+1}(v, v) = fV_t^*(v, v) \\ V_{t+1}(v+1, v) = fV_t^*(v+1, v) \end{cases}$$

hvor²

$$\begin{aligned} V_t^*(v, v) &= \frac{\int_0^{v'-v} (x+v)h(x)dx}{\int_0^{v'-v} h(x)dx} \\ &= \frac{\int_0^{v'-v} xh(x)dx}{\int_0^{v'-v} h(x)dx} + v \frac{\int_0^{v'-v} h(x)dx}{\int_0^{v'-v} h(x)dx} \\ &= v + \frac{B(p+1, q)}{B(p, q)} \cdot \frac{P(X < v' - v; p+1, q)}{P(X < v' - v; p, q)}. \end{aligned} \tag{B.5}$$

Tilsvarende vil

$$\begin{aligned} V_t^*(v+1, v) &= \frac{\int_{v'-v}^1 (x+v)h(x)dx}{\int_{v'-v}^1 h(x)dx} \\ &= v + \frac{B(p+1, q)}{B(p, q)} \cdot \frac{1 - P(X < v' - v; p+1, q)}{1 - P(X < v' - v; p, q)}. \end{aligned} \tag{B.6}$$

Innsetningsvekt:

1. Denne trunkeringen har vist seg å ikke være tilstrekkelig i alle tilfeller. Derfor har vi implementert en noe strengere versjon, som faktisk er for streng.

2. Betatettheten kan skrives som

$$h(x) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} x^{p-1}(1-x)^{q-1} = x^{p-1}(1-x)^{q-1}/B(p, q).$$

Derfor vil

$$xh(x; p, q) = x \cdot x^{p-1}(1-x)^{q-1}/B(p, q) = \frac{B(p+1, q)}{B(p, q)} x^p(1-x)^{q-1}/B(p+1, q) = \frac{B(p+1, q)}{B(p, q)} h(x; p+1, q).$$

Vi lar innsetningsvekten variere med sesong:

$$v_0 : \text{vekten p\aa utsatt laks,} \quad (B.7)$$

$$v_0(t) = g \left(\beta_{v_0}^0 + \beta_{v_0}^1 \sin \left(\frac{2\pi t}{12} \right) + \beta_{v_0}^2 \cos \left(\frac{2\pi t}{12} \right) \right).$$

$g(\cdot)$ er definert i ligning (B.2). Vi antar at fisken i gjennomsnitt settes inn i midten av måneden. Den rekker derfor ikke \aa vokse en hel måned. For \aa forenkle modellen, er v_0 lik vekten paa innsatt fisk som har vokst en halv måned.

Total biomasse: Total biomasse $B_{t+1}(v)$ i vektklasse v vil være

$$\begin{aligned} B_{t+1}(v) &= V_{t+1}(v)N_{t+1}(v) \\ &= v_0(t)N_{t+1}^I(v) \\ &\quad - V_{t+1}(v, v-1)N_{t+1}^L(v, v-1) \\ &\quad - V_{t+1}(v, v)N_{t+1}^L(v, v) \\ &\quad + V_{t+1}(v, v-1)N_{t+1}(v, v-1) \\ &\quad + V_{t+1}(v, v)N_{t+1}(v, v) \end{aligned} \quad (B.8)$$

$N_{t+1}^I(v)$ vil typisk være 0 for alle $v > 0$, siden det stort sett ikke settes inn gammel fisk.

Her antar vi at fisken som dør, rømmer og slaktes har samme gjennomsnittsvekt som de andre i vektklasse v . Det er kanskje ikke en urimelig antagelse for de som dør og rømmer, men det stemmer muligens ikke helt for slakt. I tillegg antar vi implisitt at fisk som er satt inn ikke kan bytte vektklasse med en gang de er satt inn (noe som er en rimelig antagelse).

Fra ligning (B.8) får vi at gjennomsnittsvekt $V_{t+1}(v)$ er gitt ved

$$V_{t+1}(v) = \frac{B_{t+1}(v)}{N_{t+1}(v)}.$$

B.2.1 Vekstfunksjon

Den preliminnære dataanalysen (Løland et al., 2006) viste at veksten er avhengig av (minst) fire identifiserbare faktorer; gjennomsnittsvekt, sjøtemperatur, daglengde og tid paa året.

Vekstfunksjonen eller vekstfaktoren beskriver hvordan fisk vokser fra en måned til neste, gitt gjennomsnittsvekt, daglengde og sjøtemperatur. Veksten kan ikke bli negativ, men i teorien kan vekstfunksjonen bli lik 1 (altså ingen vekst). Veksten er avhengig av

gjennomsnittsvekt, sjøtemperatur, daglengde og tid på året. For vektklasser over 0 er

$$\begin{aligned}
 f &= 1 + f' \\
 \log(f') &= \log(f'_t(ST_t, D_t)) \\
 &= \beta_{(v)} \\
 &\quad + \beta_1 \sin\left(\frac{2\pi t}{12}\right) + \beta_2 \cos\left(\frac{2\pi t}{12}\right) \\
 &\quad + \beta_3 ST_t \\
 &\quad + \beta_4 (ST_t - c)^2 \\
 &\quad + \beta_5 D_t \\
 &\quad + \beta_6 (D_t - 12)^2,
 \end{aligned} \tag{B.9}$$

der c er en konstant (\approx årlig gjennomsnittstemperatur) som ikke estimeres, mens parametrene (β -ene) estimeres basert på historiske data (se appendiks B.6). $\beta_{(v)}$ har forskjellig verdi for hver vektklasse. For vektklasse 0 er veksten anderledes (for de som er i vektklasse 0 fra forrige måned, ikke de som er satt inn);

$$\begin{aligned}
 f &= 1 + f' \\
 \log(f') &= \log(f'_t(V_t(0), ST_t, D_t)) \\
 &= \beta_1 \sin\left(\frac{2\pi t}{12}\right) + \beta_2 \cos\left(\frac{2\pi t}{12}\right) \\
 &\quad + \beta_3^0 ST_t \\
 &\quad + \beta_4^0 (ST_t - c)^2 \\
 &\quad + \beta_5 D_t \\
 &\quad + \beta_6 (D_t - 12)^2 \\
 &\quad + \beta_7^0 (V_t(0) - 0,5) \\
 &\quad + \beta_8^0 (V_t(0) - 0,5)^2.
 \end{aligned}$$

Parametre med superskript 0 gjelder bare for vektklasse 0. Forskjellen fra ligning (B.9) er at veksten er avhengig av gjennomsnittsvekt, samt at den temperaturavhengige veksten er anderledes for vektklasse 0 og de øvrige vektklassene.

B.3 Ukontrollert frafall: død, rømning, tap og annet

Vi antar at summen av død, rømning, tap og annet kan modelleres som funksjon av vektklasse, tid på året og andel av stående beholdning.

La

$$a_t^X = \frac{N_t^X}{N_{t-1}^X} \tag{B.10}$$

angi andelen som blir borte (hovedsakelig ved død). I gjennomsnitt er denne andelen noe over 1% per måned.

Vi modellerer en transformasjon av denne andelen ved

$$\frac{a_t^X + \xi}{1 - a_t^X + \xi} = f^X \cdot e^{\varepsilon_t^X}, \tag{B.11}$$

hvor ξ er en konstant (positiv, men nesten lik null for å unngå numeriske problemer). f^X er avhengig av måned, og vi har en modell for hver vektklasse v :

$$\log(f^X(v)) = c_{\text{måned}}^X(v).$$

$c_{\text{måned}}^X(v)$ er parametre. $\varepsilon_t^X(v) \sim \mathcal{N}(0, \sigma_X^2(v))$, så variansen i modellen er forskjellig fra vektklasse til vektklasse. Transformasjonen i ligning (B.11) sørger for at andelen ligger mellom 0 og 1.

Gjennomsnittsvekt på ukontrollert frafall settes lik gjennomsnittsvekt i vektklassen.

B.3.1 Bevaring av forventningen

Når vi bruker modellen for andelen som faller fra, vil vi sørge for at forventningen til de modellerte andelene (B.10) er lik forventningen til de historiske andelene. For å predikere hvor mange laks som slaktes, trekker vi tilfeldige tall fra normalfordelingen $\varepsilon_t^X(v) \sim \mathcal{N}(0, \sigma_X^2(v))$ for hvert framtidige tidspunkt. For å bevare den historiske forventningen krever vi at

$$a^{X,\text{sim}} = a^{X*} \frac{\overline{a^X}}{\overline{a^{X*}}}.$$

Her er a^{X*} simulert verdi fra modellen, $\overline{a^X}$ er gjennomsnittet av historiske andeler og $\overline{a^{X*}}$ er gjennomsnittet over mange simuleringer for de månedene vi har i dataene.

B.4 Slakt og utkast

Tilsvarende som for ukontrollert frafall modellerer vi slakt som funksjon av vektklasse, tid på året og andel av stående beholdning.

La

$$a_t^S = \frac{N_t^S}{N_{t-1}}$$

angi andelen som slaktes.

Vi modellerer en transformasjon av denne andelen ved

$$\frac{a_t^S + \xi}{1 - a_t^S + \xi} = f^S \cdot e^{\varepsilon_t^S}, \quad (\text{B.12})$$

hvor ξ er en konstant (positiv, men nesten lik null for å unngå numeriske problemer). f^S er avhengig av måned:

$$\log(f^S(v)) = c_{\text{måned}}^S(v).$$

$c_{\text{vektklasse}}^S$ og $c_{\text{måned}}^S$ er parametre. $\varepsilon_t^S(v) \sim \mathcal{N}(0, \sigma_S^2(v))$, så variansen i modellen er forskjellig fra vektklasse til vektklasse. Transformasjonen i ligning (B.12) sørger for at andelen ligger mellom 0 og 1.

Vi krever at forventningen til de modellerte andelene er lik forventningen til de historiske andelene, som i appendiks B.3.1.

Slaktevekt (på log-skala) er lik gjennomsnittsvekt i vektclassen det slaktes pluss et avvik og et støyledd,

$$\log V_t^S(v) = \log V_t(v) + \mu(v) + z_t^S,$$

$z_t^S \sim \mathcal{N}(0, \sigma_S^2(v))$. Dette betyr at vi sier at slaktevekt vanligvis er ulik gjennomsnittsvekt i den vektclassen de slaktes i. Dette kan skyldes forskjeller på rund og sløyd vekt, men også at det generelt slaktes lettere eller tyngre fisk en gjennomsnittsvekten i en vektclass. I tillegg kan det være at gjennomsnittsvekten i en vektclass generelt under- eller overestimeres noe.

B.4.1 Utkast

Utkast modelleres tilsvarende som slakt, men som andel av slakt, ikke beholdning. Implisitt antar vi at antall i utkast ikke kan være større enn antall som slaktes. Det virker ikke som en urimelig antagelse, men det kan tenkes å være urimelig i noen tilfeller.

La

$$a_t^U = \frac{N_t^U}{N_t^S}$$

angi andelen utkast.

$$\frac{a_t^U + \xi}{1 - a_t^U + \xi} = f^U \cdot e^{\varepsilon_t^U}.$$

f^U er avhengig av måned:

$$\log(f^U(v)) = c_{\text{måned}}^U(v).$$

Vi krever også at forventningen til de modellerte andelene er lik forventningen til de historiske andelene, som i appendiks B.3.1.

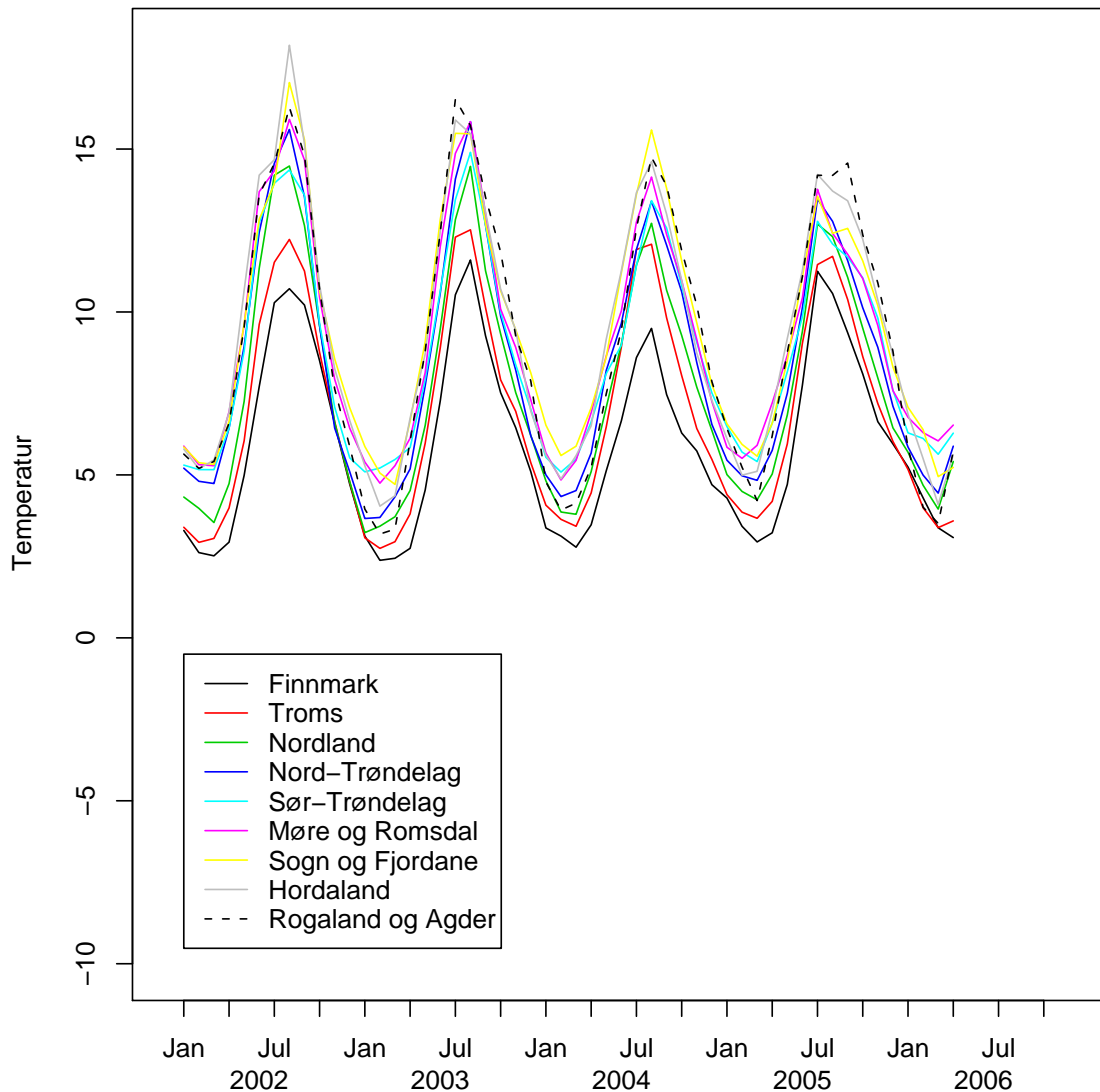
Vi antar at gjennomsnittsvekt på utkast er lik slaktevekt.

B.5 Eksterne variable

B.5.1 Sjøtemperatur

Sjøtemperatur er en viktig forklaringsvariabel til vekst. Vanligvis gir høyere sjøtemperatur høyere vekst.

I Havbruksdata registreres sjøtemperatur per lokasjon per måned. I dataene (vist i figur B.2), er temperaturene et gjennomsnitt over alle lokasjoner i fylket. Dersom det er relativt store temperaturforskjeller innad i et fylke i en måned, kan derfor dette gi lokalt stor vekst, som ikke kan forklares ved hjelp av gjennomsnittstemperaturen alene. Videre ser vi det klare sesongmønsteret, at det generelt er kaldere i vannet lenger nord, og at noen år er varmere eller kaldere enn vanlig. 2005 ser ut til å ha vært et år med relativt små temperaturforskjeller mellom fylkene.



Figur B.2. Sjøtemperatur per fylke.

B.5.2 Modell for sjøtemperatur

Modellen under sier at temperaturen varierer tilfeldig med en sesongtrend, og at temperaturen ikke beveger seg for langt bort fra denne trenden.

$$ST_t = \Lambda_t + \varepsilon_t.$$

Her er

$$\varepsilon_t = \alpha\varepsilon_{t-1} + w_t,$$

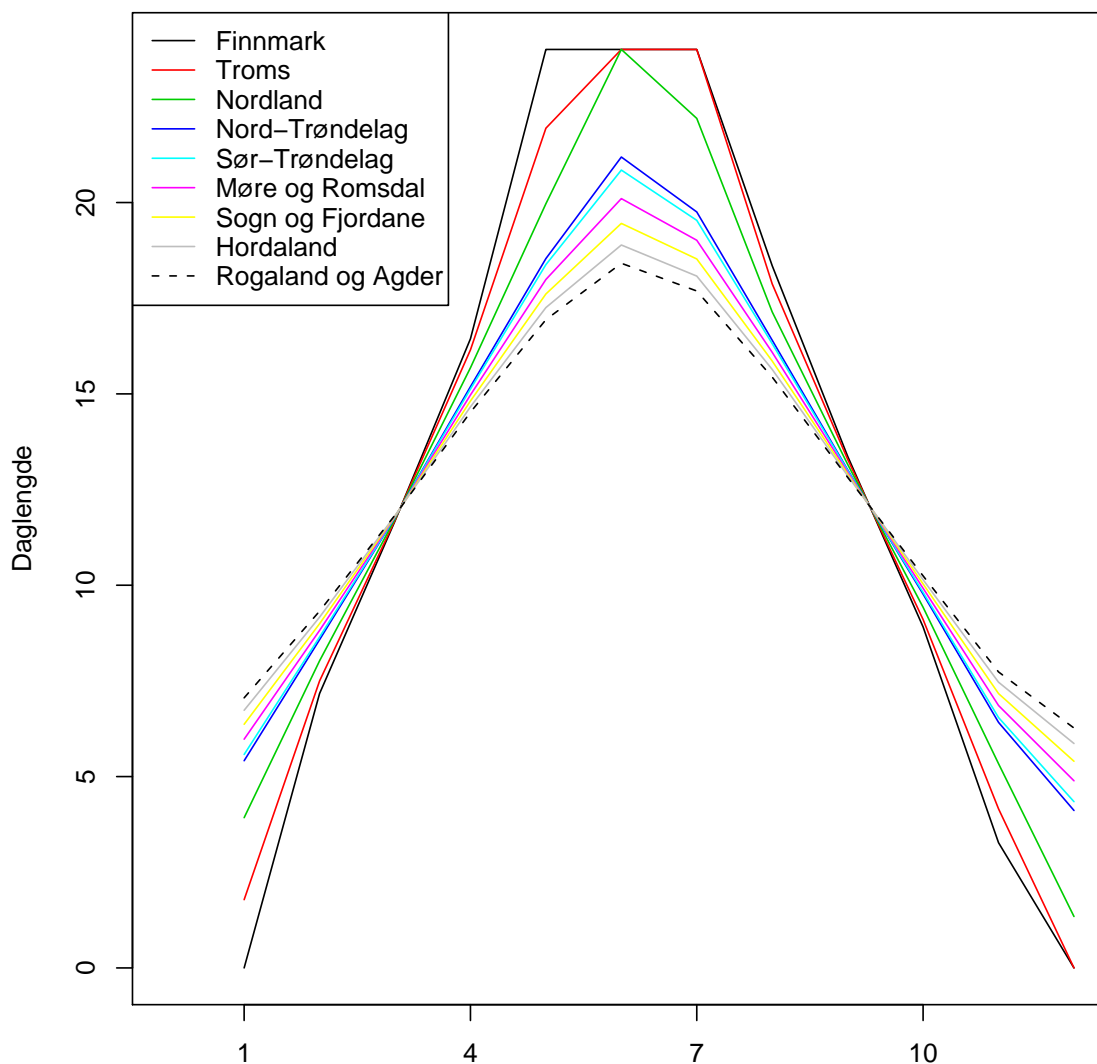
hvor $w_t \sim \mathcal{N}(0, \sigma_{ST}^2)$. Λ_t er en sesongtrend beskrevet ved hjelp av tre parametre:

$$\Lambda_t = \beta^0 + \beta^1 \sin\left(\frac{2\pi t}{12}\right) + \beta^2 \cos\left(\frac{2\pi t}{12}\right).$$

For å simulere sjøtemperatur, starter vi simuleringene i siste historiske temperatur. I resultatkapittelet har vi brukt gjennomsnittlig temperatur over alle fylkene, siden vi ser på data for hele Norge.

B.5.3 Daglengde

Løland et al. (2006) viste at vekst delvis kan forklares av daglengden. I figur B.3 har vi beregnet daglengden³ i timer for et midtpunkt (langs kysten) for hvert fylke. Daglengden er beregnet for en dag midt i måneden. I Finnmark er det for eksempel midnattssol på sommeren, og daglengden er på 24 timer. I resultatene har vi brukt gjennomsnittlig



Figur B.3. Daglengde i timer per fylke per måned (1–12).

daglengde over alle fylkene, sidene vi ser på data for hele Norge.

3. Hentet fra the National Oceanic and Atmospheric Administration, <http://www.srrb.noaa.gov/>.

B.6 Estimering av parametre

Først litt notasjon:

$$N_t = \sum_{v=0}^{10} N_t(v) \quad (\text{totalt antall fisk på tidspunkt } t),$$
$$\bar{N} = \frac{1}{T} \sum_{t=1}^T N_t \quad (\text{gjennomsnittlig totalt antall fisk}).$$

T er antall historiske måneder i datasettet.

Parametrene i modellene estimeres basert på historiske data. Dette gjøres ved å minimere avviket mellom predikerte verdier og faktiske verdier. Avviket er definert i ligning (B.13) og får lavere verdi jo bedre våre prediksjoner av data stemmer med historiske data. Vi sammenligner prediksjonene for neste måned med data for neste måned. Tilsvarende sammenligner vi prediksjonene for to til tolv måneder fram med observerte verdier to måneder til tolv fram. Dette gjør vi for at prediksjonene skal være gode både på kort og lang sikt.

$$\sum_{j=1}^{12} \left(\sqrt{\frac{1}{T \cdot 11} \sum_{t=1}^T \sum_{v=0}^{10} \left(\frac{N_t^{(j)}(v) - N_t^{(j)}(v)^{\text{obs}}}{\bar{N}} \right)^2} \right. \\ \left. + \sqrt{\frac{1}{T \cdot 11} \sum_{t=1}^T \sum_{v=0}^{10} \frac{N_t(v)}{N_t} \left(V_t^{(j)}(v) - V_t^{(j)}(v)^{\text{obs}} \right)^2} \right). \quad (\text{B.13})$$

For avviket mellom predikert antall og observert antall, har vi normalisert med gjennomsnittlig antall over hele perioden, slik at vi ser på relative avvik for antall. For gjennomsnittsvekt veker vi med vektklasseandel av totalt antall per måned, fordi det er viktigst å få riktig gjennomsnittsvekt i de vektklassene som har et høyt antall fisk.

Merk at uttrykket (B.13) er uavhengig av modellen vår. Vi kunne med andre ord foreslått en annen modell, men fortsatt minimert (B.13). Denne måten å estimere på har vist seg å være relativt robust mot modellfeil og numeriske optimeringsproblemer.

B.6.1 Praktiske detaljer i estimeringen

Dersom observert antall i en vektklasse, $N_t(v)^{\text{obs}}$, er veldig lavt, kan vi risikere at prediksjonen for neste måned også blir lav. Dersom vi så trekker fra kjent antall slakt og tap, kan vi ende opp med negative prediksjoner for antall i enkelte (marginale) vektklasser. Dette kan delvis skyldes inkonsistenser i data. Modellen vil nemlig aldri predikere negativt antall fisk. Vi trunkerer antallet, slik at det alltid er positivt.

Vekstfunksjonen tillater i utgangspunktet at fisk kan vokse over mer enn en vektklasse per måned. Vi tillater ikke dette. (Det er heller ikke realistisk at fisk vokser over mer enn en vektklasse på en måned.) Det gjør vi ved å trunkere vekstfunksjonen.

B.6.2 Eksempel på modelltilpasning

I figur B.4 ser vi en grafisk framstilling av modelltilpasningen for vektklasse 1 (1–2 kg). Her har vi antatt at slakt, utkast og ukontrollert frafall er kjent. Generelt ser prediksjonene ut til å være relativt gode. Tolvmånedersprediksjonene for antall ser ut til å være

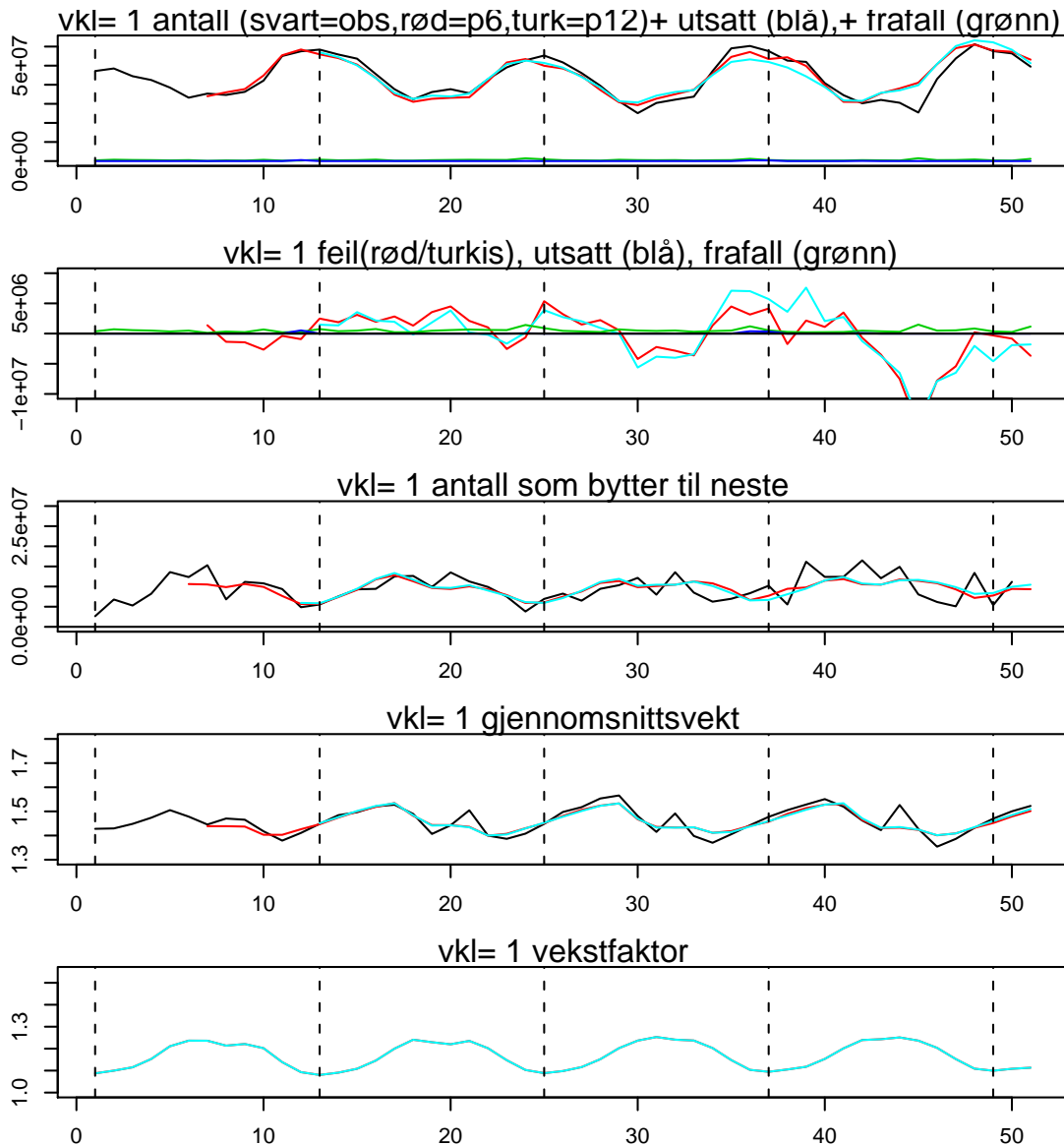
noe dårligere enn seksmånedersprediksjonene. Dette er ventet, siden det er vanskeligere å predikere lengre fram i tid. Vi ser i tillegg en tendens til at prediksjonene er glattere enn historiske data. Også dette er ventet. Her har vi i tillegg det problemet at dataene er beheftet med feil, det vil si at antallsbalansen ikke stemmer helt, for eksempel på grunn av under- eller feilrapportering. Dette kan trolig forklare noen av de relativt store avvikene for antall. Den estimerte veksten (ligning (B.9)) er høyest fra mai til oktober. Vi ser også at veksten varierer noe fra år til år. Det skyldes at temperaturen endrer seg.

B.6.3 Estimering av modellene for innsett, ukontrollert frafall, slakt og utkast

Modellene for innsett, ukontrollert frafall, slakt og utkast estimeres ved hjelp av minste kvadraters-regresjon og regresjon for generaliserte lineære modeller (McCullagh og Nelder, 1989).

B.7 Prediksjon

For å predikere framtidig biomasse, slakt, utkast og ukontrollert frafall, simulerer vi stokastisk fra modellene to år framover, betinget på dagens observasjoner. Det betyr at for hvert ledd i modellen som kommer fra en statistisk fordeling (for eksempel normalfordelingen) trekker vi tilfeldige tall. Gitt disse, kan vi framskrive modellen måned for måned. Prediksjonene er lik gjennomsnittet av simuleringene på hvert framtidige tidspunkt. Usikkerheten kan også finnes fra simuleringene. Et 90% konfidensintervall kan finnes ved å finne kvantilene 5% og 95% fra simuleringene på ethvert framtidig tidspunkt vi ønsker å predikere.



Figur B.4. Modelltilpasning for vektklasse 1. Stiplet linje angir årsskifter. Øverst: Historisk antall, samt predikert antall seks (rød) og tolv (turkis) måneder fram. Nest øverst: Observert minus predikert antall (rød/turkis). I midten: Historisk antall og våre prediksjoner av antall som bytter til neste vektklasse. Merk at vi ikke kjenner eksakt historisk antall som bytter til neste vektklasse, men vi kan anslå det fra historiske data. Nest nederst: Historisk gjennomsnittsvekt, samt predikert gjennomsnittsvekt seks (rød) og tolv (turkis) måneder fram. Nederst: Estimert vekstfaktor per måned for denne vektlassen.