Examensarbete inom teknik

Grundnivå, 15 hp

Klustringsanalys av strömmingbestånd i Östersjön

ELIAS NYGREN

Stockholm, Sverige 2024

Sammanfattning

Målet med denna studie var att identifiera områden i Östersjön vars strömming-populationer har avvikande längd och/eller vikt. Sådana områden är av intresse för bevarandeinsatser samt för att med vidare studier av områdena få bättre förståelse för vilka miljöer som passar för strömming.

Undersökningen använder data insamlad av Baltic International Acoustic Survey (BIAS) mellan 2007 och 2023. Datapunkter delades in i geografiska områden som sedan klustrades utifrån normaliserad längd och/eller vikt med hjälp av en spektral klustringsmetod. Resultatet av klustringen undersöktes för att identifiera mönster och områden som bryter mönster som därav är av intresse.

Tre avvikande områden identifierades: den västra delen av Bottniska viken, området öster om Skåne och området vid Litauens västkust. Fiskar i de första två av dessa områden hade fiskar med relativt hög längd och låg vikt jämfört med närliggande områden. Vid Litauens västkust hittades fiskar med hög vikt, men ingen avvikande längd gick att se.

Tack

Jag vill tacka David Gilljam på SLU Aqua för intressanta insikter strömmingens roll i ekosystemet i Östersjön, information om hur datan är insamlad och hjälp med att bedöma resultatens rimlighet.

Jag vill även tacka min handledare Elias Jarlebring för hjälp med den matematiska delen av projektet.

Innehåll

1 Introduktion **1**
1.1 Bakgrund 1
1.2 Mål och syfte 1

2 Metod **2**
2.1 Datakällor 2
2.1.1 Fiskbestånd 2
2.1.2 Karta 5
2.2 Normalisering 5
2.2.1 Längd 5
2.2.2 Vikt 7
2.2.3 Position 9
2.3 Klustring 10
2.3.1 Avståndsfunktion 10
2.3.2 k-means 11
2.3.3 Spektral klustring 12
2.3.4 Rekursiv klustring 14

3 Resultat **15**
3.1 Normaliseringskurvor 15
3.2 Klustring 16
3.2.1 Klustring efter längd 16
3.2.2 Klustring efter vikt 17
3.2.3 Klustring efter längd och vikt 19

4 Diskussion **20**
4.1 Klustring efter längd 20
4.2 Klustring efter vikt 20
4.3 Klustring efter längd och vikt 20
4.4 Årlig variation i data 21

5 Slutsats **23**

6 Referenser **24**

1 Introduktion

1.1 Bakgrund

Strömming (Clupea harengus) är en viktig fiskart i Östersjön, då den är en betydande matkälla för såval människor som andra arter i Östersjön [1]. Dessa andra arter inkluderar säl, skarv, gädda, torsk och andra rovfiskar. Speciellt torsk är viktig, då torskpopulationen i Östersjön gått nedåt under senaste decennierna. God tillgång till sin huvudsakliga födkälla är en förutsättning för att mängden torsk ska kunna återgå till tidigare nivåer.

Skyddet av strömming är alltså väsentligt både för att bevara ekosystemet i Östersjön och för att möjliggöra fortsatt fiske. Dock har mängden strömming i Östersjön minskat kraftigt på grund av överfiske och klimatförändring. Dessutom har fiskarnas längd och vikt minskat över tid, vilket tyder på försämrade förhållanden för arten [7][6]. Strömming är därför en viktig art för bevarandeinsatser, som behöver kunskap om variationer inom arten för att avgöra hur resurser bör prioriteras. I vissa delar av Östersjön har bevarandeinsatser redan bidragit till att strömmingpopulationerna börjat återhämta sig, men i större delen av Östersjön går trenden fortfarande åt fel håll trots decennier av ansträngningar [9].

Det finns i nuläget väldigt få studier som undersöker småskaliga geografiska skillnader i Östersjön. Många studier undersöker långtida förändringar och behandlar då Östersjön som helhet eller indelad i ett fåtal stora områdan, vilket inte ger möjlighet att identifiera mindre områden med avvikande populationer. På grund av detta finns stora brister i kunskap om vilka miljöer som främjar strömming och migrationsmönster för arten, information som är viktig för att identifiera och motverka de förändringar som orsakat nedgången för strömming i Östersjön.

1.2 Mål och syfte

Målet med denna studie är att undersöka skillnader i längd vid ålder och vikt vid längd för strömming i olika delar av Östersjön. Speciellt sökes områden med avvikande värden, då dessa populationer är av intresse för vidare studier där orsaker till skillnaderna identifieras. Populationer med avvikande värden som beror på genetiska skillnader är av intresse för eventuella bevarandeinsatser. Om avvikande värden istället beror på den lokala miljön är informationen fortfarande av vikt, då den bidrar med bättre förståelse för vilka miljöer som passar bra för strömming. Sådan information är också betydande för bevarandeinsatser och hjälper att förutsäga vilken påverkan klimatförändring har på strömming.

Avvikande områden kan även vara av vikt för fiske, då fetthalt (som påverkar vikt) är relevant för fisk som ska ätas. Dock är naturligtvis antalet fiskar också av stor vikt för fiske, något som denna studie inte undersöker.

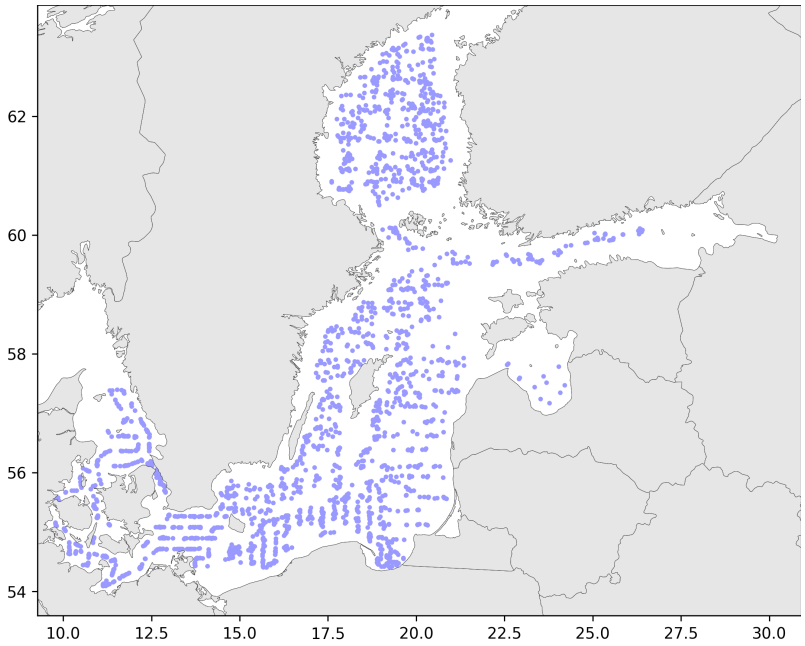
2 Metod

2.1 Datakällor

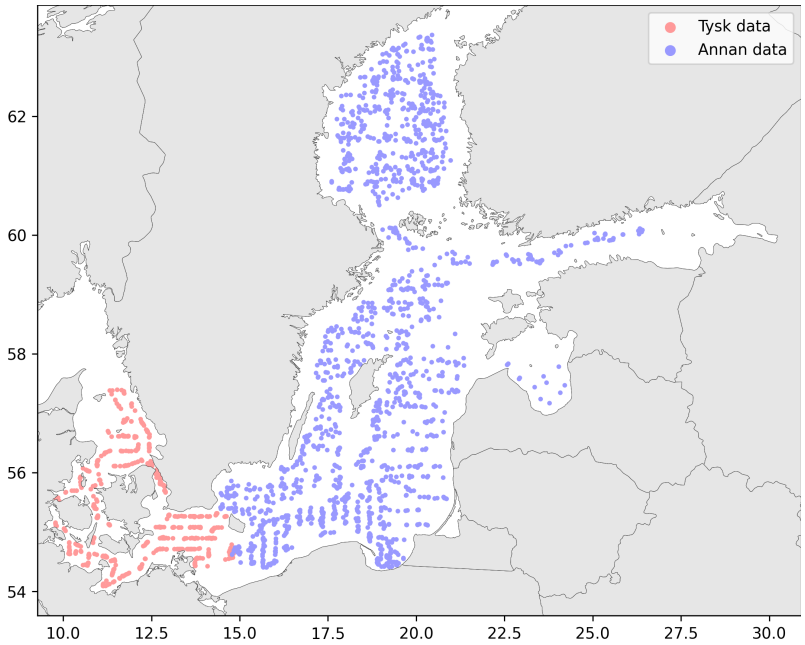
2.1.1 Fiskbestånd

Data om fiskbestånd är insamlad av Baltic International Acoustic Survey (BIAS) och publicerad av International Council for the Exploration of the Sea (ICES). All data är insamlad i september och oktober årtalen 2007 till och med 2023 [10]. Denna data innehåller information om längd, vikt, ålder och position för strömming fångad i Östersjön. Området som täcks går att se i figur 1 nedan. Denna figur visar position för varje tråldrag, där varje tråldrag innehåller data för ett flertal fiskar. BIAS-data är insamlad av Finland, Lettland, Litauen, Polen, Ryssland, Sverige och Tyskland. Dock exkluderas data insamlad av Tyskland (se figur 2) då närheten till Nordsjön leder till högre salthalt i det området, vilket har en betydande påverkan på fiskarnas storlek. Då studien fokuserar på skillnader inom Östersjön måste därför denna data exkluderas. Detta ger en mängd data vars geografiska spridning går att se i figur 3.

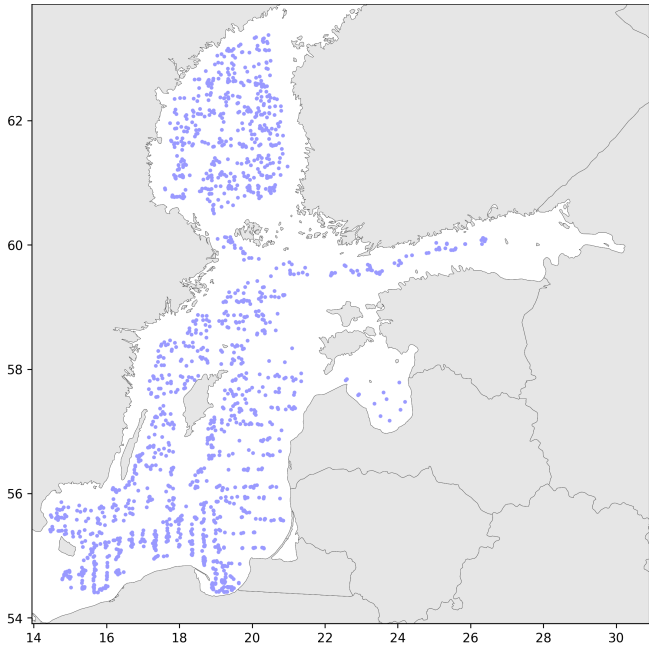
Notera att tätheten av datapunkter varierar inom området. För att undvika att detta påverkar klustringen delas datan in i rektangulära zoner, där datapunkterna i varje rektangel ersätts av en datapunkt i mitten av rektangeln. Längd, vikt och ålder för denna datapunkt ges av medelvärdet av den normaliserade datan för alla fiskar inom rektangeln. Denna indelning går att se i figur 4.



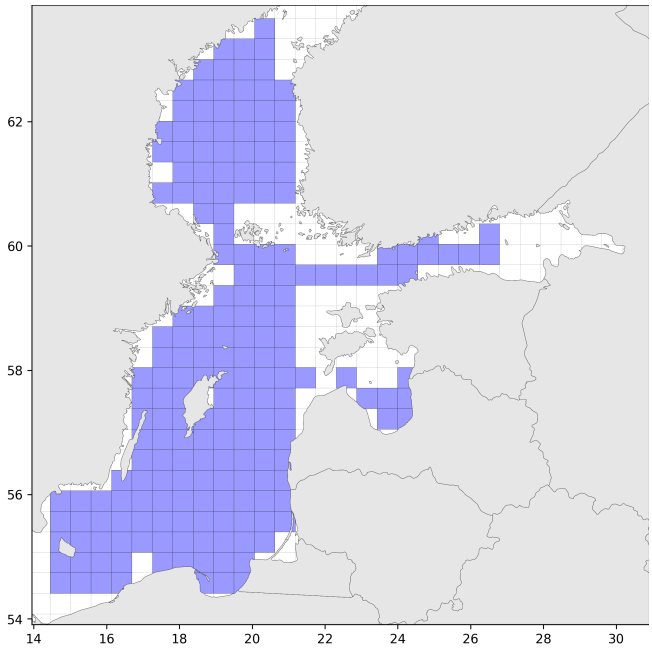
Figur 1: Område täckt av BIAS-data



Figur 2: Område täckt av BIAS-data insamlad av Tyskland



Figur 3: Område täckt av BIAS-data, utan data från Tyskland



Figur 4: Indelning av BIAS-data i rektangulära områden

2.1.2 Karta

Den geografiska data som används för att rita kartor i figurer är hämtad från Natural Earth [11]. Denna data används endast för figurer och påverkar därmed inte klustringen.

2.2 Normalisering

2.2.1 Längd

Längd på en fisk beror kraftigt på fiskens ålder, vilket innebär att områden med avvikande ålder jämfört medel också förväntas ha avvikande medellängd. Eftersom studien önskar hitta mönster i längd, inte i ålder, behöver fiskens längd mätas relativt den förväntade längden för fiskens ålder. För att beräkna den förväntade längden vid en given ålder anpassas en tillväxtkurva till alla datapunkter. Denna tillväxtkurva, kallad von Bertalanffy Growth Curve eller VBGF [4], har formen

$$L(t) = L_\infty \left(1 - e^{-K(t-t_0)}\right) \tag{2.2.1.1}$$

där $L(t)$ är medellängd för en fisk med ålder t. L_∞, K och t_0 är parametrar. För att bestämma dessa parametrar minimeras medelkvadratfelet för datan, alltså med givna längder $L_1, \ldots, L_n$ och åldrar $t_1, \ldots, t_n$ så minimeras

$$E_L = \sum_{i=1}^{n} (L_i - L(t_i))^2. \tag{2.2.1.2}$$

I den använda datan är ålder avrundad till hela år, vilket innebär att många datapunkter har samma ålder. Om datapunkterna grupperas i grupper $1, \ldots, m$ med åldrar $t_1, \ldots, t_m$ och $n_1, \ldots, n_m$ längder i varje grupp så kan summan skrivas om enligt

$$\begin{aligned} E_L &= \sum_{i=1}^{n} (L_i - L(t_i))^2 \\ &= \sum_{i=1}^{m} \sum_{j=1}^{n_i} (L_{ij} - L(t_i))^2 \\ &= \sum_{i=1}^{m} \sum_{j=1}^{n_i} L_{ij}^2 - 2L_{ij}L(t_i) + L(t_i)^2 \\ &= \sum_{i=1}^{m} \left(n_i L(t_i)^2 + \sum_{j=1}^{n_i} L_{ij}^2 - 2L_{ij}L(t_i) \right). \end{aligned} \tag{2.2.1.3}$$

Låt $\bar{L_i}$ vara medellängden i grupp i och låt

$$\frac{1}{n_i} \sum_{j=1}^{n_i} L_{ij}^2 = \hat{L}_i. \tag{2.2.1.4}$$

Då är

$$
\begin{aligned}
E_L &= \sum_{i=1}^{m} \left(n_i L(t_i)^2 + \sum_{j=1}^{n_i} L_{ij}^2 - 2L_{ij}L(t_i) \right) \\
&= \sum_{i=1}^{m} \left(n_i L(t_i)^2 - 2n_i \bar{L}_i L(t_i) + n_i \hat{L}_i \right) \qquad (2.2.1.5) \\
&= \sum_{i=1}^{m} n_i \left(L(t_i)^2 - 2\bar{L}_i L(t_i) + \hat{L}_i \right).
\end{aligned}
$$

Låt

$$r_i = \sqrt{n_i \left(L(t_i)^2 - 2\bar{L}_i L(t_i) + \hat{L}_i \right)}. \qquad (2.2.1.6)$$

Då ska alltså

$$E_L = \sum_{i=1}^{m} r_i^2 \qquad (2.2.1.7)$$

minimeras. Detta görs med Gauss-Newtons metod [2], som utifrån en startgissning

$$\boldsymbol{\beta}^{(0)} = [L_\infty^{(0)} \ K^{(0)} \ a_0^{(0)}]^T = [1 \ 1 \ 1]^T \qquad (2.2.1.8)$$

gör iterationen

$$\boldsymbol{\beta}^{(s+1)} = \boldsymbol{\beta}^{(s)} - \delta(\boldsymbol{J}^T\boldsymbol{J})^{-1}\boldsymbol{J}^T\boldsymbol{r}(\boldsymbol{\beta}^{(s)}) \qquad (2.2.1.9)$$

där δ är en konstant, $\boldsymbol{r} = [r_1 \dots r_m]^T$ och Jacobimatrisen $\boldsymbol{J}$ har element J_{ij} enligt

$$J_{ij} = \frac{\partial r_i}{\partial \beta_j}. \qquad (2.2.1.10)$$

Låt

$$p = e^{-K(t-t_0)} \qquad (2.2.1.11)$$

så att

$$L(t) = L_\infty \left(1 - e^{-K(t-t_0)} \right) = L_\infty(1-p). \qquad (2.2.1.12)$$

Då är

$$r_i^2 = n_i \left(L_\infty^2 (1-p)^2 - 2\bar{L}_i L_\infty (1-p) + \hat{L}_i \right). \qquad (2.2.1.13)$$

Derivatorna i Jacobimatrisen ges då av

$$\frac{\partial r_i}{\partial \beta_1} = \frac{\partial r_i}{\partial L_\infty} = \frac{1}{2r_i}\frac{\partial r_i^2}{\partial L_\infty} = \frac{n_i}{r_i}\left(L_\infty(1-p)^2 - \bar{L}_i(1-p) \right), \qquad (2.2.1.14)$$

$$\frac{\partial r_i}{\partial \beta_2} = \frac{\partial r_i}{\partial K} = \frac{1}{2r_i}\frac{\partial r_i^2}{\partial K} = \frac{n_i(t_i - t_0)p}{r_i}\left(L_\infty^2(1-p) - \bar{L}_i L_\infty \right), \qquad (2.2.1.15)$$

$$\frac{\partial r_i}{\partial \beta_3} = \frac{\partial r_i}{\partial t_0} = \frac{1}{2r_i}\frac{\partial r_i^2}{\partial t_0} = \frac{n_i K p}{r_i}\left(L_\infty^2(p-1) + \bar{L}_i L_\infty \right). \qquad (2.2.1.16)$$

Termen $(\boldsymbol{J}^T\boldsymbol{J})^{-1}\boldsymbol{J}^T\boldsymbol{r}(\boldsymbol{\beta}^{(s)})$ i Gauss-Newton-iterationen skrivs om för ökad numerisk stabilitet. Detta görs genom att QR-faktorisera $\boldsymbol{J}$:

$$\boldsymbol{J} = \boldsymbol{Q}\boldsymbol{R} = [\boldsymbol{Q}_1 \ \boldsymbol{Q}_2] \begin{bmatrix} \boldsymbol{R}_1 \\ \boldsymbol{0} \end{bmatrix} = \boldsymbol{Q}_1\boldsymbol{R}_1 \tag{2.2.1.17}$$

där $\boldsymbol{Q}_1$ har ortogonala kolonner och $\boldsymbol{R}_1$ är övertriangular. Denna faktorisering ger att

$$\begin{aligned}
(\boldsymbol{J}^T\boldsymbol{J})^{-1}\boldsymbol{J}^T\boldsymbol{r}(\boldsymbol{\beta}^{(s)}) &= ((\boldsymbol{Q}_1\boldsymbol{R}_1)^T(\boldsymbol{Q}_1\boldsymbol{R}_1))^{-1}(\boldsymbol{Q}_1\boldsymbol{R}_1)^T\boldsymbol{r}(\boldsymbol{\beta}^{(s)}) \\
&= (\boldsymbol{R}_1^T\boldsymbol{Q}_1^T\boldsymbol{Q}_1\boldsymbol{R}_1)^{-1}\boldsymbol{R}_1^T\boldsymbol{Q}_1^T\boldsymbol{r}(\boldsymbol{\beta}^{(s)}) \\
&= (\boldsymbol{R}_1^T\boldsymbol{R}_1)^{-1}\boldsymbol{R}_1^T\boldsymbol{Q}_1^T\boldsymbol{r}(\boldsymbol{\beta}^{(s)}) \\
&= \boldsymbol{R}_1^{-1}\boldsymbol{R}_1^{-T}\boldsymbol{R}_1^T\boldsymbol{Q}_1^T\boldsymbol{r}(\boldsymbol{\beta}^{(s)}) \\
&= \boldsymbol{R}_1^{-1}\boldsymbol{Q}_1^T\boldsymbol{r}(\boldsymbol{\beta}^{(s)})
\end{aligned} \tag{2.2.1.18}$$

där det nyttjats att $\boldsymbol{Q}_1^T\boldsymbol{Q}_1 = \boldsymbol{I}$. Därmed görs iterationen

$$\boldsymbol{\beta}^{(s+1)} = \boldsymbol{\beta}^{(s)} - \delta\boldsymbol{x} \tag{2.2.1.19}$$

där $\boldsymbol{x}$ ges av

$$\boldsymbol{R}_1\boldsymbol{x} - \boldsymbol{Q}_1^T\boldsymbol{r}(\boldsymbol{\beta}^{(s)}) = \boldsymbol{0}. \tag{2.2.1.20}$$

Gauss-Newton-iterationen avslutas när

$$||\boldsymbol{x}|| < 10^{-10}. \tag{2.2.1.21}$$

Detta ger en funktion $L(t)$ för medellängd vid en ålder t, vilket gör det möjligt att normalisera längd till en relativ avvikelse från medel enligt

$$l_i = \frac{L_i - L(t_i)}{L(t_i)} \tag{2.2.1.22}$$

där l_i är den normaliserade längden för fisk i.

2.2.2 Vikt

Vikten för en fisk beror på fiskens längd. För att undvika att områden med avvikande längd hittas när områden med avvikande vikt sökes måste därför vikten normaliseras med avseende på längd. Likt längden görs detta genom att hitta en funktion som anger medelvikt vid en viss längd och sedan ersätta fiskens vikt med relativ avvikelse från funktionen. Funktionen för vikt V har formen

$$V(L) = aL^b \tag{2.2.2.1}$$

där L är fiskens längd och parametrarna a och b bestäms sådant att kvadratfelet

$$E_V = \sum_{i=1}^{n}(V_i - V(L_i))^2 \tag{2.2.2.2}$$

minimeras, där V_i och L_i är onormaliserad vikt och längd för fisk i. På grund av avrundning har många punkter samma längd. Därför delas datapunkterna in i grupper $1, \ldots, m$ med längder $L_1, \ldots, L_m$ och $n_1, \ldots, n_m$ fiskar i varje grupp. För varje grupp definieras också

$$\bar{V}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} V_{ij} \tag{2.2.2.3}$$

och

$$\hat{V}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} V_{ij}^2. \tag{2.2.2.4}$$

Med denna indelning kan kvadratfelet skrivas om enligt

$$\begin{aligned} E_V &= \sum_{i=1}^{m} \sum_{j=1}^{n_i} (V_{ij} - V(L_i))^2 \\ &= \sum_{i=1}^{m} \sum_{j=1}^{n_i} (V_{ij}^2 - 2V_{ij}V(L_i) + V(L_i)^2) \\ &= \sum_{i=1}^{m} n_i \hat{V}_i - 2n_i \bar{V}_i V(L_i) + n_i V(L_i)^2 \\ &= \sum_{i=1}^{m} n_i \left(\hat{V}_i - 2\bar{V}_i V(L_i) + V(L_i)^2 \right) \\ &= \sum_{i=1}^{m} n_i \left(\hat{V}_i - 2\bar{V}_i a L_i^b + a^2 L_i^{2b} \right). \end{aligned} \tag{2.2.2.5}$$

Låt

$$r_i = \sqrt{n_i \left(\hat{V}_i - 2\bar{V}_i a L_i^b + a^2 L_i^{2b} \right)} \tag{2.2.2.6}$$

så att

$$E_V = \sum_{i=1}^{m} r_i^2. \tag{2.2.2.7}$$

Då kan E_V minimeras med Gauss-Newtons metod [2], som utifrån en startgissning

$$\boldsymbol{\beta} = [a \ b]^T = [1 \ 1]^T \tag{2.2.2.8}$$

gör iterationen

$$\boldsymbol{\beta}^{(s+1)} = \boldsymbol{\beta}^{(s)} - \delta(\boldsymbol{J}^T \boldsymbol{J})^{-1} \boldsymbol{J}^T \boldsymbol{r}(\boldsymbol{\beta}^{(s)}) \tag{2.2.2.9}$$

där δ är en konstant, $\boldsymbol{r} = [r_1 \ldots r_m]^T$ och Jacobimatrisen $\boldsymbol{J}$ har element J_{ij}

$$J_{ij} = \frac{\partial r_i}{\partial \beta_j} \tag{2.2.2.10}$$

som beräknas enligt

$$\frac{\partial r_i}{\partial \beta_1} = \frac{\partial r_i}{\partial a} = \frac{1}{2r_i}\frac{\partial r_i^2}{\partial a} = \frac{n_i}{r_i}\left(aL_i^{2b} - \hat{V}_i L_i^b\right), \tag{2.2.2.11}$$

$$\frac{\partial r_i}{\partial \beta_2} = \frac{\partial r_i}{\partial b} = \frac{1}{2r_i}\frac{\partial r_i^2}{\partial b} = \frac{n_i \log(L_i)}{r_i}\left(a^2 L_i^{2b} - aL_i\right). \tag{2.2.2.12}$$

Likt vid normalisering av längd skrivs termen $(\boldsymbol{J}^T\boldsymbol{J})^{-1}\boldsymbol{J}^T\boldsymbol{r}(\boldsymbol{\beta}^{(s)})$ om för ökad numerisk stabilitet. Detta görs genom att QR-faktorisera $\boldsymbol{J}$:

$$\boldsymbol{J} = \boldsymbol{Q}\boldsymbol{R} = [\boldsymbol{Q}_1 \ \boldsymbol{Q}_2]\begin{bmatrix}\boldsymbol{R}_1 \\ \boldsymbol{0}\end{bmatrix} = \boldsymbol{Q}_1\boldsymbol{R}_1 \tag{2.2.2.13}$$

där $\boldsymbol{Q}_1$ har ortogonala kolonner och $\boldsymbol{R}_1$ är övertriangular. Denna faktorisering ger att

$$\begin{aligned}(\boldsymbol{J}^T\boldsymbol{J})^{-1}\boldsymbol{J}^T\boldsymbol{r}(\boldsymbol{\beta}^{(s)}) &= ((\boldsymbol{Q}_1\boldsymbol{R}_1)^T(\boldsymbol{Q}_1\boldsymbol{R}_1))^{-1}(\boldsymbol{Q}_1\boldsymbol{R}_1)^T\boldsymbol{r}(\boldsymbol{\beta}^{(s)}) \\ &= (\boldsymbol{R}_1^T\boldsymbol{Q}_1^T\boldsymbol{Q}_1\boldsymbol{R}_1)^{-1}\boldsymbol{R}_1^T\boldsymbol{Q}_1^T\boldsymbol{r}(\boldsymbol{\beta}^{(s)}) \\ &= (\boldsymbol{R}_1^T\boldsymbol{R}_1)^{-1}\boldsymbol{R}_1^T\boldsymbol{Q}_1^T\boldsymbol{r}(\boldsymbol{\beta}^{(s)}) \\ &= \boldsymbol{R}_1^{-1}\boldsymbol{R}_1^{-T}\boldsymbol{R}_1^T\boldsymbol{Q}_1^T\boldsymbol{r}(\boldsymbol{\beta}^{(s)}) \\ &= \boldsymbol{R}_1^{-1}\boldsymbol{Q}_1^T\boldsymbol{r}(\boldsymbol{\beta}^{(s)})\end{aligned} \tag{2.2.2.14}$$

där det nyttjats att $\boldsymbol{Q}_1^T\boldsymbol{Q}_1 = \boldsymbol{I}$. Därmed görs iterationen

$$\boldsymbol{\beta}^{(s+1)} = \boldsymbol{\beta}^{(s)} - \delta\boldsymbol{x} \tag{2.2.2.15}$$

där $\boldsymbol{x}$ ges av

$$\boldsymbol{R}_1\boldsymbol{x} - \boldsymbol{Q}_1^T\boldsymbol{r}(\boldsymbol{\beta}^{(s)}) = \boldsymbol{0}. \tag{2.2.2.16}$$

Gauss-Newton-iterationen avslutas när

$$||\boldsymbol{x}|| < 10^{-10}. \tag{2.2.2.17}$$

Detta ger en funktion $V(L)$ för medelvikt vid en längd L, vilket gör det möjligt att normalisera vikt till en relativ avvikelse från medel enligt

$$v_i = \frac{V_i - V(L_i)}{V(L_i)} \tag{2.2.2.18}$$

där v_i är den normaliserade längden för fisk i.

2.2.3 Position

I den valda datan är positionen för varje tråldrag angiven med start- och slutposition. Som position för en given datapunkt används medelvärdet av dessa två positioner. Positionerna är angivna i grader latitud och longitud, vilket är problematiskt vid avståndsberäkning då en grad latitud inte nödvändigtvis

motsvarar samma fysiska avstånd som en grad longitud (de är endast lika längs ekvatorn). Det innebär att skillnader i latitud och longitud inte enkelt kan sättas samman till ett fysiskt avstånd. För att lösa detta problem används storcirkelavståndet, som är det kortaste avståndet längs en storcirkel mellan två punkter. Storcirkelavståndet d mellan två punkter ges av

$$d = \cos^{-1}\left(\sin(\phi_1)\sin(\phi_2) + \cos(\phi_1)\cos(\phi_2)\cos(|\lambda_1 - \lambda_2|)\right) \tag{2.2.3.1}$$

där ϕ_1, λ_1 och ϕ_2, λ_2 är punkternas longituder och latituder [8]. Detta är avståndet (mätt i grader) som används vid beräkning av likhet för två punkter (se avsnitt 2.3.1).

2.3 Klustring

Klustring innebär att en mängd datapunkter delas in i grupper (kluster) sådant att punkter som är lika placeras i samma kluster. Denna likhet representeras av en graf som för n datapunkter är $n \times n$, där element i, j är ett icke-negativt tal som anger avståndet mellan punkterna i och j. Varje element i matrisen bestäms av en avståndfunktion, som är beskriven nedan. Matrisen används sedan i en klustringsalgoritm för att bestämma en klustring enligt givna kriterier.

I detta avsnitt beskrivs två klustringsalgoritmer: k-means och en spektral klustringsalgoritm. k-means är del av den spektrala klustringsalgoritmen, men kan också användas enskilt för att klustra punkter. Dock har den en del stora nackdelar jämfört med spektrala klustringsalgoritmer, då k-means är begränsad i formen på de kluster som hittas (endast konvexa kluster är möjliga) och algoritmen är känslig för valet av ursprungliga centrum till varje kluster. Känsligheten för ursprungscentrum går att minska, men inte helt eliminera, genom att använda väl valda ursprungscentrum (som beskrivet i avsnitt 2.3.2 nedan). Den begränsade formen på kluster går å andra sidan inte att påverka i k-means-algoritmen, vilket är anledningen till att denna undersökning istället använder spektral klustring.

2.3.1 Avståndsfunktion

Avståndsfunktionen $A(p_1, p_2)$ representerar avståndet mellan två punkter p_1 och p_2, där punkter med lägre avstånd bör placeras i samma kluster. Låt

$$L_A(p_1, p_2) = |L(p_1) - L(p_2)| \tag{2.3.1.1}$$

$$V_A(p_1, p_2) = |V(p_1) - V(p_2)| \tag{2.3.1.2}$$

där $L(p)$ och $V(p)$ är normaliserad längd respektive vikt (se avsnitt 2.2.1 och 2.2.2) och låt $D(p_1, p_2)$ vara det fysiska avståndet mellan två punkter (mätt i grader) enligt avsnitt 2.2.3. Avståndet $A(p_1, p_2)$ mellan punkterna ges då av funktioner på formen

$$A(p_1, p_2) = (1 + \alpha L_A(p_1, p_2) + \beta V_A(p_1, p_2))D(p_1, p_2)^2 \tag{2.3.1.3}$$

där koefficienterna α och β väljs beroende på önskad klustring. Notera att skalning av $A(p_1, p_2)$ inte är nödvändig då detta hanteras i omvandlingen från avståndsfunktion till viktfunktion enligt ekvation 2.3.3.1.

Denna avståndsfunktion syftar till att hitta sammanhängande geografiska områden med liknande längd eller vikt. Att multiplicera med kvadraten av fysiskt avstånd gör att avståndsfunktionen blir mycket hög för punkter långt ifrån varandra, vilket främjar sammanhängande kluster. Den konstanta termen garanterar att detta gäller även om L_A och V_A är mycket små, men naturligtvis i lägre grad än om L_A eller V_A är stor.

2.3.2 k-means

k-means är en enkel klustringsmetod som används som del av spektral klustring. Klustring med k-means görs enligt följande [5]:

1. Välj k ursprungliga centrum $c_1, \ldots, c_k$.
2. För varje datapunkt p, låt p tillhöra klustret C_i om p är närmre c_i än c_j för alla $i \neq j$.
3. För varje kluster C_i, låt $c_i = \frac{1}{|C_i|} \sum_{p \in C_i} p$.
4. Upprepa steg 2 och 3 tills klustringen inte längre förändras.

Notera att detta sätt att dela in punkter i kluster, som tidigare nämnt, innebär att alla kluster är konvexa.

k-means är också väldigt känslig för det ursprungliga valet av centrum i steg 1; det finns inte någon garanti att iterationen konvergerar mot ett globalt minimum. Metoden för att välja ursprungliga punkter som använts är k-means++, som är beskriven nedan.

k-means: val av ursprungscentrum

Ursprungliga centrum till k-means väljs normalt sett genom att slumpmässigt välja k datapunkter och låta dessa vara centrum. Att välja centrum från en uniform fördelning leder dock ofta till dåliga resultat, då exempelvis mycket lika datapunkter (som bör vara del av samma kluster) kan väljas som ursprungliga centrum till olika kluster. För att förbättra sannolikheten att hitta en bra klustring kan andra algoritmer användas för att väljs ursprungliga centrum.

k-means++ [5] är en sådan algoritm, som med större sannolikhet hittar bra ursprungliga centrum. k-means++ väljer centrum slumpmässigt från mängden datapunkter, men viktar dessa punkter beroende på närhet till redan valda centrum. Låt $D(p)$ vara avståndet från punkten p till närmsta redan valda centrum och låt P vara mängden av alla datapunkter. Då väljer k-means++ centrum enligt följande:

1. Välj en slumpmässig datapunkt p från P och låt $c_1 = p$.
2. För nästa centrum c_i, välj $c_i = p'$ med sannolikhet $\frac{D(p')^2}{\sum_{p \in P} D(p)^2}$.

3. Upprepa steg 2 tills k centrum valts.

Som nämnt förbättrar detta sannolikheten för ett bra val av ursprungliga centrum, men slump har fortfarande en betydande påverkan. Därför upprepas k-means-steget 1000 gånger och den bästa klustringen enligt kostnadsfunktionen (ekvation 2.3.3.6) väljs.

2.3.3 Spektral klustring

Spektral klustring används för att dela in en graf i kluster, där noder i grafen motsvarar datapunkter och (viktade) kanter mellan dessa noder bestäms av nodernas avstånd i avståndsmatrisen. För denna grafrepresentation är kanternas vikter en transformation av avståndet, då vikten önskas vara hög för en kant mellan noder som har kort avstånd. För en avståndmatris A med element $a_{i,j}$ ges viktmatrisen W med element $w_{i,j}$ givna av

$$w_{i,j} = \exp\left(-\frac{a_{i,j}}{2\sigma^2}\right) \tag{2.3.3.1}$$

där σ är en konstant som avgör spridningen för varje kluster. Ett möjligt sätt att klustra dessa noder är genom att minimera summan av vikterna av kanterna mellan de olika klustrena. En klustring med k kluster kan alltså hittas genom att beräkna

$$\min_{A_1,\ldots,A_k \subset V} \sum_{i=1}^{k} W(A_i, \bar{A}_i) \tag{2.3.3.2}$$

där

$$W(A, \bar{A}) = \sum_{i \in A, j \in \bar{A}} w_{i,j} \tag{2.3.3.3}$$

och $\bar{A} = V \backslash A$. Detta leder dock till att avvikande datapunkter delas in i kluster för sig (eftersom avvikande punkters avstånd är höga, vilket leder till låga vikter). Ett exempel på hur detta kan se ut går att se i figur 5a. Spektral klustring syftar till att lösa detta problemet genom att även ta hänsyn till storleken $S(A)$ på varje kluster och istället beräkna

$$\min_{A_1,\ldots,A_k \subset V} \sum_{i=1}^{k} \frac{W(A_i, \bar{A}_i)}{S(A_i)}. \tag{2.3.3.4}$$

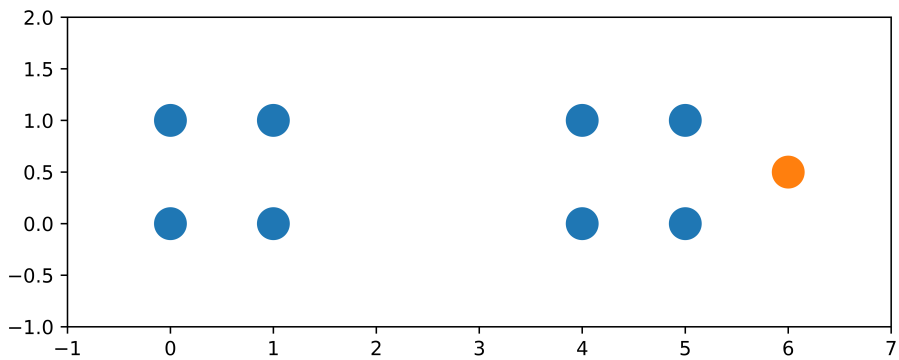
Denna storlek kan mätas på två sätt: antingen har varje nod samma storlek, så att storleken är antalet noder i klustret, eller så är storleken på varje nod summan av vikterna av alla kanter till noden och klustrets storlek är summan av alla noders storlek. Den senare av dessa, alltså

$$S(A) = \mathrm{vol}(A) = \sum_{i \in A} d_i = \sum_{i \in A} \sum_{j=1}^{n} w_{i,j} \tag{2.3.3.5}$$

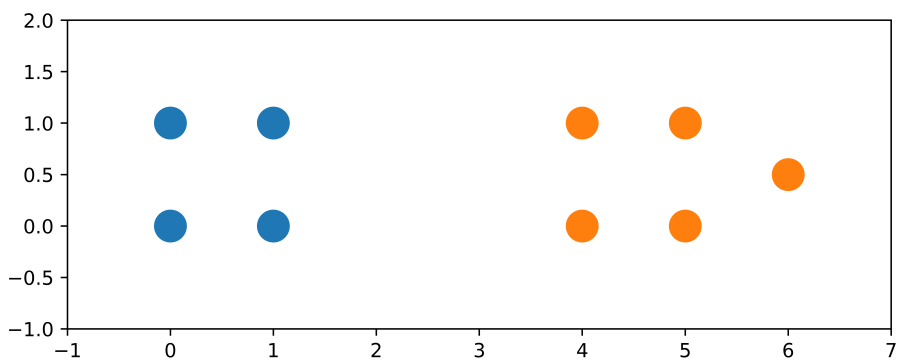
används för denna undersökning. Detta på grund av att d_i är lågt för avvikande datapunkter, vilket ytterligare straffar kluster med enskilda avvikande datapunkter. Klustringen görs då genom att beräkna

$$\min_{A_1,\ldots,A_k \subset V} \sum_{i=1}^{k} \frac{W(A_i, \bar{A}_i)}{\mathrm{vol}(A_i)}. \tag{2.3.3.6}$$

Figur 5 illustrerar hur detta förbättrar klustringen jämfört med ekvation 2.3.3.2.



(a) Optimal klustring enligt ekvation 2.3.3.2.



(b) Optimal klustring enligt ekvation 2.3.3.6.

Figur 5: Jämförelse mellan klustring enligt ekvation 2.3.3.2 och 2.3.3.6 med vikter $w_{i,j} = \frac{1}{||p_i - p_j||}$.

På grund av formen av $A_1, \ldots, A_k$ är denna minimering mycket svår att beräkna exakt. Det är dock möjligt att approximera klustringen enligt följande [3]: låt

$$D = \begin{bmatrix} d_1 & 0 & \ldots & 0 \\ 0 & d_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & d_n \end{bmatrix} \quad \text{där} \quad d_i = \sum_{j=1}^{n} w_{i,j} \tag{2.3.3.7}$$

och låt

$$L = D^{-1/2} W D^{-1/2}. \tag{2.3.3.8}$$

Beräkna sedan egenvektorerna $e_1, \ldots, e_k$ för de k största egenvärdena till L och skala dessa egenvektorer så att deras norm är 1. Genom att tolka raderna i matrisen

$$E = [e_1, \ldots, e_k] \tag{2.3.3.9}$$

som datapunkter kan dessa rader klustras med k-means (eller annan klustringsmetod) till k kluster. Detta ger en klustring av ursprungsproblemet där punkt i i ursprungsproblemet tillhör kluster j om rad i i E tillhör kluster j i resultatet av k-means-klustringen.

2.3.4 Rekursiv klustring

Trots viktningen efter klusterstorlek i ekvation 2.3.3.6 leder spektral klustring inte alltid till jämnstora kluster. För att uppnå en jämn fördelning av klusterstorlek används därför rekursiv klustring. Detta innebär att datan först delas in i k kluster av varierande storlek. Därefter delas varje kluster med fler än ett givet maximalt antal datapunkter C_{max} in i två mindre kluster. Detta upprepas tills inget kluster är större än C_{max}.

3 Resultat

3.1 Normaliseringskurvor

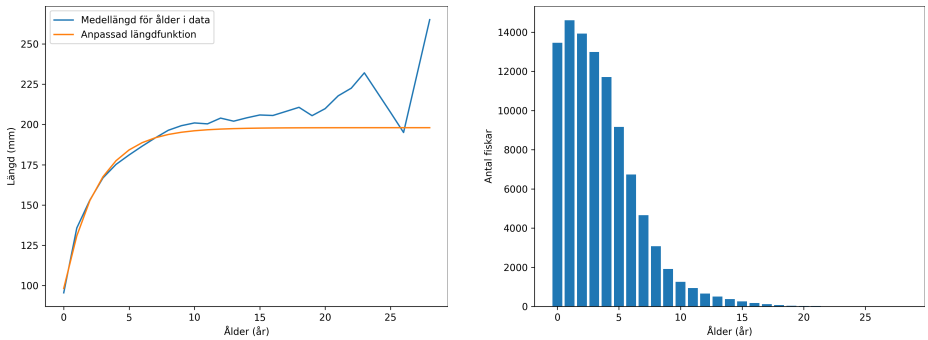
Anpassning av ekvation 2.2.1.1 till datan (enligt avsnitt 2.2.1) ger att medellängden $L(t)$ vid ålder t är

$$L(t) = L_\infty \left(1 - e^{-K(t-t_0)}\right) \tag{3.1.1}$$

där

$$\begin{aligned} L_\infty &\approx 198.0, \\ K &\approx 0.3955, \\ t_0 &\approx -1.735. \end{aligned} \tag{3.1.2}$$

Denna funktion går att se i figur 6 tillsammans med medellängd i datan för varje ålder. Notera att längder vid högre åldrar innehåller få datapunkter, vilket är anledningen till att medellängden för de åldrarna varierar mer samt att den anpassade kurvan inte följer den delen av datan lika nära som de låga åldrarna.



(a) Anpassad funktion och medellängd

(b) Antal fiskar för varje ålder

Figur 6: Anpassad funktion för längd vid given ålder

Vikt som funktion av längd anpassas på liknande vis till datan (se avsnitt 2.2.2), med en funktion enligt ekvation 2.2.2.1:

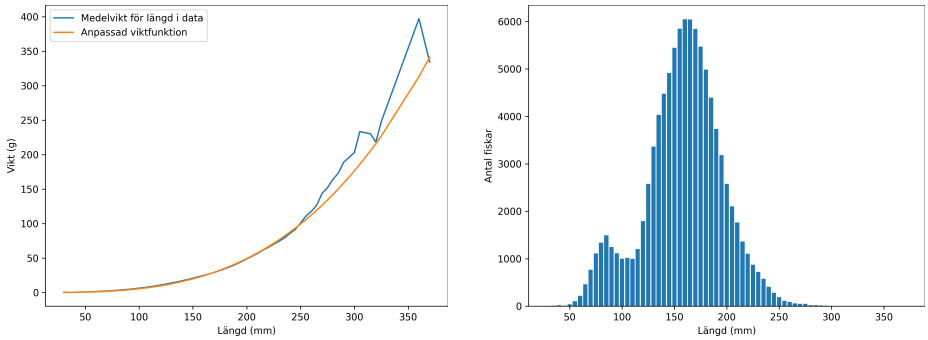
$$V(L) = aL^b \tag{3.1.3}$$

där

$$\begin{aligned} a &\approx 2.858 \\ b &\approx 3.145 \end{aligned} \tag{3.1.4}$$

vilket ger en funktion som går att se i figur 7 tillsammans med medelvikt i datan för varje längd. Likt längd varierar antalet datapunkter för varje medelvärde, vilket är anledningen till den ojämna medelvikten i datan för långa fiskar.

Dessa kurvor används för normalisering enligt ekvationer 2.2.1.22 och 2.2.2.18.



(a) Anpassad funktion och medelvikt

(b) Antal fiskar för varje längd

Figur 7: Anpassad funktion för vikt vid given längd

3.2 Klustring

3.2.1 Klustring efter längd

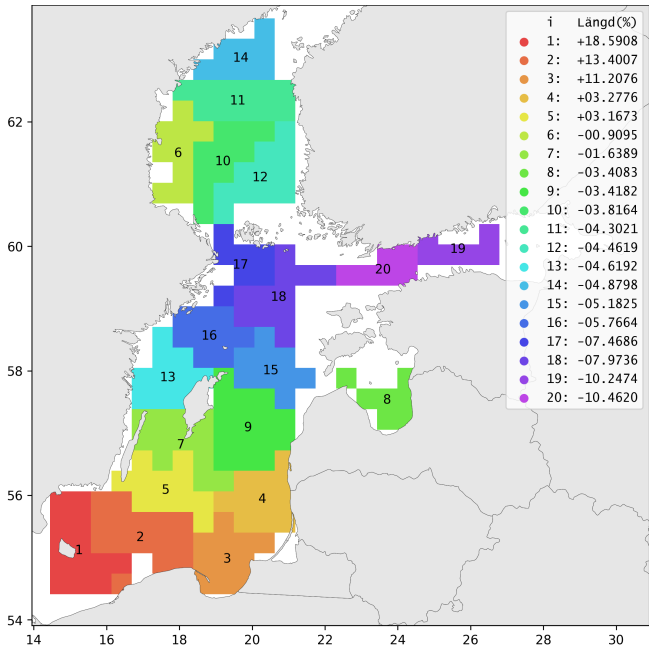
Klustring efter längd görs med avståndfunktion enligt ekvation 2.3.1.3 med $\alpha = 70$ och $\beta = 0$, alltså

$$A(p_1, p_2) = (1 + 70L_A(p_1, p_2))D(p_1, p_2)^2. \tag{3.2.1.1}$$

Denna omvandlas till en viktfunktion enligt ekvation 2.3.3.1 med $\sigma = 1$. Antalet kluster k är 16 och rekursiv klustring enligt avsnitt 2.3.4 tillämpas med högst $C_{max} = 16$ rektangulära områden per kluster. Resultatet av klustringen går att se i figur 8 tillsammans med normaliserad medellängd för varje kluster.

Denna klustring visar att fiskar i Österskön är längst i den sydvästra delen av Östersjön (kluster 1-3) och kortast i Finska viken (kluster 19 och 20). Mellan dessa områden finns en tydlig gradient, där längd ökar längre söderut.

Längre norrut håller dock inte mönstret med längre fiskar längre söderut. Istället är fiskar i Bottniska viken (kluster 6, 10, 11, 12 och 14) längre än de fiskar som finns i Finska viken och dess närhet. Variationerna i detta område är relativt små, med undantag för kluster 6 i den västra delen där fiskarna är relativt stora jämfört med resten av Bottniska viken.



Figur 8: Klustring med $\alpha = 70$, $\beta = 0$, $\sigma = 1$, $k = 16$, $C_{max} = 16$

3.2.2 Klustring efter vikt

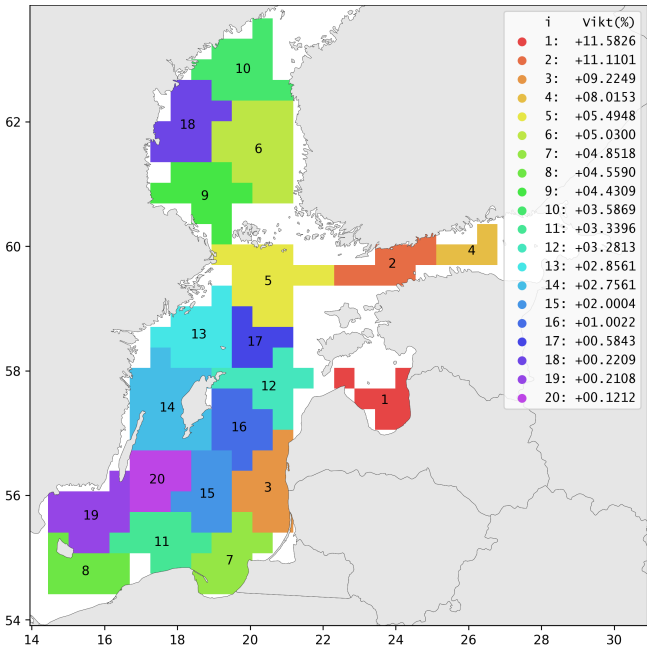
Klustring efter vikt görs med avståndfunktion enligt ekvation 2.3.1.3 med $\alpha = 0$ och $\beta = 25$, alltså

$$A(p_1, p_2) = (1 + 25V_A(p_1, p_2))D(p_1, p_2)^2. \tag{3.2.2.1}$$

Denna omvandlas till en viktfunktion enligt ekvation 2.3.3.1 med $\sigma = 1$. Antalet kluster k är 16 och rekursiv klustring enligt avsnitt 2.3.4 tillämpas med högst $C_{max} = 16$ rektangulära områden per kluster. Resultatet av klustringen går att se i figur 9 tillsammans med normaliserad medelvikt för varje kluster.

Till skillnad från klustringen efter längd visar denna klustring inte några tydliga mönster över stora områden. Dock finns flera kluster som skiljer sig betydligt från närliggande områden. De minsta klustrena (19 och 20) finns vid den södra delen av Sverige, nära de betydligt större klustrena (7, 8 och 11) i den mest södra delen av Östersjön. Kluster 18 sticker också ut som mycket litet i den västra delen av Bottniska viken, återigen avvikande från närliggande kluster (6, 9 och 10).

Tunga fiskar (för sin längd) går att hitta i finska viken, Rigabukten och vid Litauens västkust. Finska viken innehåller två kluster med tunga fiskar (2 och 4) och även där viken mynnar ut i resten av Östersjön (kluster 5) är fiskarna tunga. Rigabukten sticker ut med de allra tyngsta fiskarna (kluster 1), men är på grund av hål i insamlingsområden går det inte att jämföra med angränsande



Figur 9: Klustring med $\alpha = 0$, $\beta = 25$, $\sigma = 1$, $k = 16$, $C_{max} = 16$

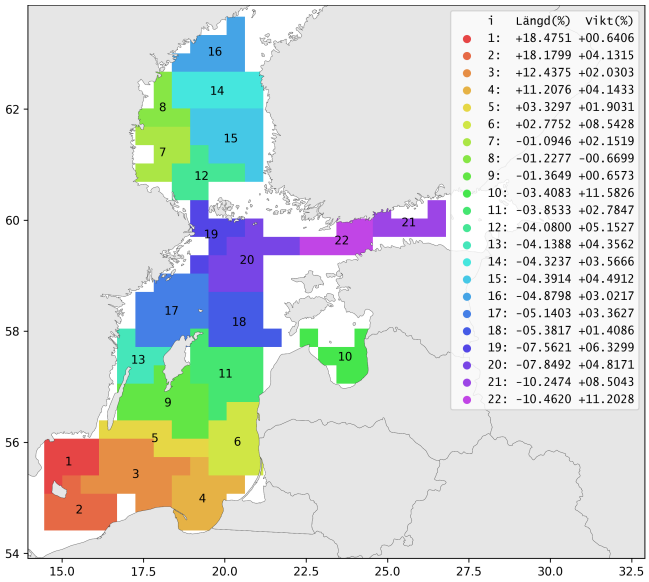
kluster. Slutligen finns tunga fiskar i kluster 3 vid Litauens västkust, som sticker ut då angränsande kluster (7, 12, 15 och 16) innehåller relativt lätta fiskar.

3.2.3 Klustring efter längd och vikt

För klustring efter både längd och vikt används en avståndfunktion som är medelvärdet av funktionerna för klustring efter enbart längd eller vikt (ekvationer 3.2.1.1 och 3.2.2.1). Detta ger en avståndfunktion enligt ekvation 2.3.1.3 med $\alpha = 35$ och $\beta = 12.5$, alltså

$$A(p_1, p_2) = (1 + 35L_A(p_1, p_2) + 12.5V_A(p_1, p_2))D(p_1, p_2)^2. \tag{3.2.3.1}$$

Denna omvandlas till en viktfunktion enligt ekvation 2.3.3.1 med $\sigma = 1$. Antalet kluster k är 16 och rekursiv klustring enligt avsnitt 2.3.4 tillämpas med högst $C_{max} = 16$ rektangulära områden per kluster. Resultatet av klustringen går att se i figur 10 tillsammans med normaliserad medellängd och medelvikt för varje kluster. Sorteringen och färgläggningen av kluster baseras på längd.



Figur 10: Klustring med $\alpha = 35$, $\beta = 12.5$, $\sigma = 1$, $k = 16$, $C_{max} = 16$

Denna klustring avgränsar tydligare de områden som identifierats som avvikande i både längd och vikt. Dessa inkluderar den västra delen av Bottniska viken, som nu delats in i två kluster (7 och 8) som visar att speciellt den nordvästra delen (kluster 8) är avvikande, med hög längd och låg vikt jämfört med närliggande kluster. Även kluster 7 avviker åt samma håll, men med lite högre vikt.

Ett annat område med hög längd och låg vikt är det som markeras av kluster 1. Närliggande kluster har antingen lika hög längd med också hög vikt (kluster 2) eller ligger närmre i både längd och vikt (kluster 3 och 5) men ändå med både kortare och tyngre fiskar än i kluster 1.

Notera även att det inte finns något tydligt samband mellan avvikande vikt och avvikande längd.

4 Diskussion

4.1 Klustring efter längd

Klustringen efter längd (se avsnitt 3.2.1) visade framförallt stora mönster, som tyder på att längd beror framförallt på geografi. Casini m.fl. [6] har hittat liknande mönster i data insamlad mellan 1992 och 2008, vilket stärker dessa resultat. Dock inkluderar den studien endast data för den södra delen av Östersjön (söder om 60 grader latitud, men utan finska viken). Klustringen visar att resultaten hittade av Casini m.fl. inte går att extrapolera till den norra delen av Östersjön, då längd ökar längre norrut (tvärtemot mönstret söder om 60 grader).

Den västra delen av Bottniska viken (kluster 6 i figur 8) kan vara av speciellt intresse, då detta område skiljer sig betydligt från närliggande kluster. Det är därför möjligt att området innehåller en avvikande population av strömming.

4.2 Klustring efter vikt

Klustring efter vikt pekade ut flera områden med potentiellt avvikande populationer, som beskrivna i avsnitt 3.2.2. Både Rigabukten och området vid Litauens västkust hade fiskar med relativt hög vikt, där Litauens västkust är av intresse för de stora skillnaderna jämfört angränsande området. Rigabukten är ganska skilt geografiskt från resten av Östersjön och har därför inga angränsande kluster som går att jämföra med. Denna geografiska isolering ger möjlighet till skillnader både i miljö och genetik, men som beskrivet i avsnitt 4.4 är det också möjligt att avvikelsen i just detta område beror på sättet datan är insamlad.

Även möjliga populationer med lätta fiskar identifierades, ett sådant i västra delen av Bottniska viken och ett nära Sveriges södra kust. Båda dessa avviker från närliggande kluster, men området i Bottniska viken är troligen mer intressant då det båda avviker kraftigare från angränsande områden och täcker ett mindre geografiskt område.

4.3 Klustring efter längd och vikt

Klustringen efter längd och vikt (avsnitt 3.2.3) används för att tydligare avgränsa områden som avviker i både längd och vikt. Ett sådant område är den västra delen av Bottniska viken, som stack ut både i klustringen efter enbart längd och efter enbart vikt. Detta område motsvarar kluster 7 och 8 i figur 10, där kluster 8 kan vara av extra intresse med de allra lättaste fiskarna trots längre fiskar än närliggande områden. Även kluster 7 sticker ut på samma vis, men med lite tyngre fiskar. Dessa kluster tyder på att det finns en avvikande population i västra delen av Bottniska viken.

Ett annat kluster som kan vara av intresse är kluster 1, öster om Skåne. Detta område pekades ut som ett med lätta fiskar i klustringen efter vikt och klustringen efter både längd och vikt visar att området trots högst längd har mycket låg vikt, med nämnvärd skillnad gentemot närliggande kluster (2, 3 och 5). Det är alltså möjligt att detta område har en avvikande population.

Av intresse är också att det inte går att se något samband mellan hög längd och hög vikt, vilket tyder på att variationer beror på andra orsaker än tillgång till föda (eftersom tillgång till föda bör påverka både längd och vikt).

4.4 Årlig variation i data

Som nämnt i avsnitt 2.1.1 är den använda datan insamlad under september och oktober mellan 2007 och 2023. Med data insamlad under så lång tid finns en risk att långtida förändringar eller årliga variationer i längd eller vikt har en påverkan på de resultat som uppnåtts.

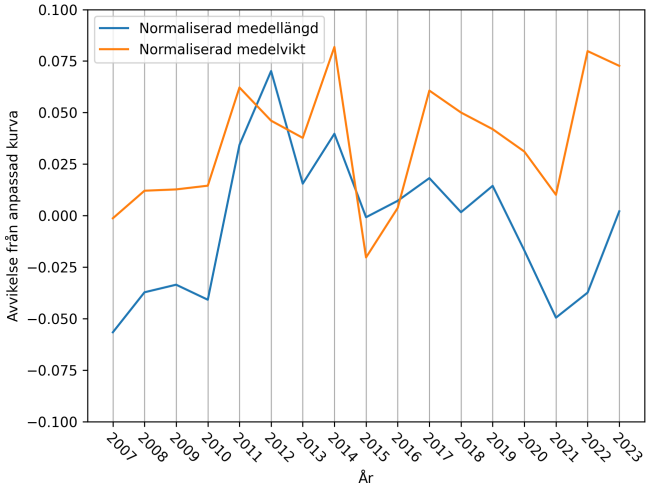
Beräkning av normaliserad medellängd för varje år (se figur 11) antyder att det mycket riktigt finns en betydande variation mellan olika år, även om inga tydliga långtida förändringar går att se. Det är dock sannolikt att skillnaderna i denna graf inte enbart beror på skillnader i fiskstorlek i Östersjön utan även orsakas av skillnader i insamlingsområden. Inte alla deltagande länder har samlat in data varje år, vilket leder till att delar av Östersjön saknar data vissa år. Detta påverkar förstås medellängden. Dessutom finns skillnader i antalet fångade fiskar i olika delar av östersjön, som också har en påverkan.

Ett exempel på skillnaderna i insamlingsområden går att se i figur 12, där 2011 jämfört med 2010 har ytterligare data i den södra delen av Östersjön. I figur 11 motsvaras detta av en kraftig ökning i medellängd för 2011 jämfört med 2010, vilket är förväntningen utifrån områdena eftersom den södra delen av Östersjön (enligt figur 8) har långa fiskar.

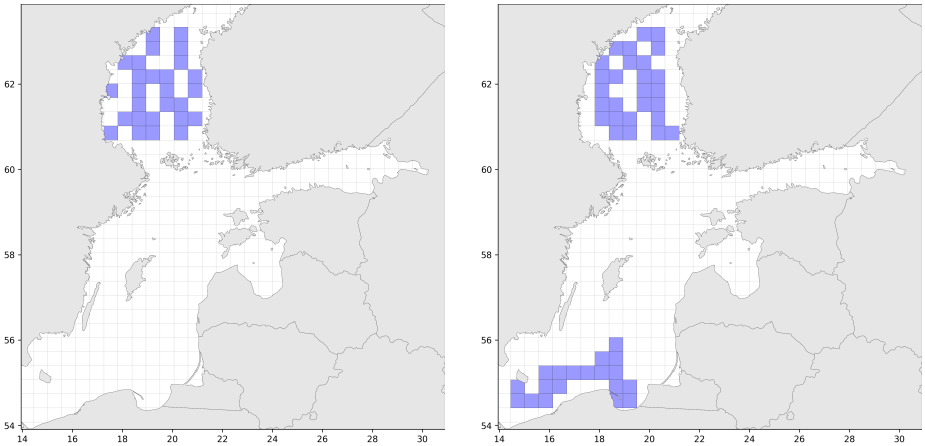
Dock går inte alla årliga variationer att förklara med insamlingsområden. Från 2014 och framåt finns data från hela Östersjön varje år, med undantag för 2018 som saknar data i Finska viken och Rigabukten där data endast finns från 2022 och 2023. Trots det finns stora variationer, som sannolikt orsakas av varierande tillgång till föda. Notera att variationer i längd och vikt för det mesta håller samma riktning; en ökning i medellängd sker samtidigt som en ökning i medelvikt och vice versa. Detta stöder teorin att variationerna orsakas av tillgång till föda.

Att alla år 2014 och framåt har data från nästan hela Östersjön (med undantag för Rigabukten) innebär att årliga variationer sannolikt jämnas ut, men det är möjligt att avvikelser under åren med lite data leder till att årliga variationer istället tolkas som avvikande områden. Områden som kunnat påverkas på detta vis är den södra delen av Östersjön samt Bottniska viken, som går att se i insamlingsområdena för 2011 (se figur 12). Klustring med årtal före 2014 exkluderade visar emellertid liknande skillnader som går att se när all data inkluderas (se figur 8, 9 och 10). Detta tyder på att de årliga variationerna samt skillnaderna i insamlingsområden inte har en betydande påverkan på resultaten.

Bristen på data i Rigabukten före 2022 är dock något som sannolikt har en betydande påverkan på resultaten för det området. Klustringen visar att fiskar i Rigabukten är de tyngsta i Östersjön, men som går att se i figur 11 så är 2022 och 2023 två årtal med hög medelvikt. Det är således sannolikt att den höga vikten i Rigabukten åtminstone delvis orsakas av insamlingsåren snarare än en avvikande population.



Figur 11: Normaliserad medellängd och medelvikt för varje årtal



(a) 2010 (b) 2011

Figur 12: Skillnad i insamlingsområden mellan 2010 och 2011

5 Slutsats

Denna klustringsanalys visar att det finns småskaliga variationer i både längd och vikt inom Östersjön, något som det i nuläget finns bristande kunskap om. Det är möjligt att andra metoder eller datamängder kan användas för att identifiera andra områden eller tydligare hitta mönster i variationer, vilket öppnar upp en möjlighet för vidare studier. Dessa andra metoder skulle exempelvis kunna vara mer avancerade klustringsmetoder eller statistiska metoder.

I denna studie har fyra avvikande områden har identifierats: den västra delen av Bottniska viken, området öster om Skåne, området vid Litauens västkust och Rigabukten. De avvikande värdena i Rigabukten beror troligtvis på skillnader i insamlingsår mellan området och resten av Östersjön (se avsnitt 4.4). De resterande tre områdena förklaras dock inte av insamlingsområden, vilket innebär att det sannolikt finns avvikande populationer av strömming i de områdena. Speciellt området i Bottniska viken och vid Skåne är intressanta då dessa skilde sig i både längd och vikt.

Dessa områden bör därför studeras vidare, med målet att identifiera orsaker till skillnaderna. Om avvikelserna beror på genetiska skillnader kan populationerna vara av speciellt intresse för bevarandeinsatser.

6 Referenser

[1] H. Sparholt, "Fish Species Interactions in the Baltic Sea," *Dana*, årg. 10, maj 1994.

[2] Å. Björck, *Numerical Methods for Least Squares Problems*. Society for Industrial och Applied Mathematics, 1996. DOI: 10.1137/1.9781611971484.

[3] A. Y. Ng, M. I. Jordan och Y. Weiss, "On spectral clustering: analysis and an algorithm," i *Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic*, ser. NIPS'01, Vancouver, British Columbia, Canada: MIT Press, 2001, s. 849–856.

[4] G. M. Cailliet, W. de Smith, H. F. Mollet och K. J. Goldman, "Age and growth studies of chondrichthyan fishes: the need for consistency in terminology, verification, validation, and growth function fitting," *Environmental Biology of Fishes*, årg. 77, s. 211–228, 2006. DOI: 10.1007/s10641-006-9105-5.

[5] D. Arthur och S. Vassilvitskii, "K-Means++: The Advantages of Careful Seeding," *Proc. of the Annu. ACM-SIAM Symp. on Discrete Algorithms*, årg. 8, s. 1027–1035, jan. 2007. DOI: 10.1145/1283383.1283494.

[6] M. Casini, G. Kornilovs, M. Cardinale m. fl., "Spatial and temporal density dependence regulates the condition of central Baltic Sea clupeids: compelling evidence using an extensive international acoustic survey," *Population Ecology*, årg. 53, nr 4, s. 511–523, 2011. DOI: https://doi.org/10.1007/s10144-011-0269-2.

[7] Z. Pekcan-Hekim, A. Gårdmark, A. M. L. Karlson, P. Kauppila, M. Bergenius och L. Bergström, "The role of climate and fisheries on the temporal changes in the Bothnian Bay foodweb," *ICES Journal of Marine Science*, årg. 73, nr 7, s. 1739–1749, mars 2016, ISSN: 1054-3139. DOI: 10.1093/icesjms/fsw032.

[8] L. Råde och B. Westergren, *Mathematics Handbook for Science and Engineering*. Studentlitteratur, 2019, ISBN: 9789144128436.

[9] T. Raid och E. Sepp, "Variability in the structure of pelagic fish communities: A pitfall for management of herring in the Baltic Sea?" I *Trends in Maritime Technology and Engineering: Proceedings of the 6th International Conference on Maritime Technology and Engineering*, T. Santos och C. G. Soares, utg., 1. utg., vol. 2, CRC Press, 2022, ISBN: 9781003320289. DOI: 10.1201/9781003320289-65.

[10] ICES, *Baltic International Acoustic Survey*, Hämtad 2024-03-04. URL: https://acoustic.ices.dk/submissions.

[11] Natural Earth, *1:10m Cultural Vectors, Admin 0 - Countries*, Version 5.1.1, Hämtad 2024-02-01. URL: https://www.naturalearthdata.com/downloads/10m-cultural-vectors/.