

Statistical modelling of sea lice count data from salmon farms in the Faroe Islands

H Gislason 

Faculty of Science and Technology,
University of the Faroe Islands, Nóatún 3,
FO-100 Tórshavn, Faroe Islands &
Fiskaaling - Aquaculture Research Station
of the Faroes, Við Áir 11, FO-430 Hvalvík,
Faroe Islands

Correspondence

Hannes Gislason, The Faculty of Science and
Technology, University of the Faroe Islands,
Tórshavn, Faroe Islands.
Email: hannesg@setur.fo

Abstract

Fiskaaling regularly counts the number of sea lice in the attached development stages (chalimus, mobiles and adult) for the salmon farms in the Faroe Islands. A statistical model of the data is developed. In the model, the sea-lice infection is represented by the chalimus (or mobile) lice developing into adult lice and is used to simulate past and current levels of adult lice—including treatments—as well as to predict the adult sea lice level 1–2 months into the future. Time series of the chalimus and adult lice show cross-correlations that shift in time and grow in size with temperature. This implies in situ the temperature-dependent development times of about 56 down to 42 days and the inverted development times (growth rates) of 0.018 up to 0.024 lice/day at 8–10°C. The temperature dependence $D(T) = \alpha_1(T + \alpha_2)^{\alpha_3} = 17,840(T + 7.439)^{-2.128}$ is approximated by $D_1(T) = 105.2 - 6.578T \approx 49$ days at the mean temperature 8.5°C—similar to $D_{cha}(T) = 100.6 - 6.507T \approx 45$ days from EWOS data. The observed development times at four sites for a year (2010–11) were 49, 50, 51 and 52 days, respectively. Finally, we estimate the sea lice production from fish farms to discuss approaches to control the sea lice epidemics—preferably by natural means. This study is useful for understanding sea lice levels and treatments, and for in situ analysis of the sea-lice development times and growth rates.

KEYWORDS

lagged correlations, sea-lice development, statistical modelling

1 | INTRODUCTION

Modelling the host–parasite system of sea lice on farmed salmon is motivated by the economic importance of the aquaculture industry and the large costs of treatments against sea lice (Liu & Bjelland, 2014), as well as the environmental conservation interests in wild salmon (Vollset et al., 2016). An extensive literature on modelling the sea lice and salmon epidemiology using different types of models was recently reviewed (Groner et al., 2016). Examples include modelling the host density effects on sea lice levels (Jansen et al., 2012), space–time modelling of the spread between and within salmon farms (Aldrin, Storvik, Kristoffersen, & Jansen, 2013), modelling parasite dynamics on farmed salmon for conservation of wild salmon (Rogers et al., 2013), deterministic modelling of infection pressure based on lice monitoring data (Kristoffersen et al., 2014), modelling resistance to chemotherapeutants (McEwan, Groner, Fast,

Gettinby, & Revie, 2015) and modelling sea lice dynamics in the seasonally changing environment (Rittenhouse, Revie, & Hurford, 2016).

A revised life cycle for sea lice consisting of eight stages was recently proposed (Hamre et al., 2013), which includes three free-living planktonic stages: nauplius (I, II) and the infectious copepod stage followed by five attached stages: chalimus (I, II), preadult (I, II) mobile males and females and the adult males and females. The total generation time of sea lice depends on water temperature and is about 98 days at 7°C (Heuch, Nordhagen, & Schram, 2000). Under laboratory conditions, the adult female can live up to 191 days and produce up to 11 pairs of egg strings each containing a mean number of 285 eggs at 7.2°C (Heuch et al., 2000). The nauplius and the infectious copepod stages last about 7 and 12 days at 7°C, respectively (Dalvin, 2016). This results in about 79 days (98 – 19) at 7°C from the infection until the first pair of egg strings are developed.

The Faroese salmon industry is troubled by high sea lice levels leading to expensive treatments against sea lice. The Aquaculture Research Station of the Faroes (Fiskaaling) regularly counts the number of sea lice in the attached development stages (chalmus, mobiles and adult) for the salmon farms in the Faroe Islands. The lice-counting regulations require counts to be made by Fiskaaling on each site about every 14 days, where a sample of 10 fish are examined from each cage. The previously permitted sea lice limit was a mean number of 2 adult females/fish, which has now been reduced to 1.5 adult females/fish.

The sea lice counts are reported to the Faroese Food and Veterinary Authority and used by the aquaculture industry for timing treatments against sea lice. Previously, the sea lice treatments were mainly based on chemotherapeutants (e.g., Alphamax/Betamax, Salmosan, Slice, Ektobann, Releeze and Interlox Paramove/hydrogen peroxide), but recently fresh water treatments and cleaner fish are increasingly being used instead of chemotherapeutants.

To better understand and help reduce the sea lice levels, the sea lice counts are also used for research at Fiskaaling. Sea lice research at Fiskaaling includes in situ-based estimation (á Norði, Simonsen, & Patursson, 2016) and modelling (Kragesteen, 2016; Patursson, Simonsen, Visser, & Patursson, 2017) of the density and dispersion of the free-living stages (nauplius, copepod) of the sea lice to estimate the sea lice self-infection and the external infection pressure within and between sites.

A method was developed for estimating the salmon louse nauplii production at fish farms (á Norði et al., 2016). At the site Sundalagið, Faroe Islands, plankton surveys were conducted around the farm with a plankton net, and at different depths in a fish cage using a plankton pump. The copepod density was low, and the highest observed copepod density was 0.3 cop./m³—similar to the density found in open waters. Nauplii dominated the planktonic stages, and the nauplii production was estimated to 26–68 naup. female⁻¹ day⁻¹ at 7.8°C.

Patursson et al. (2017) applied multiple linear regression to rank the exposure of Faroese salmon farms by their different physical geography, freshwater exchange, and tidal dispersion currents leading to varying strengths of the self- and externally driven infection pressure. Increased exposure was found to correspond to a lower rate of self-infection, but could also increase the external infection pressure (Patursson et al., 2017).

Kragesteen (2016) performed numerical simulations of the sea lice dispersal between all salmon farms in the Faroe Islands. The numerical model was based on tidal forcing and connectivity matrices between the farms were constructed. The numerical results indicated variable connection patterns between the farms—classified as being main emitters, receivers or isolated. The infection time between the farms was predicted to be within a few days up to about 2 weeks (Kragesteen, 2016).

In this study, a statistical model of the sea lice count data was developed to understand and simulate (1) the adult sea lice levels, (2) the treatments against sea lice and (3) to help reduce the sea lice levels. The development times from chalmus to adult female are about 1–2 months in Faroese waters (6–11°C) as inferred from

EWOS growth curves for sea lice as a function of temperature. Therefore, the chalmus counts can potentially be used to simulate past and current levels of adult lice, and to predict the adult sea lice level 1–2 months into the future. Additionally, time-series analysis is used to investigate the sea lice dynamics, and to independently check the in situ development times from chalmus to adult females—to be compared with the development times derived from the EWOS data (Myhr et al., 2009).

2 | MATERIALS AND METHODS

2.1 | Count data

The count data consist of time series of sea-lice counts in the different development stages (attached, mobiles and adult) found on salmon farms in the Faroe Islands. The counts are typically made at 2- to 3-week intervals, where the sea lice on ten fish are counted from each selected cage, and the overall mean count is derived for each lice stage at the site.

Counts are made of both *Lepeophtheirus salmonis* (the salmon sea lice) and *Caligus elongatus*, but the count data do not distinguish between the chalmus stage of salmon sea lice and *C. elongatus*, which are classified as one common category of attached lice.

2.2 | Statistical model

The principle behind the sea lice modelling (Figure 1) is based on the assumptions that the number of adult sea lice mainly depend on:

- The variations in the attached chalmus lice infection from local and remote sources of the free-swimming stages of sea lice.
- The different sea temperatures that lead to different development times to adult lice.
- The accumulation of adult lice from the chalmus infections.
- The survival from chalmus to adult, which is a scaling factor assumed to be constant pr. modelling.

In the model, a simulated sea-lice infection is represented by the chalmus lice developing into adult lice that accumulate by an integration of their arrival probability (Figure 1). Similarly, the count data are discrete samplings of the real infection and its accumulation to adult. The adult counts at each t_i are therefore estimated by a sum of chalmus counts $c(t_j)$ between t_1 and $t_k < t_i$ scaled by the survival (Figure 1, formula). The time t_k is set below t_i to account for the mean development time μ_d between t_k and t_i that depends on the sea temperature. For the Faroe Islands, the development times were inferred from EWOS growth curves (Myhr et al., 2009) for sea lice as a function of sea temperatures in Norway (7–17°C). This was performed by linear regression of a subset of the EWOS-development times for sea temperatures (7–11°C) that are comparable with the sea temperatures (6–11°C) observed in the Faroe Islands. The obtained linear regression formula is then used to estimate the daily development times for sea lice in the Faroe Islands from daily sea temperatures at one site

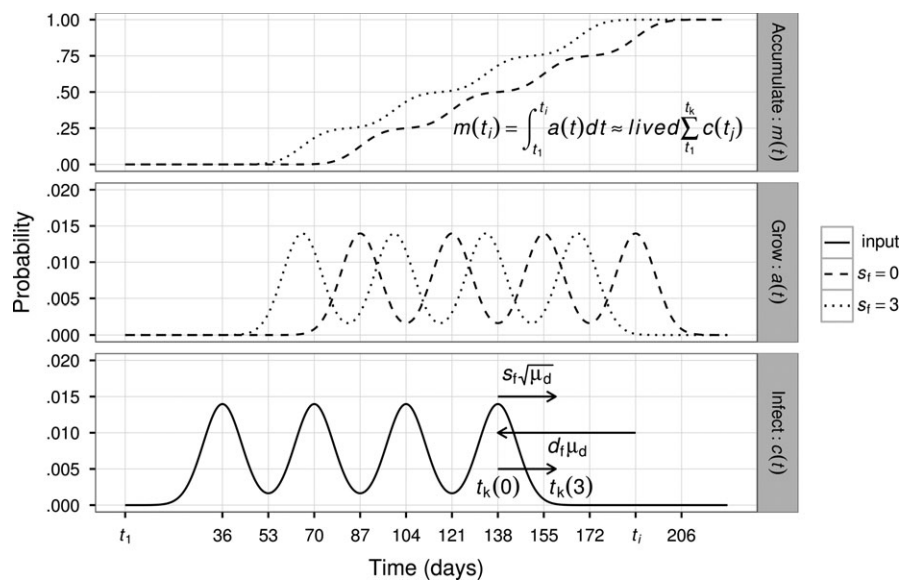


FIGURE 1 The model structure given a simulated infection probability $c(t)$ of chalimus lice (bottom panel) with four peaks each contributing 0.25 (25%) to the total infection (100%). The adult lice peaks (middle panel, dashed curve) arrive one development time later and accumulate (top panel, dashed curve) to the model estimate $m(t_i)$. It is an integration of the arrival probability of the adult lice $a(t)$ between t_1 and t_i that increases to 1 (100%) as the four infection peaks develop into adult lice—assuming all survive. For count data, a sum of chalimus counts $c(t_j)$ between t_1 and $t_k < t_i$ scaled by the survival is used (top panel, formula). The time t_k is set below t_i to account for the mean development time μ_d between t_k and t_i that depends on sea temperature, and $t_k = t_i - d_f \mu_d + s_f \sqrt{\mu_d}$ (bottom panel, annotations) is used to adjust the timing of the lice accumulation ($d_f = 1$, $s_f = 0, 3$ for dashed and dotted curves, respectively)

(location of Nesvík, 2007, Faroe Islands), which are used as model temperatures for all the modelled sites in the Faroe Islands.

The statistical modelling was performed with R/RStudio. The counts of chalimus lice are used to simulate past and current counts of adult lice, and to predict the adult sea lice counts one development time into the future. For each count date t_i in the data series of adult lice $a(t_i)$, we construct the model estimate $m(t_i)$ of adult lice by applying a cumulative summing function to the past chalimus lice $c(t_j)$ (Appendix A).

This is performed by including in the sum all the chalimus lice $c(t_j)$ for the count dates t_j in the interval $[t_1; t_k < t_i]$, where t_1 is the first count date in our series and t_k is the last count date to be included to give all the chalimus lice at t_k enough time to develop into adult lice at t_i (Figure 1).

How far t_k should be set below t_i depends on the daily development times between t_k and t_i that vary with the changing sea temperature. We set $t_k = t_i - (d_f \mu_d - s_f \sqrt{\mu_d})$, where μ_d is the mean development time between t_k and t_i and $\sqrt{\mu_d}$ is the standard deviation. The d_f factor may be used to adjust the development times if they are inaccurately estimated and the s_f factor is the number of standard deviations used to account for the variability of the development times. Application of a single parameter may also account for the variability (Appendix B).

Finally, a fitted survival factor is applied to fit the model estimates to the past adult lice data. The possible presence of *C. elongatus* in the chalimus stage of the sea lice counts affects the survival factor, which we therefore interpret with caution. This caution is especially relevant when large levels of mobile and adult stages of

C. elongatus are observed in the sea lice counts indicating that *C. elongatus* may distort the chalimus counts. In these cases, we can still apply our chalimus-based model, but the survival factor of chalimus will tend to be underestimated.

2.3 | Modelling from mobiles

The mobile stages of *L. salmonis* and *C. elongatus* are distinguishable. Therefore, to avoid the survival-factor issue of the model based on chalimus, a modelling from mobiles has also been implemented. The modelling from mobiles is analogous to the modelling from chalimus. The main difference being that a shorter development time is applied that also leads to a shorter prediction interval into the future.

2.4 | Model parameters

The main modelling parameters are as follows: the sea lice data file, the sea lice model type (chalimus or mobiles) and the sea lice natural survival (Table 1). Optionally, we can modify the development and standard deviation factors from their default values ($d_f = 1$, and $s_f = 3$). The scaling factor related to natural survival was fitted to obtain the lowest errors for a given set of the development and standard deviation factors. As the chalimus counts include both sexes—and may include distortion from *C. elongatus*—while the adult counts only represent adult females of salmon sea lice, the scaling factor of 0.5 corresponds roughly to a survival of 100%.

Optionally, we can simulate sea lice treatments by selecting the timing (data point) of treatment and simulate the treatment effect

TABLE 1 The model parameters

Parameters	Description	Values	Default
<i>data</i>	Sea lice count data	Excel files	NA
<i>model</i>	Sea lice model type	Chalimus or mobiles	Chalimus
<i>lived</i>	Proportion transformed to adult females	0–1 (expected <0.5)	1
d_f	Development factor (option)	0–1.5	1
s_f	St. deviation factor (option)	–3 to 3	3
<i>treat</i>	Data point of treatment (option)	Data index	0 (off)
<i>effect</i>	Proportion removed (if treatment)	0–1	0 (off)

between 0 and 1 (0% and 100%). When applied, the treatment simulates that a percentage of the adult sea lice are removed corresponding to the treatment effect, while the chalimus (or mobile) lice are assumed to be more or less unaffected by the treatment. The sea lice simulation is assumed applicable nearly independently of the treatment method, as most treatments are not very effective against the younger lice stages. However, if a treatment also removes the younger lice stages, then it is not simulated in the current version of the model.

The remaining input parameters of the user interface (Appendix A) serve as options to filter the input data set, to switch the prediction into the future on or off, or selecting which tables to output.

In the end, the performance of the model is demonstrated by simulations of 6- to 12-month-long time series of adult sea lice counts for different farming sites and years in the Faroe Islands.

2.5 | Time-series analysis

For independent checks of the development times that we used in the statistical modelling, a time-series analysis was applied to the count data from different sites used for salmon aquaculture in the Faroe Islands (Appendix C).

The results to be presented later will suggest that the mean time differences at lag –3 can be used to infer the mean development time from chalimus to adult lice.

The time-series methodology was initially only used on the full length of the given data series of lice counts at each site. However, time series for different sites usually cover different periods of time, and are therefore not completely comparable as they cover different seasonal temperature regimes.

An approach to better compare cross-correlation plots across sites is therefore to restrict the time series to the same time period, for example, a whole year. This should probe similar temperature regimes at each site, and we use this approach to estimate the mean development time during a complete year.

Another approach is to generate multiple subsets of the time series to study different periods of the year, and hence probe slightly different temperature regimes (Appendix C).

The lagged correlations between chalimus and adult lice provide estimates of the development times between the lice stages, which we compare with the corresponding development times inferred from the EWOS growth curves.

The time-series analysis of the chalimus and adult lice are also used to derive the three parameters ($\alpha_1, \alpha_2, \alpha_3$) in the temperature-dependence formula used by Stien, Bjørn, Heuch, and Elston (2005): $D(T) = \alpha_1(T + \alpha_2)^{\alpha_3}$ —using the previously described subsets of time series—each probing a slightly different temperature regime. Data of mean time differences at lags versus the corresponding mean temperatures are plotted and the resulting data used to find the parameters (Appendix D).

The chalimus lice develop into adult lice in the mean development time. Therefore, as usual in both infectious disease and parasite modelling (Rittenhouse et al., 2016; Vynnycky & White, 2011), the inverted development time (1/Development) is interpreted as a mean growth rate (arrival rate) of lice/day, and we show how this growth rate changes with the mean sea temperature. Further, we study the size of the cross-correlations in two plots, one versus temperature and one versus the inverted development time.

3 | RESULTS

The main results of this study are presented in two parts: the first part (Figures 2–6) presents the development times inferred from EWOS data (Figure 2) and the modelling of the adult lice from the

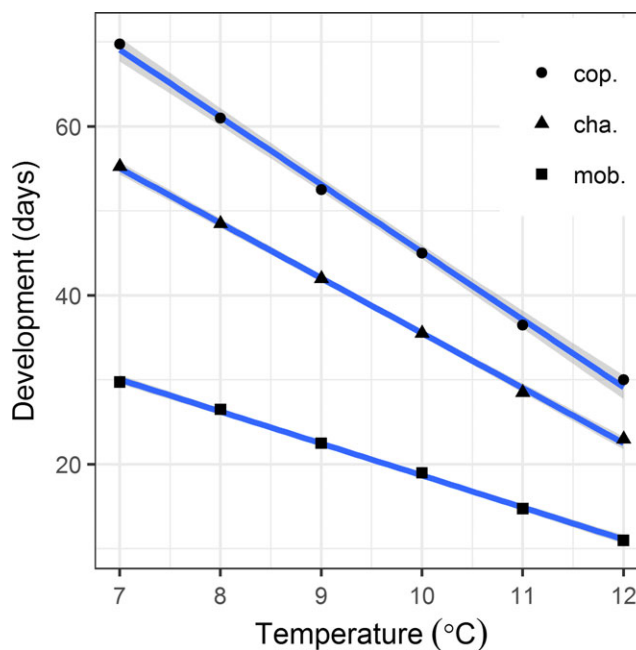


FIGURE 2 The development time (black points) to adult females inferred from EWOS growth curves similar to Myhr et al. (2009) and the linear regression (blue lines) for different sea temperatures from the lice stages copepod, chalimus and mobiles: $D_{\text{cop}}(T) = 125.1 - 7.993T$, $D_{\text{cha}}(T) = 100.6 - 6.507T$ and $D_{\text{mob}}(T) = 56.55 - 3.786T$ (all three $R^2 \approx 1$ and slope p -values = $1e-6, 2e-7, 5e-7$) [Colour figure can be viewed at wileyonlinelibrary.com]

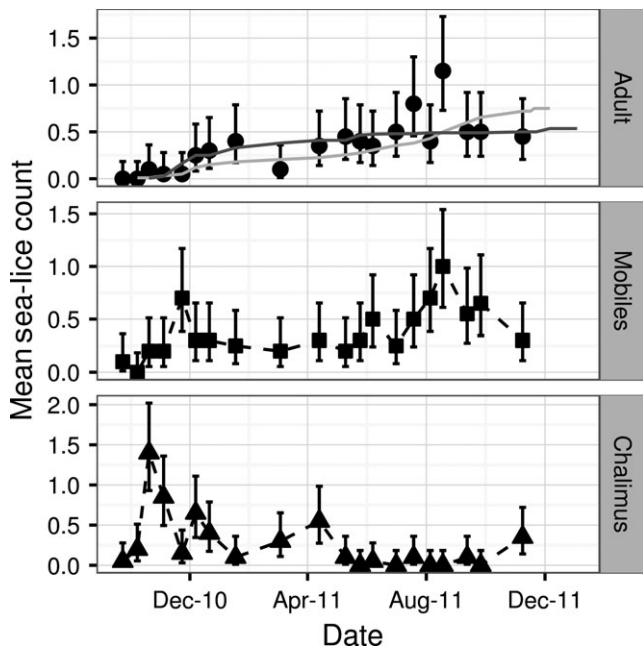


FIGURE 3 The sea-lice counts (black points) of chalimus, mobiles and adult lice and the modelling of adult lice (dark and light grey lines) from chalimus and mobile lice, respectively, for site 1, 2010–11, $n = 20$ fish in each count. For both models, the development and standard deviation factors are 1 and 3, respectively, and the natural survival factor is 0.1

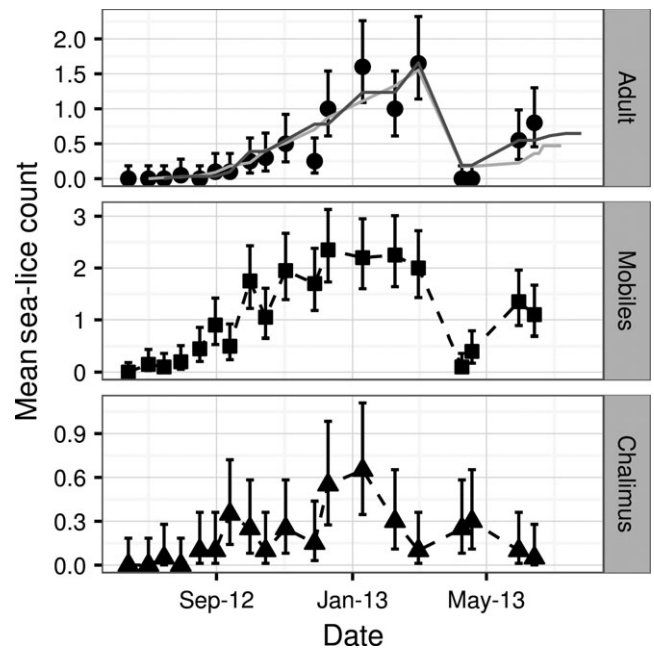


FIGURE 5 The sea-lice counts and the modelling of adult lice from chalimus and mobile lice for site 1, 2012–13, $n = 20$ fish in each count. The development and standard deviation factors are 1 and 3 for both models, and the survival factor is 0.65 and 0.1 for the chalimus and mobile models, respectively. A treatment is simulated at 9 April 2013 (data point 16) with a treatment effect of 0.90

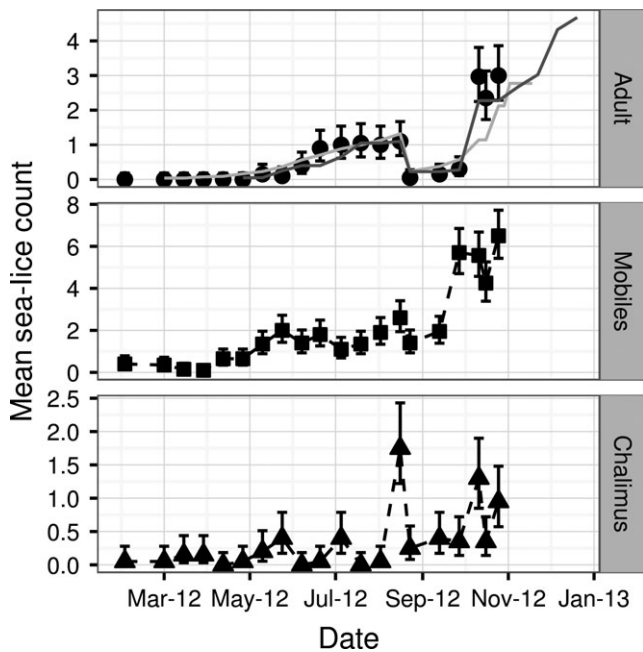


FIGURE 4 The sea-lice counts and the modelling of adult lice from chalimus and mobile lice for site 2, 2012, $n = 20$ fish in each count. The development factor is 1 for both models, and the standard deviation factors are -2 and 3 and the survival factor is 1 and 0.1 for the chalimus and mobile models, respectively. A treatment is simulated at 23 August 2012 (data point 15) with a treatment effect of 0.85

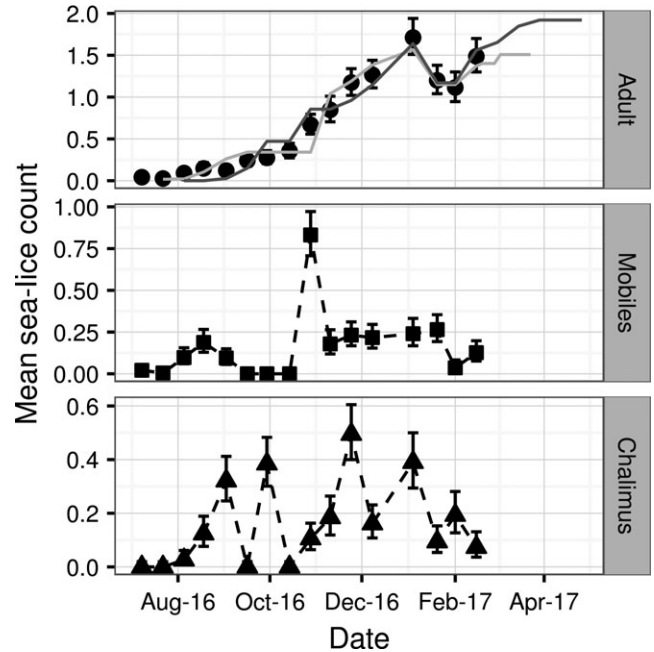


FIGURE 6 The sea-lice counts and the modelling of adult lice from chalimus and mobile lice for site 1, 2016–17, $n = 140$ –190 fish in each count. The development and standard deviation factors are 1 and 3 for both models, and the survival factor is 1 and 0.84 for the chalimus and mobile models, respectively. A treatment is simulated at 20 January 2017 (data point 14) with a treatment effect of 0.35

attached chalimus and mobile stages (Figures 3–6), and the second part (Figures 7–12) presents the time-series analysis. Finally, for the purpose of the discussion, one extra plot (Figure 13) is presented, which shows a nonlinear relation between the adult female counts and the number of released sea lice eggs fish⁻¹ day⁻¹.

3.1 | Development times used in the modelling

The development times obtained from EWOS data for the three lice stages, copepod, chalimus and mobiles, decrease linearly with the sea temperature (Figure 2). As the modelling is made from the attached stages of chalimus and mobiles, the development times used are $D_{cha}(T) = 100.6 - 6.507T$ for the modelling from chalimus lice and $D_{mob}(T) = 56.55 - 3.786T$ for the modelling from mobile lice.

The modelling plots to be presented show three panels of the sea lice count data: one for each lice stage of chalimus, mobiles and adult lice versus count time together with the modelling (solid lines) shown in the adult panel (Figures 3–6) for the obtained model parameters (Table 2).

3.2 | Modelling from chalimus

The count data for site 1 (2010–11), site 2 (2012) and site 1 (2012–13) are typical for this time periods, in which typically only a sample of 20 fish were examined from each cage, and this leads to relatively large error bars on the data (Figures 3–5) as compared to the increased number of fish (140–190) in each counting (Figure 6). Another common characteristic is the behaviour of the chalimus counts, which jump up and down from the 0-count baseline of no signal, while the adult counts tend to increase (accumulate) with time, and do not come down again unless a treatment is performed.

Site 1 (2010–11) shows a strong chalimus signal in November 2010, with smaller signals between December 2010 and April 2011 (Figure 3). Around December 2010, or roughly 1 month after the first chalimus signal, the adult lice signal starts to grow. In this case, however, the growth in the adult sea lice counts tends to level off at about 0.5 lice/fish, because the chalimus signal is very low from May 2011.

The results of the modelling from the chalimus to the adult stage are shown as dark grey lines in the adult panel of the plots, which in

general capture the tendency of the data within the accuracy of the counting's and the model. The prediction effect of the model is illustrated by the lines that extend beyond the last count data due to the development time from the last chalimus counts into the future (Figures 3–6).

An example of modelling from chalimus for site 2, 2012, shows an increase in the adult lice until about 1 lice/fish, where they suddenly drop down to about 0 in late August 2012, before rising again to a higher level of about 2.5–3 lice/fish around one and a half month later. A treatment is simulated at 23 August 2012, with a treatment effect of 0.85 (85%) for removing the adult lice, assuming no effect on the chalimus lice (Figure 4).

This behaviour is characteristic: that the adult signal sometimes shows a rapid increase soon after a treatment to similar or even higher values than before the treatment. This can mainly be explained from the chalimus count, which shows a large signal (1.75 lice/fish) about the same time as the treatment is performed (Figure 4). This chalimus signal contributes to the large increase in both the mobile and the adult counts shortly after treatment, because the treatments applied are less effective against the chalimus lice, which can grow into new mobiles and adults in a time corresponding to the development times (Figure 2). In this case, the mobile lice already present at the time of treatment are possibly also contributing (Figure 4), as they seem largely unaffected by this particular treatment and are virtually free to grow into adults.

Another example from site 1, 2012–13, show the adult lice increase from low levels in mid-June to mid-September 2012, up to 1.65 lice/fish on 1 March 2013. On 9 April 2013, the adult lice have decreased to 0, followed by a new increase to 0.55 lice/fish on 30 May and 0.8 on 13 June 2013. A treatment is simulated at 9 April 2013, with a treatment effect of 0.90 (Figure 5).

Here, we also see an increase of the adult count shortly after treatment, but to a lower level than before the treatment. Again, this can mainly be explained by the chalimus counts, which are at moderate levels at the point of treatment (Figure 5). This leads to a more moderate increase in the adult lice counts shortly after treatment (as compared to Figure 4). This particular treatment seems to remove both the adult lice and most of the mobile lice. Consequently, mainly chalimus lice at the time of treatment are left to survive into new mobiles and adults.

TABLE 2 Parameter values for Figures 3–6

Site	From	To	Figures	<i>n</i>	Model	Lived	<i>d_f</i>	<i>s_f</i>	Treat	Effect	Date
1	2010-09-23	2011-11-08	3	20	Chalimus	0.10	1	3	0	0.00	NA
1	2010-09-23	2011-11-08	3	20	Mobiles	0.10	1	3	0	0.00	NA
2	2012-02-02	2012-10-25	4	20	Chalimus	1.00	1	-2	15	0.85	2012-08-23
2	2012-02-02	2012-10-25	4	20	Mobiles	0.10	1	3	15	0.85	2012-08-23
1	2012-06-14	2013-06-13	5	20	Chalimus	0.65	1	3	16	0.90	2013-04-09
1	2012-06-14	2013-06-13	5	20	Mobiles	0.10	1	3	16	0.90	2013-04-09
1	2016-07-08	2017-02-15	6	140–190	Chalimus	1.00	1	3	14	0.35	2017-01-20
1	2016-07-08	2017-02-15	6	140–190	Mobiles	0.84	1	3	14	0.35	2017-01-20

The development and standard deviation factors (Table 2, d_t , s_f) were at their default values 1 and 3 (Figures 3, 5 and 6). The development and standard deviation factors were usually kept at 1 and 3, respectively, during modelling, but the standard deviation factor of -2 was used for the chalimus model in one case to fit the timing of the large increase in adult counts shortly after treatment (Figure 4). This corresponds to a slower development than expected and may also be modelled by setting the standard deviation factor to 0 and using a development factor of about 1.28 (Appendix B).

Remember that the scaling factor of 0.5 is assumed to correspond roughly to a survival of 100% for adult females, because the scaling is applied to a sum of chalimus counts that include both sexes, while the adult counts only include females. For the presented examples of chalimus modelling, we found scaling factors of 0.1, 0.65 and 1 (Table 2, lived), and in other examples (not shown) most scaling factors were between 0.6 and 1. The lowest scaling factor of 0.1 may correspond to about 20% survival for adult females if we neglect any distortion of the chalimus counts. The scaling factors exceeding 0.5 up to 1 may indicate that the number of chalimus lice have been underestimated by up to 50%, as they are more difficult to find than the larger adult females.

In any case, the scaling factors related to survival should be interpreted with great caution. Besides the model assumptions, the difficulties in counting the small chalimus lice and the possible presence of *C. elongatus* in the chalimus stage may affect the scaling factor.

3.3 | Modelling from mobiles

The modelling based on mobiles is also implemented (Figures 3–6, light grey lines), which gives similar fits although the modelling from chalimus (Figures 3–6, dark grey lines) is slightly better fitted to the data.

An example of lice counts for site 1, 2016–17, has much smaller error bars compared to the previous plots (Figures 3–5) because of the larger number of fish in each counting (Figure 6). The adult counts increase up to about 1.75 lice/fish at which point a treatment is applied. The treatment does not seem to affect neither the chalimus nor the mobile lice, and only moderately the adult lice. This may indicate resistance to the treatment (Figure 6). The treatment is simulated at 20 January 2017, with a treatment effect of 0.35. Again, the effect of the treatment is short-lived, and about one and a half month after the treatment the adult sea lice level is back at the sea lice limit of 1.5 lice/fish. Here, both the chalimus and the mobile lice are at moderate levels at the time of treatment, but combined with the ineffective treatment, these levels are high enough to restore the adult count back to its level before treatment in about one development time.

The modelled values from mobiles are generally in good agreement with the adult counts as observed from the plots (Figures 3–6). A summary of the errors between the $n = 15$ pairs of the adult counts and the modelled values results in: -0.33 (*Min.*), -0.06 (1^{st} *Qu.*), -0.009 (*Median*), -0.007 (*Mean*), 0.08 (3^{rd} *Qu.*) and 0.19 (*Max*) lice/fish; and the standard deviation was 0.13 (Figure 6). These errors are smaller than errors obtained for Figures 3–5, but similar to the errors for the

corresponding modelling from chalimus for this same site and the same data (Figure 6). Therefore, the better performance is likely due to the better data quality from the increased number of fish (140–190) in each counting (Figure 6), as compared to only 20 (Figures 3–5).

The errors are fairly symmetrically distributed about mean 0 with the standard deviation of 0.13, and the maximum numerical error was 0.33 lice/fish (Figure 6). Assuming a nearly normal distribution of errors about mean 0, all errors should be within ± 3 *SD*, which in this case is about ± 0.4 lice/fish. This error margin is the maximum prediction errors of the model for this case, while 95% prediction intervals for the errors are about ± 0.3 lice/fish.

For the presented examples of modelling from mobiles, we found scaling factors of 0.1 and 0.84 (Table 2, lived). In one case, the scaling factor of 0.1 is the same as for the modelling from chalimus (Figure 3), while in two cases the scaling factor of 0.1 from mobiles is much lower as compared to 1 and 0.65 from chalimus (Figures 4–5). This suggests that the survival of chalimus is overestimated for Figures 4–5, probably due to difficulties in counting the small chalimus lice. In one case, the scaling factor of 0.84 from mobiles is above 0.5 and slightly lower than 1 from chalimus (Figure 6), indicating distorted scaling factors that should be treated with caution. The common scaling factor of 0.1 from mobiles (Figures 3–5) may indicate a survival of about 20% between the mobile and adult female stages in these cases.

A test to predict the sea-lice levels on this (and another) site used by a fish farm was performed on 5 January 2017. Only having the sea lice data until 8 December 2016, for the site, the model predicted the lice levels in early January to be 1.5 lice/fish (Figure 6). The fish farm reported back that they just made a new counting of 1.7 lice/fish on 4 January 2017, confirming the prediction within an error of 0.2 lice/fish. For the other site (not shown), by only having the sea-lice data until 12 December 2016, for that site, the model predicted the lice levels in early January to be 0.75 lice/fish. Here, the company reported back that they made a new counting of 0.57 lice/fish on 4 January 2017, also confirming the prediction within an error of 0.2 lice/fish. The purpose of this test was primarily to estimate the difference in survival at these two sites, in which the later site with the counting of 0.57 lice/fish in early January used cleaner fish in most sea cages in addition to thermalizing treatments, while these treatments were much less used in the former site (Figure 6). The results from the fitted scaling factor of about 0.84 and 0.1 suggest a very different survival for these sites. As the modelling in this test was performed from mobiles, the scaling factor should ideally be less inflated by *C. elongatus* and by the difficulties in counting the small chalimus lice. However, the scaling factor above 0.5 indicates some distortion for this case (Figure 6).

3.4 | Time-series results

We now turn to the second part of the results regarding the time-series analysis (Figures 7–12). The mean time differences versus mean temperature between lagged time series of chalimus lice and adult lice for four sites in Faroe Islands, 2009–11, are compared with the EWOS-inferred development times from chalimus to adult.

Multiple subsets of the time series are used to find the mean time differences at different mean temperatures. The time differences for lags -4 to 0 with cross-correlations (*ccf*) between chalimus and adult lice (*ccf*-lag mapped to point colour, *ccf*-value mapped to point size) suggest that the EWOS-development times $D_{cha}(T)$ are similar to the mean time differences at lag -3 (Figure 7).

This plot indicates variability in both the mean time differences and in the *ccf*-values with temperature. For site A, the only significant correlations are found at lag -3 at the lowest temperatures down to about 7.8°C . With increasing temperature, other correlations are observed for site A: first at lag -3 and -2 , and then for lag -3 , -2 , -1 and 0 at the temperature of 9.5°C . Site B is

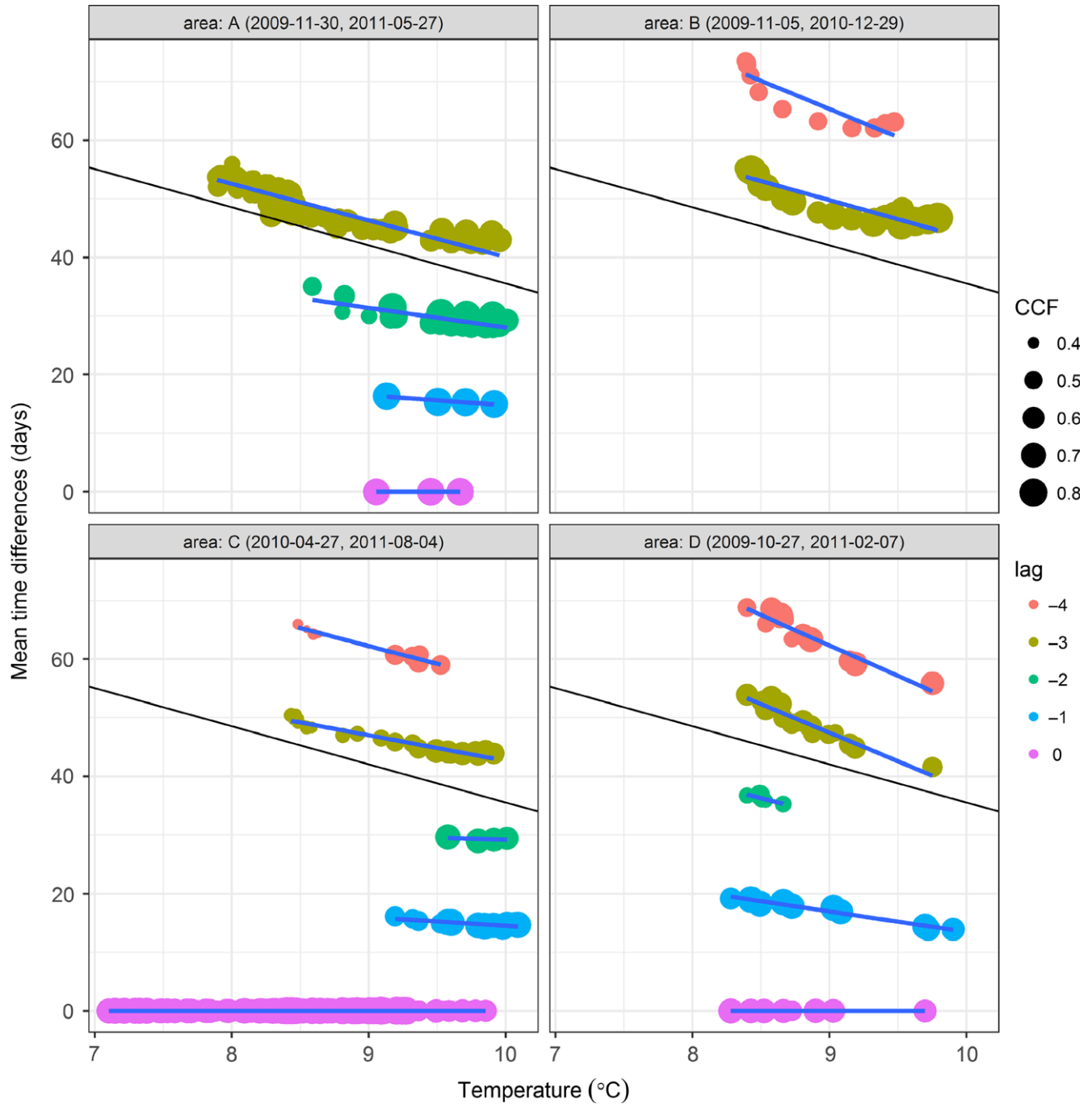


FIGURE 7 Mean time differences versus mean temperature between lagged time series of chalimus lice as compared to the time series of adult lice, for four sites in Faroe Islands, 2009–11. The plotted time differences are for lags -4 to 0 with significant cross-correlations between chalimus and adult lice (*ccf*-lag mapped to point colour, *ccf*-value mapped to point size). Multiple subsets of the time series are used to find the mean time differences at different mean temperatures. The temperature dependence is indicated by best-fit linear regression (blue lines). For comparison, the chalimus to adult development time inferred from EWOS is also shown (black line), which are similar to the mean time differences at lag -3 [Colour figure can be viewed at wileyonlinelibrary.com]

different, as the only significant correlations for this site are at lag -3 and -4 . Site C is mainly dominated by correlations at lag 0 for the lowest temperatures down to about 7.2°C , while at about 8.5°C additional correlations are at lag -4 and -3 , and at 9.5°C they are observed at lags -4 , -3 , -2 , -1 and 0. Also, most of these correlations seem to increase with temperature, although the correlation at lag 0 seems to be smallest at the highest temperatures larger than 9.3°C . Finally, site D shows significant correlations at lag -1 and 0 at the lowest temperature about 8.2°C , and at lag -4 , -3 , -2 , -1 and 0 for 8.5°C . The correlation at lag -2 disappears at higher temperatures, while the others remain.

In general, the *ccf* data are complex and challenging to interpret. Sometimes, significant correlations are observed for positive lags, either only at positive lags, or in addition to the correlations at negative lags. Occasionally, when looking at the evolution of the correlations with time, the first significant correlations are observed at positive lags, then later at negative lags, to finally be observed at both negative and positive lags.

Statistics (counts) of the correlations at the four sites in Faroe Islands, 2009–11, describe the number of correlations for both positive and negative lags. Cross-correlation at large lags are typically hard to detect and the largest lags we observed were found at -7 and 6 for site C and A, respectively (Figure 8). For lag < 0 , the largest count of correlations is found for lag -3 for all four sites (Figure 8). For site A and D, the number of correlations is maximum for lag -3 . For site B, the largest counts are relatively similar at lag -3 and lag 4, which has the maximum count. Site C has the largest number of correlations at lag 0, followed by the correlations at lag -3 .

Considering densities of counts larger than 5%, the main part of the correlations are found at lags -3 , -2 , 3, and 4 (site A), -4 , -3 , 1 and 4 (site B), -3 , -1 , 0 and 3 (site C), -4 , -3 , -2 , -1 , 0 and 2 (site D). Hence, the number of lags with count densities larger than 5% is 4 for the sites A–C and six for site D. Considering all the counts (Figure 8), the number of lags with correlations is ten (site A), four (site B), eight (site C) and seven (site D).

The counts and number of lags with significant correlations (Figure 8) is not fully understood. However, one possible suggestion is that the correlations at different lags are caused by correlations of the adult lice counts with series of subpopulations of chalimus lice generated by the 10–11 egg strings laid by an adult female in its lifetime of about 191 days (Appendix C; Figure 12).

The temperature-dependent development time (Figure 9a) inferred from the mean time difference at lag -3 in the lagged correlations between chalimus and adult lice is modelled by the formula (Appendix D):

$$D(T) = \alpha_1(T + \alpha_2)^{\alpha_3} = 17,840(T + 7.439)^{-2.128}.$$

$D(T)$ tries to capture the curvature of the temperature dependence. However, the curvature of the data seems a bit stronger than predicted by this relation. To compare with the linear formula derived from EWOS data, the first-order Taylor polynomial $D_1(T)$ was

derived. At the mean temperature 8.5°C in the Faroe Islands, $D(T)$ is approximated by:

$$D_1(T) = 105.2 - 6.578T.$$

$D_1(T)$ predicts a mean development time of 49 days, which is similar to the 45 days predicted by $D_{\text{cha}}(T) = 100.6 - 6.507T$ from EWOS data. Note that the slope of $D_1(T)$ and $D_{\text{cha}}(T)$ are almost identical, so the difference of about 4 days between these two predictions stays constant.

We also present the inverted development times as the rate $1/\text{Development}$ (Figure 9b). In this case, only a smoothed curve is plotted, instead of also inverting the $D(T)$, $D_1(T)$ and $D_{\text{cha}}(T)$ relations. This plot illustrates how the arrival rate increases with temperature: initially a fast increase between about 8 and 8.75°C followed by a change in curvature to a more moderate increase with temperature as the temperature approaches 10°C . The mean temperatures in the Faroe Islands do not go much beyond 10°C , so the arrival rate of nearly 0.024 lice/day at 10°C may be regarded as the maximum rate. The mean yearly rate of about 0.021 lice/day is found at 8.5°C . These rates are in good agreement with population growth rates obtained from the slope of log-linear plots of salmon lice over time (Patursson et al., 2017).

The size of the cross-correlations between chalimus and adult lice at site 2, Faroe Islands, 2009–11, show linear associations between *ccf*-value and temperature (Figure 10a; slope p -value = $3\text{e-}22$, $R^2 = 0.72$), and between *ccf*-value and the arrival rate $1/\text{Development}$ (Figure 10b; slope p -value = $1\text{e-}17$, $R^2 = 0.62$). In this plot, the significance limit for the *ccf*-values was raised a bit by subtracting three *na*-values in the differences at lag -3 from the denominator of $\pm ciline \approx \pm 1.96/\sqrt{n}$ (Appendix C).

Therefore, 77 points are included in this analysis (Figure 10) as opposed to using all the 96 significant points, if not subtracting the three *na*-values. This change was not required to establish the linear associations—it only raises R^2 . The linear association between the *ccf*-values and temperature, and between the *ccf*-values and arrival rate is not surprising, but confirms our trust in these correlations. Apparently significant, but spurious cross-correlations are known to happen in time-series analysis, for example, if both time series are autocorrelated. However, it seems unlikely to find such clear and biologically meaningful associations if we were tricked by spurious correlations.

The cross-correlations at four sites over a whole year (2010–11) suggest very similar development times: 49, 50, 51 and 52 days, respectively, inferred from the mean time differences at lag -3 (Figure 11). This is because the mean sea temperature is similar between Faroese fjords in the same year at about 8.5°C . Also, the *ccf*-values at lag -3 are similar at about 0.5–0.6 for site A, B and D as expected from Figure 10a for temperatures of 8.5°C , while the *ccf*-value at lag -3 is somewhat lower (about 0.4) for site C. Remember that the observed variability in Figure 10 was obtained after raising the confidence limit for significant correlations. Some

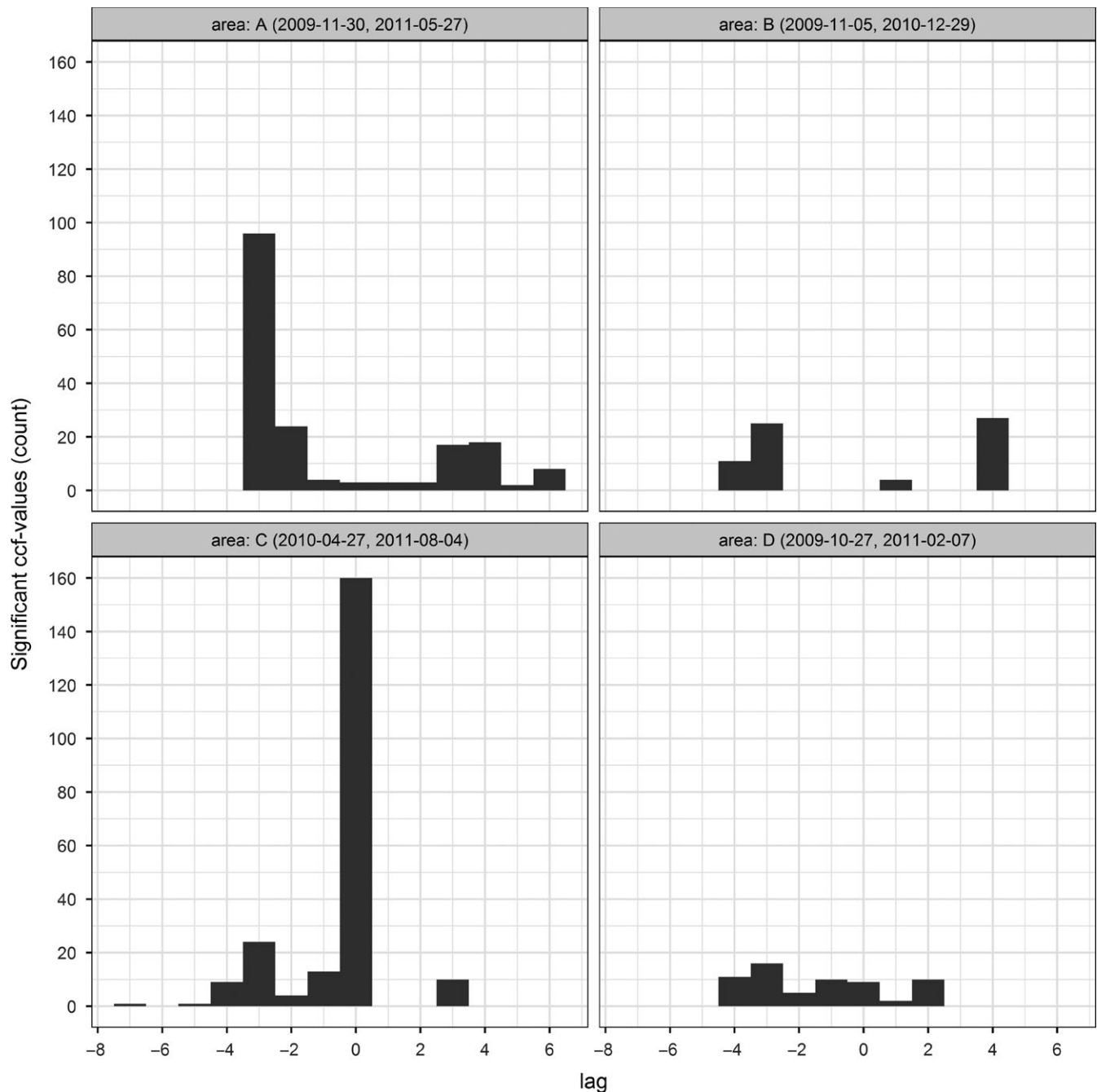


FIGURE 8 Statistics (counts) of the significant lagged correlations for the multiple subsets of the time series used to find the mean time differences at different mean temperatures at four sites in the Faroe Islands, 2009–11. In this plot, both negative and positive lags are included. Significant correlations are observed for both positive and negative lags. For lag < 0, the largest count of correlations is observed for lag -3

larger variability in the *ccf*-values is therefore expected for such low correlations as observed at lag -3 for site C at 8.5°C.

These cross-correlation plots were not selected to present the strongest correlations, which occur at higher temperatures, but they were meant to present cross-correlation plots for a complete year (Figure 11). At higher temperatures, the *ccf*-value at lag -3 for site C increased to above 0.6 and was similar in size to the correlation at lag 0 (plot not shown). Similarly, the cross-correlations at the other sites increase with temperature, as expected from Figure 10a. At

higher temperatures, the correlations just below the significance limit tend to increase above the limit, for example, so we observe cross-correlations at both positive and negative lags for site A and D.

3.5 | Reproduction

Stochastic simulation modelling of the adult female levels based on Poisson distributions suggests a nonlinear relation between the adult female level and the number of adult females that

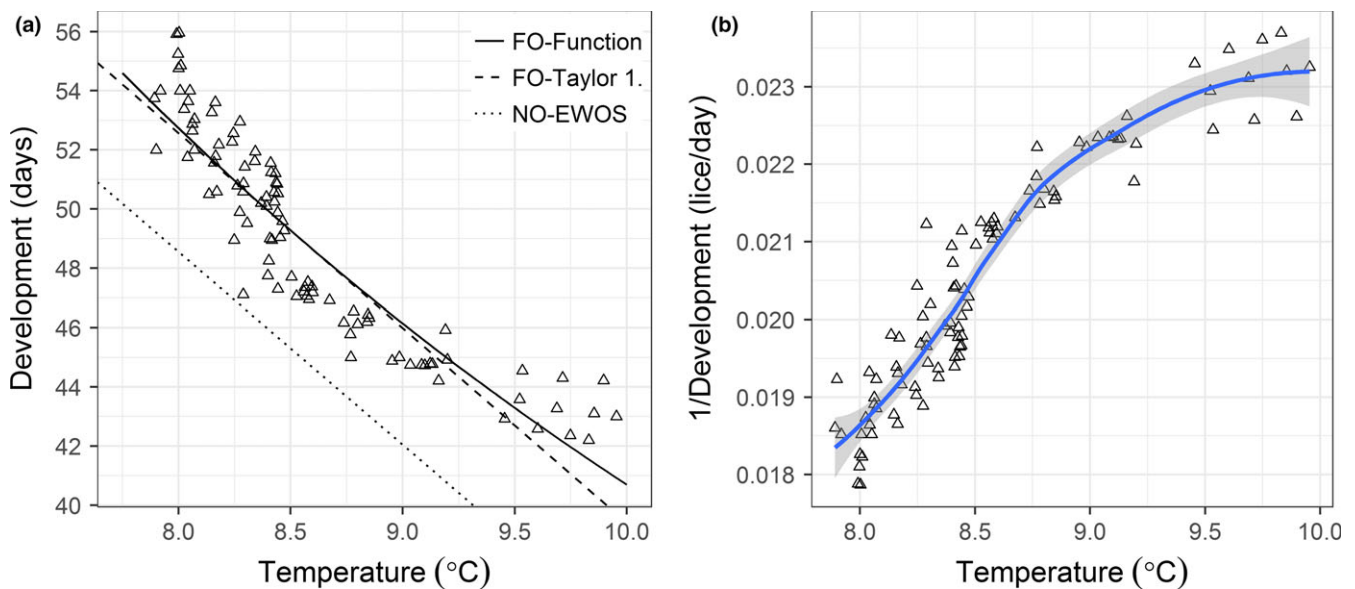


FIGURE 9 Development times (a) and inverted development times $1/\text{Development}$ (b) for chalmus to adult females (open points) versus temperature for site 2, Faroe Islands, 2009–11. The in situ development time is inferred from the mean time difference at lag -3 in the lagged correlations (*ccf*) between chalmus and adult lice. The three lines in (a) are as follows: (1) the modelled temperature dependence for this site (solid line: $D(T) = \alpha_1(T + \alpha_2)^{\alpha_3} = 17,840(T + 7.439)^{-2.128}$), (2) the first-order Taylor polynomial approximation at the mean temperature (dashed line: $D_1(T) = 105.2 - 6.578T$) and (3) the temperature dependence inferred from EWOS (dotted line: $D_{\text{cha}}(T) = 100.6 - 6.507T$). A linear regression for plot (b) slope p -value = $1e-30$, $R^2 = 0.86$, is not shown, as the relationship seems curved and better fitted with the smoothed (blue) curve [Colour figure can be viewed at wileyonlinelibrary.com]

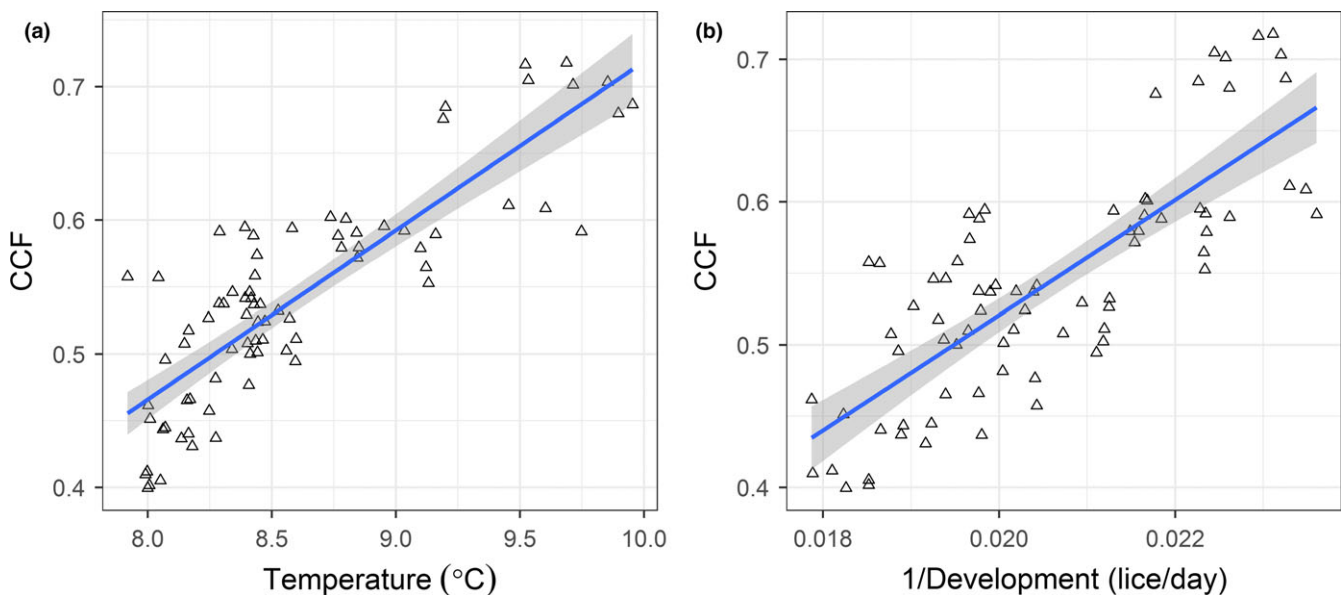


FIGURE 10 Cross-correlations between chalmus and adult lice at site 2, Faroe Islands, 2009–11. Plot (a) and (b) show the significant *ccf*-values at lag -3 versus mean sea temperature and versus $1/\text{Development}$, respectively, for which development in (b) are the mean time differences at lag -3 . Linear regression lines (blue) show significant linear relationships: (a) slope p -value = $3e-22$, $R^2 = 0.72$ and (b) slope p -value = $1e-17$, $R^2 = 0.62$ [Colour figure can be viewed at wileyonlinelibrary.com]

reproduce; for example, for the average lice count of 0.5 adult females per fish, 40%–60% of adult females reproduce, while at 1.5 adult females per fish, 80%–90% of the females reproduce (Stormoen, Skjerve, & Aunsmo, 2013). A nonlinear relation between the adult female level and the number of released sea

lice eggs per fish per day is therefore presented (Figure 13) for use in the Discussion section.

Recall that under laboratory conditions, the adult female can live up to 191 days and produce up to 11 pairs of egg strings each containing a mean number of 285 eggs at 7.2°C (Heuch et al., 2000),

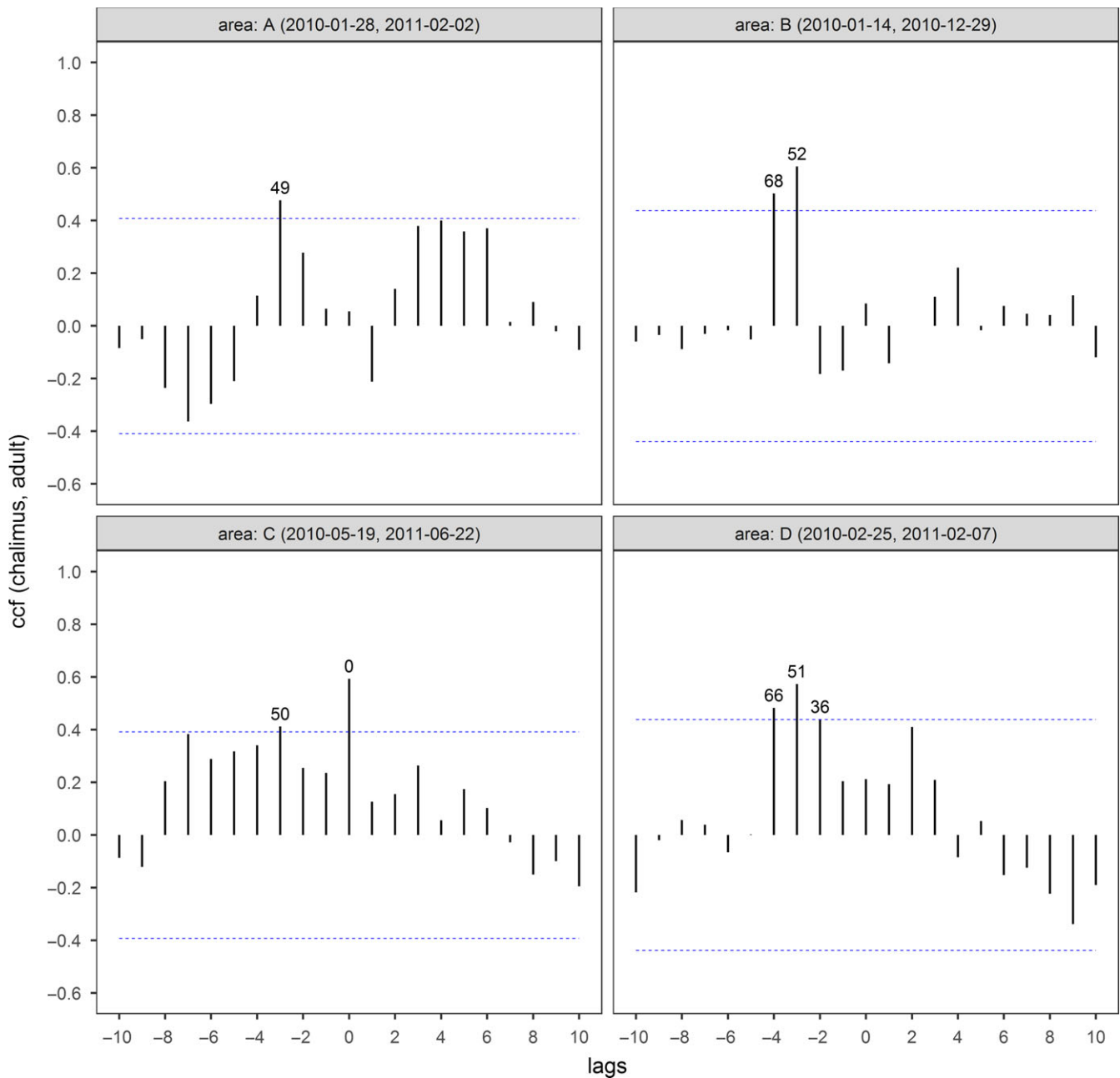


FIGURE 11 Lagged correlations between time series of chalimus lice as compared to the time series of adult lice for four sites in Faroe Islands. The labelled mean time differences are for lags with significant correlations (*ccf*). This plot covers a whole year (2010–11) to find the mean development time for a full year, which is inferred from the mean time differences of 49–52 days at lag -3 [Colour figure can be viewed at wileyonlinelibrary.com]

thus producing up to $2 \times 285 \times 11 = 6,270$ eggs, that is, on average $6270/191$ (eggs adult female $^{-1}$ day $^{-1}$).

Given the middle reproduction curve by Stormoen et al. (2013) and denoting it $r(a_f)$, it is asymptotically limited by $r(a_f) \lesssim 1$ for increasing a_f (Figure 13a, concave curve). Denoting the lice level a_f (adult female/fish) in the Faroe Islands, each salmon fish releases on average: $S_e(a_f) = (6270/191)r(a_f)a_f \approx 32.8r(a_f)a_f$ (eggs fish $^{-1}$ day $^{-1}$) at 7.2°C. S_e ranges between 0 and about 60 (eggs fish $^{-1}$ day $^{-1}$) for the adult lice levels between 0 and 2 (Figure 13b); for example, for the lice counts of $a_f = 0.1, 0.25, 0.5, 0.75, 1.0, 1.5$ and 2.0 (adult female/

fish), each salmon fish releases on average $S_e(a_f) = 0.33, 2.1, 8.2, 15, 24, 42$ and 59 (eggs fish $^{-1}$ day $^{-1}$), respectively.

Consider the gain term $r(a_f)a_f$, which is asymptotically limited by $r(a_f)a_f \lesssim a_f$ for large a_f and approximated by $r(a_f)a_f \approx a_f^2$ for small a_f (Figure 13a, convex curve). If $r(a_f)a_f > 1$, it amplifies the number of released eggs per fish per day, and if $r(a_f)a_f < 1$ the output is reduced, as compared to the unamplified number of 32.8 eggs fish $^{-1}$ day $^{-1}$ for a reproductive adult female at 7.2°C ($r(a_f)a_f = 1$). The adult female lice count a_f should always be kept small (<0.5 adult females/fish) to let the quadratic term $r(a_f)a_f \approx a_f^2$

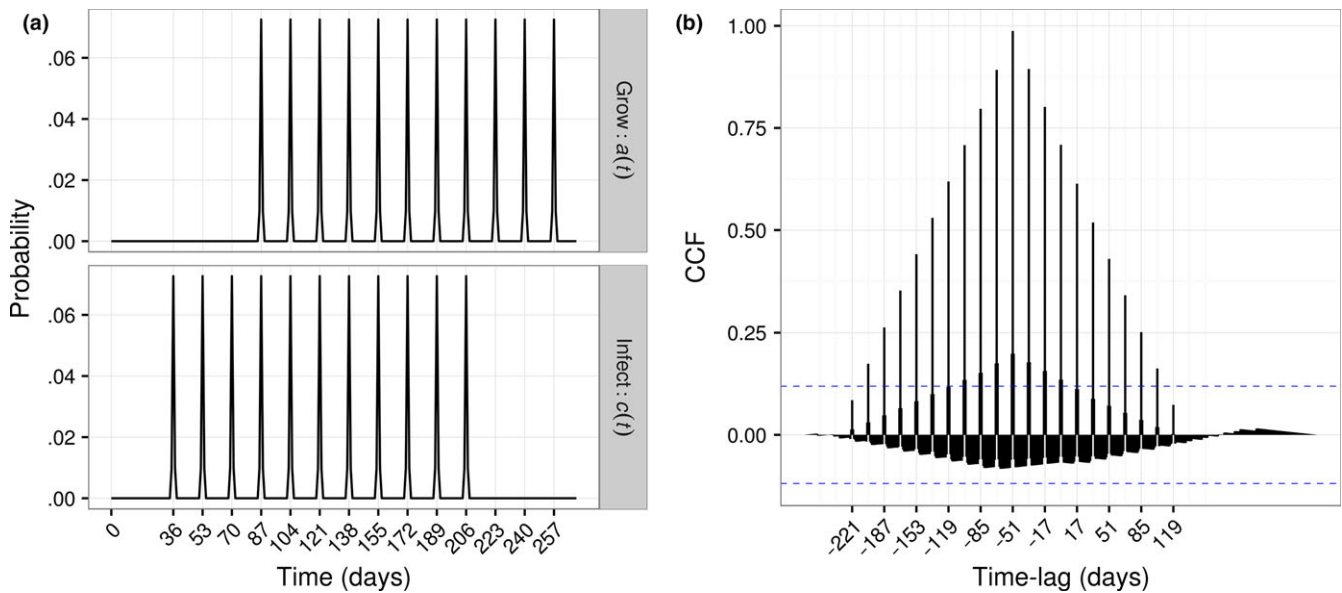


FIGURE 12 (a) Simulated infection with 11 chalimus subpopulations beginning at day 36 and spaced in time by 17 days (bottom panel)—as an adult female can live up to 191 days and produce up to 11 pairs of egg strings that may lead to chalimus infections spaced on average by 17 days at 7.2°C. These grow into adults in 51 days in this simulation with the first arriving at day 87 (upper panel). The standard deviation of the peaks was small (0.5 days) to get distinct peaks. (b) Lagged correlations between the simulated data. They are centred around -51 days that shows the strongest correlation as all the 11 adult peaks correlate with all the 11 chalimus peaks 51 days back in time. This corresponds to a lag of -3 in units of the subpopulation spacing (17 days). The lowest correlations are found at -221 (36 - 257) and 119 (206 - 87) days as in plot (a) only one peak at 257 can correlate with the one at 36 for lag -13, and only one peak at 87 can correlate with the one at 206 for lag 7. The number of peaks that may correlate increases in steps of 1 for each lag from lag -13 (1) to lag -3 (11) and then decreases by 1 for each lag down to lag 7 (1) [Colour figure can be viewed at wileyonlinelibrary.com]

for small a_f reduce $S_e(a_f)$ much below 32.8 eggs fish⁻¹ day⁻¹; for example, for $a_f = 0.4, 0.2$ and 0.1 , $S_e(a_f)$ is reduced to 16%, 4% and 1% of 32.8 eggs⁻¹ fish⁻¹ day⁻¹, respectively.

The mean number F_e of eggs/day released by a farm with n number of fish is similarly estimated by: $F_e(a_f, n) = nS_e(a_f) = 32.8 nr(a_f)a_f$ eggs/day at 7.2°C. Note that F_e contains the co-amplifying factors $nr(a_f)a_f$; for example, for a fish farm with a million fish, F_e will range between 0 and about 60 million eggs/day for adult lice levels between 0 and 2 (adult female/fish).

4 | DISCUSSION

The sea lice parasite on farmed salmon is one of the major unsolved challenges to the aquaculture industry. The high sea lice levels lead to expensive treatments and an increased handling inflicting stress on the fish. A decreased growth and increased losses of fish may be observed. In the worst-case scenario, the stressed fish may become less resistant to outbreaks of various fish diseases. Environmental issues are also at stake. The potential spread of sea lice to wild salmon, sea trout and other marine species and the increased use of chemotherapeutants are two of the main concerns. The recent use of cleaner fish combined with various forms of water treatments is better for the environment. However, these treatments are costly and may still stress the fish.

The aim of this study was to develop a statistical model of the sea lice data in order to understand and simulate the adult sea lice

levels, and the ineffective treatments previously used against sea lice. A better understanding could help us reduce future sea lice levels. The focus was on modelling data sets from sea lice surveys at Faroese salmon farms. A time-series analysis was applied to study lagged correlations in the count data to infer the sea lice dynamics on salmon farms, for example, the temperature dependence of such important characteristics as the sea lice development times and the arrival rates.

A statistical model of the sea lice data was developed that reproduces the general trends in past sea lice counts and predicts future sea lice levels 1–2 months into the future (Figures 3–6). The predictions into the future were based on the sea lice development times from chalimus to adult, which depend on the sea temperature (Figure 2). The model describes how lice infections accumulate from the young lice into adult lice, and how the adult lice counts tend to grow until lice treatments are applied; reducing the sea lice level (Figures 4–6). However, the effect of the treatments is typically short-lived, and within 1–2 months, the sea lice return, sometimes even to higher levels than before the treatment (Figure 4). This is mainly explained by high chalimus counts at the time of treatment. The chalimus lice are less affected by the treatment and grow into new adults in one development time.

The modelling from chalimus is sometimes affected by *C. elongatus* not being distinguished from the attached chalimus stage of salmon sea lice in the count data. The relative prevalence of *L. salmonis* and *C. elongatus* in the attached stages of sea lice is not well known.

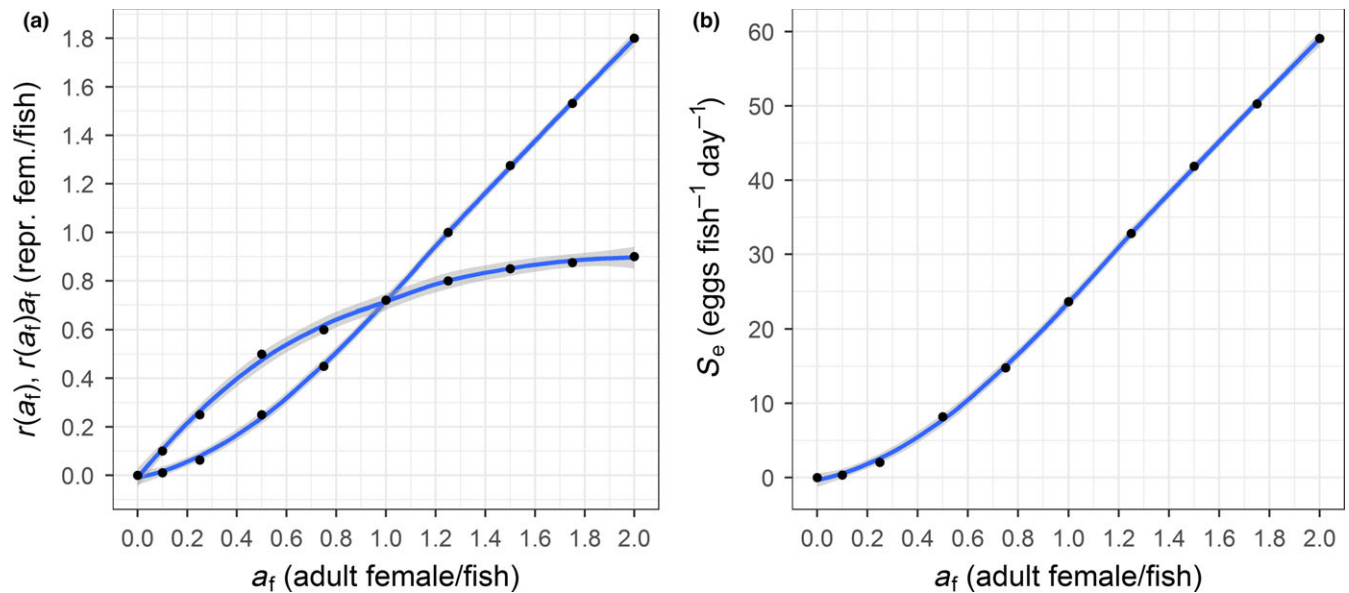


FIGURE 13 (a) The reproduction $r(a_f) \approx a_f$ for small values of a_f and asymptotically limited by $r(a_f) \lesssim 1$ for increasing a_f (concave curve, reproduced from Stormoen et al. (2013)), and the gain term $r(a_f)a_f \approx a_f^2$ for small a_f , which is asymptotically limited by $r(a_f) a_f \lesssim a_f$ for large a_f (convex curve). If $r(a_f)a_f > 1$, this turns up (amplifies) the number of released eggs per fish per day, and if $r(a_f)a_f < 1$ this turns the number down, as compared to the unamplified number of 32.8 eggs fish⁻¹ day⁻¹ for a reproductive adult female at 7.2°C. (b) The number $S_e(a_f) \approx 32.8r(a_f)a_f$ of released lice eggs per fish per day [Colour figure can be viewed at wileyonlinelibrary.com]

The prevalence of gravid *C. elongatus* in the Faroe Islands was observed to be seasonal with low counts (0–0.5) in June–August, moderate (0.5–1) in May and September and mostly higher (1–4) in the remaining months January–April and October–December (á Norði et al., 2015).

In the present study, the counts of adult *C. elongatus* (both sexes) was mostly low to moderate (<0.5) with maximum counts (about 1) in October and December for site 1, 2010–11, after 4–6 months at sea, respectively. The situation for site 1, 2012–13, was similar with maximum counts (0.5–1) in October to March after 4–9 months at sea. Similarly for site 2, 2012, the levels of *C. elongatus* were mostly between 0 and 0.5 except in August (about 1) and October (1–2) after 6–10 months at sea, respectively. For site 1, 2016–17, the levels of adult *C. elongatus* were high (2–4) in September to January after 2–6 months at sea and lower (<1) for the remaining months.

As these levels of *C. elongatus* may distort the chalimus counts, a modelling based on mobiles was implemented (Figures 3–6). The model errors are fairly symmetrically distributed errors about mean 0 with a standard deviation of 0.13, and the maximum numerical error 0.33 lice/fish (Figure 6). The 95% prediction intervals for these errors are about ± 0.3 lice/fish, and the maximum prediction errors was estimated to ± 0.4 lice/fish.

A test performed for two farming sites to investigate sea lice levels and survival factors at these sites predicted the adult sea lice counts within an error of 0.2 lice/fish. The fitted scaling factor was very different for these two sites using very different treatments, indicating that the model may be useful both for predicting sea lice levels and for investigating the effects of different treatments.

Simulating not only the general trends in sea lice counts, but also the short-lived effect of the treatments (Figures 4–6) is an important part of this study. The simulation of the treatments reproduced the quick return of high sea-lice levels; explained by the low effect of treatments on the young lice combined with their short development times to adult.

Non-optimal timing of treatments could contribute to high sea lice levels as well as the increased resistances to chemotherapeutants that are suspected by the industry. A recently updated sea-lice regulation for the Faroe Islands states that if suspected or confirmed cases of resistance are observed the treatment has to be repeated with a different method. However, no information about suspected nor confirmed cases of resistance was available for this study. The sea lice count data and the use of chemotherapeutants for each company and farming sites in the Faroe Islands were not treated as public information, and therefore, the farming sites and companies were anonymized in this study. In addition to the mentioned update, this will change so that future information regarding sea lice will be public from 1 October 2017.

Traditionally, the treatments with chemotherapeutants against sea lice in the Faroe Islands tend to be administered when the adult sea lice counts reach the permitted limit of mean 1.5 (previously 2.0) adult females/fish. At these high sea-lice levels, however, the lice epidemic is very difficult to control. The lice counts for chalimus (and mobiles) have typically also reached high levels at this point. Consequently, treatments primarily targeting adult lice will be ineffective at reducing the overall population.

The second part of this study dealt with time-series analysis of cross-correlations in the count data (Figures 7–12). The purpose was

to check the development times used in the modelling of sea lice levels derived from EWOS (Figure 2). The results of the time-series analysis—when compared between sites—depended on the different time periods covered by the count data, which probe different temperature regimes. Therefore, a systematic study of cross-correlations for subsets of the time series from four sites was applied to probe the different temperature regimes. This revealed a close similarity between the EWOS-development times and the lagged correlations at lag -3 (Figure 7). Statistics of all the significant lagged correlations revealed that lag -3 had the highest count for all four sites for lags <0 , but also revealed various differences between the lag statistics (Figure 8).

This complexity is cautiously attributed to cross-correlations between different populations of adult and chalimus lice leading to the following relation for which lags should show correlations: $[-3 - g; 7 - g]$ for a given adult generation, A_g , $g = 0 \dots 10$, produced by self-infection of 11 chalimus subpopulations. This is lag -13 to 7 for all correlations between 11 chalimus and adult subpopulations and the correlation at lag -3 in units of the subpopulation gap corresponds to the development time—illustrated by a simulation using 51 days for the development time and 17 days for the population gap (Figure 12). Different distributions of lags may be expected depending on which subpopulations of the chalimus and adult lice dominate the lice counts. A continuous distribution of lags could indicate strong self-infection dynamics (Figure 8, sites A and D), while gaps in lags (Figure 8, sites B and C) could indicate missing subpopulations (e.g., removed copepod populations by currents) and less strong self-infection dynamics.

For the site covering the largest temperature range of cross-correlations, the temperature-dependent development time was modelled by the formula: $D(T) = \alpha_1(T + \alpha_2)^{\alpha_3} = 17,840(T + 7.439)^{-2.128}$ (Figure 9a). From $D(T)$, the first-order Taylor polynomial $D_1(T)$ was derived at the mean temperature 8.5°C in the Faroe Islands: $D_1(T) = 105.2 - 6.578T$. $D_1(T)$ predicts a mean development time of 49 days, as compared to the 45 days predicted by $D_{\text{cha}}(T) = 100.6 - 6.507T$ from EWOS data. The slope of $D_1(T)$ and $D_{\text{cha}}(T)$ are almost identical, keeping the difference of 4 days between these two linear predictions constant. For the studied temperature range of $7.8\text{--}10^\circ\text{C}$, there is not much difference in using the linear approximation $D_1(T)$ as opposed to using the nonlinear $D(T)$ formula, which seems less curved than the data (Figure 9a).

The inverted development times correspond to the arrival rate: $1/\text{Development}$ (Figure 9b). The mean rate over a year of about 0.021 lice/day is found at 8.5°C , which is the mean sea temperature over a year. The maximum rate of nearly 0.024 lice/day is found at 10°C , which is about the highest obtainable mean temperature for the time series. These rates are in good agreement with population growth rates in the Faroe Islands derived by another method (Patursson et al., 2017).

The size of the cross-correlations showed linear associations between *ccf*-value and temperature (Figure 10a), and between *ccf*-value and the rate $1/\text{Development}$ (Figure 10b). These are biologically meaningful associations as we expect shorter development

times and higher arrival rates at the highest temperatures (Figure 9a, b), which both should contribute to stronger correlations with increased temperature.

Finally, time series were compared for similar time periods for a complete year at four sites (Figure 11). This probes the mean development time over a year in the Faroe Islands, which was found to be very similar: 49, 50, 51 and 52 days at these four sites, as expected from the comparable temperatures at the sites in the same year. Also, side-band gaps of about 15–16 days are observed in this study (Figure 11) that possibly confirm the simulated correlations between subpopulations (Figure 12). The size of the cross-correlations was relatively similar as expected (from Figure 10a).

We conclude that (1) the similarity of the derived development times as compared to the EWOS-development times, (2) the similarity of the derived arrival rates as compared to population growth rates (Patursson et al., 2017), (3) the biologically meaningful linear associations for the size of the cross-correlations (Figure 10a,b) and finally (4) the similarity of cross-correlations compared over a complete year (Figure 11) all increase the reliability of these correlations.

The time-series data were checked for autocorrelations, as spurious cross-correlations are known to happen in time-series analysis if both time series are autocorrelated. In most cases, this was found not to be the case. The exception was site C (Figures 7 and 8), for which some of the subsets of the time series showed autocorrelation in both time series. However, as the results for site C seemed roughly similar to the other sites, for which autocorrelation was mainly found in the adult series and not for the chalimus series, the results for site C are also presented and estimated to be valid.

With the exception of using the development times derived from EWOS data, the study was limited to what could be inferred and modelled from the sea lice data. Any detailed information about sea-lice treatments and the number of fish at each site was not contained within the data.

The data from the lice surveys are less controlled and less well described as compared to laboratory studies, and this may to some extent complicate the conclusions derived. On the other hand, the count data at salmon farms reflect conditions that may be difficult, if not impossible to replicate under laboratory conditions. The data are not gathered by the industry itself, but by Fiskaaling—the Aquaculture Research Station of the Faroes. The data are not affected by industrial interests, but the quality may to some extent vary due to employees with different experiences in sea-lice counting.

For future control of sea-lice levels in the Faroe Islands, it must be recognized that using a high sea-lice limit (1.5 adult females/fish) is an ineffective approach, because at these levels of adult females, the reproduction of offspring is likely to be highly effective as estimated by the reproduction ratio: $r(1.5) \approx 0.85$ (Figure 13a); the number of reproductive adult females per fish given by the gain factor: $r(1.5)1.5 \approx 1.28 > 1$ (Figure 13a); the number of released eggs per fish: $S_e(1.5) = 32.8r(1.5)1.5 \approx 42$ eggs fish $^{-1}$ day $^{-1}$ (Figure 13b); and the number of released eggs by a fish farm: $F_e(1.5, n) = nS_e(1.5) \approx 42n$ eggs/day. This can lead both to a high

self-infection at farms and to ocean-current spreading of the lice infection to other farms; for example, for 1 million fish in a fish farm, about 42 million eggs will be released per day.

The adult female lice count a_f should always be kept below 0.5 adult females/fish to take advantage of the quadratic term $r(a_f)a_f \approx a_f^2$ for small a_f in order to reduce $S_e(a_f)$ considerably below $S_e(1) \approx 33$ eggs fish⁻¹ day⁻¹—and not amplify it by 1.28 reproductive adult females per fish to $S_e(1.5) \approx 42$ eggs⁻¹ fish⁻¹ day⁻¹; for example, for $a_f = 0.4, 0.2$ and 0.1 , $S_e(a_f)$ is reduced by 0.16, 0.04 and 0.01 reproductive adult females per fish to 5.2 (16%), 1.3 (4%) and 0.33 (1%) eggs⁻¹ fish⁻¹ day⁻¹, respectively, as compared to $S_e(1) \approx 33$ eggs fish⁻¹ day⁻¹ for one reproductive adult female per fish.

The reproduction discussion and the quadratic dependence at low densities is inspired by the modelling of lice reproduction for farmed salmon with lice abundance between 0 and 2 adult females/fish assuming that lice will stay with the same host in nature as opposed to under laboratory conditions where lice may transfer between hosts (Stormoen et al., 2013). Considering model limitations this can to some extent distort the conclusions derived; for example, if mobile males in fish farms can leave their hosts to seek out females, this possibly leads to a higher reproduction at low densities. At higher densities the lice fecundity—but not the lice survival—is found to be reduced (Ugelvik, Skorping, & Mennerat, 2017). However, this effect is most important at high parasite loads, above 2–4 female lice/fish, and before reaching these levels in fish farms, the various treatments reduce the densities. Consequently, for the densities mostly below 2 females/fish in our study, we do not expect much reduced fecundities with the lice loads.

Rather than only focusing on the lice limit a_f of the adult female, a better approach to be considered for future lice protocols are setting limits for the younger lice stages, because when adult females are present, the situation quickly gets out of hand. Another approach is also to limit the number of fish n at each farm in order to reduce the number of sea lice released by the farm: $F_e(a_f, n) = nS_e(a_f)$.

Presently no limits are set for the attached or mobile stages of lice in the Faroe Islands. High numbers of the younger lice stages are warnings of high lice infection pressure, which are visible 1–2 months before the adult lice levels start to rise. In the future, sea-lice limits should also be based on the planktonic stages of sea lice to account both for self-infection and for infection from other farms.

To account for the planktonic stages may not be practically feasible at the moment, but it should be when protocols for sea-lice counting of the planktonic stages are better developed and automated. Improvements and automation of sea-lice counting of the attached stages will improve the data quality to be used in the modelling of the adult sea lice.

In conclusion, the adult sea-lice counts were modelled from the attached stages of chalimus and mobile lice. It illustrated how the chalimus infections developed into the accumulation of adult lice. The ineffective treatments could be simulated and explained. Predictions 1–2 months into the future were made. In a test for two farming sites the model predicted the adult sea lice counts within an

error of 0.2 adult females/fish, which was within the estimated 95% prediction intervals of the errors (± 0.3) and the estimated maximum prediction errors (± 0.4) for this test.

Time-series analysis on the sea-lice count data was used to check the sea-lice development times and to probe the sea lice dynamics in situ. From cross-correlations at lag -3 between the chalimus and adult lice time series, the temperature-dependent sea-lice development time was derived and inverted to find the arrival rate (growth rate) of the adult lice from chalimus lice. The size of the cross-correlations showed linear associations with both temperature and the arrival rate. The different distributions of the correlations for different lags seem consistent with correlations between subpopulations of chalimus lice and adult lice. We suggest that correlations at several lags with a continuous distribution indicate strong self-infection dynamics, while correlations at few lags and gaps in lags indicate missing subpopulations and less strong self-infection dynamics.

Finally, we estimated the sea lice production from fish farms to discuss and advise on approaches to control the sea-lice epidemics—preferably by natural means. We recommend lower sea-lice limits (< 0.5 adult females/fish), and urge in particular to set limits for the younger lice stages. As the number of fish in Faroese farms sometimes is larger than one million fish, a lice limit closer to 0 than to 0.5 adult females/fish is recommended together with setting limits on the younger lice stages and on the numbers of fish at each farm.

Achieving such low lice levels may seem unrealistic when compared to the relatively high sea lice levels in the Faroe Islands. One of the arguments against setting very low lice limits was the potential impact on the industry, the fish and the environment from the expected increased use of chemotherapeutants and other treatments to keep the lice counts below a very low limit. Instead, a limit of 1.5 adult female/fish was set but enforcing this limit more strictly than the old limit (2.0 adult female/fish); for example, if a farm has more than 1.5 adult female/fish in three consecutive lice counts, it may be required to harvest the fish within 2 months.

The new lice limit is a move in the right direction although probably an inadequate one, because controlling the sea lice epidemic by accepting a limit that allows good reproductive conditions for the sea lice seems even more unrealistic than enforcing a very low lice limit that may break the reproductive cycle. Although fish in the worst affected farms are harvested, they have already generated large amounts of new sea lice larvae that may infect other farms. Lice limits of 0.3–0.5 adult female/fish are enforced in Norway and although both physical and environmental conditions are different, such levels may also be enforced in the Faroe Islands. The lice situation in the Faroe Islands is likely to improve in the future because the local aquaculture industry is continuously developing new and improved techniques to keep the lice levels well below the imposed limit.

Reflecting on the modelling approach, what we have done is similar to the modelling process described by Groner et al. (2016), as (1) we visualized and described field data of sea lice counts motivated by present phenomena in lice epidemics and we suggested hypothesis of relationships in a statistical model; (2) we used the model on

field data and inferred model parameters from past and present data to investigate our hypothesis and explain relationships; (3) we used model simulation and prediction of future lice counts to project system behaviour—also after interventions with sea-lice treatments; and (4) to some extent, we also performed model analysis by theoretical simulations to investigate and explore the mechanism and the expected patterns explained by the model.

Compared to currently used approaches (Groner et al., 2016; Hamilton-West et al., 2012; Heuch, Revie, & Gettinby, 2003; Revie, Gettinby, Treasurer, & Rae, 2002; Revie, Gettinby, Treasurer, Rae, & Clark, 2002; Revie, Robbins, Gettinby, Kelly, & Treasurer, 2005; Rittenhouse et al., 2016), the emphasis in our approach was mainly on the statistical modelling process using field data and to lesser extent on mathematical derivations and theoretical simulations—although two cases of simulations based on theoretical probabilities of sea lice counts were presented that were meant to illustrate the model principle (Figure 1) and the cross-correlation patterns (Figure 12).

The work of Hamilton-West et al. (2012), Heuch et al. (2003) and Revie, Gettinby, Treasurer, Rae, & Clark (2002) focused on large epidemiological data sets and the seasonal changes in the count data investigating correlations within and between lice stages and the environmental parameters, and characterizing the lice burdens in different geographical regions.

The modelling work by Revie et al. (2005) and Rittenhouse et al. (2016) has some resemblance to the modelling of chalimus and adult lice in our study although the approach is different and based on a deterministic host–parasite compartment model.

Most modelling approaches described by Groner et al. (2016) are (1) either population growth models or host–parasite compartment models that are deterministic and solved by difference or ordinary differential equations, respectively; (2) deterministic advection–diffusion models solved by partial differential equations; (3) age-structured population growth models described by difference equations that can include stochastic effects; (4) various forms of statistical regression models describing random effects; (5) statistical survival analysis based on survival and hazard functions that are correlational; (6) stochastic population growth models, for example, Ricker model; (7) individual-based deterministic or stochastic computer models simulating interactions of individuals within a system; and finally (8) numerically solved ocean-circulation models simulating the environment and spread of sea lice using hydrodynamic equations for transport and diffusion.

Our model may possibly be classified as related to age-structured population models and survival analysis. To our knowledge, no existing approach has modelled and predicted lice counts in fish farms, lice treatments and development times in situ in similar detail as our study, and we have not seen cross-correlation studies applied elsewhere for determination of the sea lice development times or sub-population gaps. From the presented study, we hope the methods developed and analysis derived may help improve the understanding and control of the sea-lice epidemics.

ACKNOWLEDGEMENTS

This work was commissioned by Øystein Patursson and continuously supported by Knud Simonsen, the former and present Head of Research at Fiskaaling, respectively. Fiskaaling supported the work and delivered the sea lice count data that are managed by Kirstin Eliassen. We thank the local aquaculture companies for use of the anonymized data and the quality managers, veterinaries and other personnel for willingness to participate and discuss prediction test and modelling of sea-lice levels at meetings at Fiskaaling. Gunnvør á Norði is acknowledged for information regarding planktonic sea-lice stages. Thanks to Hilmar Simonsen and Sigmundur P. Hannesarson for proofreading of the article.

ORCID

H Gislason  <http://orcid.org/0000-0003-0649-7559>

REFERENCES

- Aldrin, M., Stovrik, B., Kristoffersen, A. B., & Jansen, P. A. (2013). Space-time modelling of the spread of salmon lice between and within Norwegian marine salmon farms. *PLoS One*, 8, e64039.
- Dalvin, S. (2016). Temperaturenens innflytelse på lakseluslarver. *Rapport fra Havforskningen*, 3-2016 (pp. 1–9). Bergen, Norway: Institute of Marine Research. <http://hdl.handle.net/11250/2408806>
- Degroot, M. H., & Schervish, M. J. (2002). *Probability and statistics*. Boston, MA: Addison-Wesley.
- Groner, M. L., Rogers, L. A., Bateman, A. W., Connors, B. M., Frazer, L. N., Godwin, S. C., ... Schlägel, U. E. (2016). Lessons from sea louse and salmon epidemiology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371, 20150203.
- Hamilton-West, C., Arriagada, G., Yatabe, T., Valdes, P., Herve-Claude, L. P., & Urcelay, S. (2012). Epidemiological description of the sea lice (*Caligus rogercresseyi*) situation in southern Chile in August 2007. *Preventive Veterinary Medicine*, 104, 341–345.
- Hamre, L. A., Eichner, C., Caipang, C. M. A., Dalvin, S. T., Bron, J. E., Nilssen, F., ... Skern-Mauritzen, R. (2013). The salmon louse *Lepeophtheirus salmonis* (Copepoda: Caligidae) life cycle has only two chalimus stages. *PLoS One*, 8, e73539.
- Heuch, P. A., Nordhagen, J. R., & Schram, T. A. (2000). Egg production in the salmon louse [*Lepeophtheirus salmonis* (Krøyer)] in relation to origin and water temperature. *Aquaculture Research*, 31, 805–814.
- Heuch, P. A., Revie, C. W., & Gettinby, G. (2003). A comparison of epidemiological patterns of salmon lice, *Lepeophtheirus salmonis*, infections on farmed Atlantic salmon, *Salmo salar* L., in Norway and Scotland. *Journal of Fish Diseases*, 26, 539–551.
- Jansen, P. A., Kristoffersen, A. B., Viljugrein, H., Jimenez, D., Aldrin, M., & Stien, A. (2012). Sea lice as a density-dependent constraint to salmonid farming. *Proceedings. Biological Sciences/The Royal Society*, 279, 2330–2338.
- Kragesteen, T. J. (2016). *Modelling dispersal of salmon lice in a Tidal Energetic Island System: Faroe Islands*. MSc Thesis, pp. 1–93. Technical University of Denmark, Lyngby, Denmark.
- Kristoffersen, A. B., Jimenez, D., Viljugrein, H., Grontvedt, R., Stien, A., & Jansen, P. A. (2014). Large scale modelling of salmon lice (*Lepeophtheirus salmonis*) infection pressure based on lice monitoring data from Norwegian salmonid farms. *Epidemics*, 9, 31–39.
- Liu, Y., & Bjelland, H. V. (2014). Estimating costs of sea lice control strategy in Norway. *Preventive Veterinary Medicine*, 117, 469–477.

- McEwan, G. F., Groner, M. L., Fast, M. D., Gettinby, G., & Revie, C. W. (2015). Using agent-based modelling to predict the role of wild refugia in the evolution of resistance of sea lice to chemotherapeutants. *PLoS One*, *10*, e0139128.
- Myhr, E., Wadsworth, S., Vecino, J., Gonzales, J., Trioncoso, J., & Carr, I. (2009). EWOS integrated sea lice programme – feed as a tool in the management of sea lice. *EWOS Spotlight*, *1*, 1–24.
- Norði, G. Á., Simonsen, K., Danielsen, E., Eliassen, K., Mols-Mortensen, A., Christiansen, D. H., ... Patursson, Ø. (2015). Abundance and distribution of planktonic *Lepeophtheirus salmonis* and *Caligus elongatus* in a fish farming region in the Faroe Islands. *Aquaculture Environment Interactions*, *7*, 15–27.
- Norði, G. Á., Simonsen, K., & Patursson, Ø. (2016). A method of estimating in situ salmon louse nauplii production at fish farms. *Aquaculture Environment Interactions*, *8*, 397–405.
- Patursson, E. J., Simonsen, K., Visser, A. W., & Patursson, Ø. (2017). Effect of exposure on salmon lice *Lepeophtheirus salmonis* population dynamics in Faroese salmon farms. *Aquaculture Environment Interactions*, *9*, 33–43.
- Peacock, J. L., & Peacock, P. J. (2011). *Oxford handbook of medical statistics*. Oxford, UK: Oxford University Press.
- Revie, C. W., Gettinby, G., Treasurer, J. W., & Rae, G. H. (2002). The epidemiology of the sea lice, *Caligus elongatus* Nordmann, in marine aquaculture of Atlantic salmon, *Salmo salar* L., in Scotland. *Journal of Fish Diseases*, *25*, 391–399.
- Revie, C. W., Gettinby, G., Treasurer, J. W., Rae, G. H., & Clark, N. (2002). Temporal, environmental and management factors influencing the epidemiological patterns of sea lice (*Lepeophtheirus salmonis*) infestations on farmed Atlantic salmon (*Salmo salar*) in Scotland. *Pest Management Science*, *58*, 576–584.
- Revie, C. W., Robbins, C., Gettinby, G., Kelly, L., & Treasurer, J. W. (2005). A mathematical model of the growth of sea lice, *Lepeophtheirus salmonis*, populations on farmed Atlantic salmon, *Salmo salar* L., in Scotland and its use in the assessment of treatment strategies. *Journal of Fish Diseases*, *28*, 603–613.
- Rittenhouse, M. A., Revie, C. W., & Hurford, A. (2016). A model for sea lice (*Lepeophtheirus salmonis*) dynamics in a seasonally changing environment. *Epidemics*, *16*, 8–16.
- Rogers, L. A., Peacock, S. J., McKenzie, P., Dedominicis, S., Jones, S. R. M., Chandler, P., ... Krkošek, M. (2013). Modeling parasite dynamics on farmed salmon for precautionary conservation management of wild salmon. *PLoS One*, *8*, e60096.
- Shumway, R. H., & Stoffer, D. S. (2011). *Time series analysis and its applications: With R examples*. New York, NY: Springer.
- Stien, A., Bjørn, P. A., Heuch, P. A., & Elston, D. A. (2005). Population dynamics of salmon lice *Lepeophtheirus salmonis* on Atlantic salmon and sea trout. *Marine Ecology Progress Series*, *290*, 263–275.
- Stormoen, M., Skjerve, E., & Aunsmo, A. (2013). Modelling salmon lice, *Lepeophtheirus salmonis*, reproduction on farmed Atlantic salmon, *Salmo salar* L. *Journal of Fish Diseases*, *36*, 25–33.
- Teetor, P. (2011). *R cookbook*. Beijing, China: O'Reilly.
- Ugelvik, M. S., Skorping, A., & Mennerat, A. (2017). Parasite fecundity decreases with increasing parasite load in the salmon louse *Lepeophtheirus salmonis* infecting Atlantic salmon *Salmo salar*. *Journal of Fish Diseases*, *40*, 671–678.
- Vollset, K. W., Krontveit, R. I., Jansen, P. A., Finstad, B., Barlaup, B. T., Skilbrei, O. T., ... Dohoo, I. (2016). Impacts of parasites on marine survival of Atlantic salmon: A meta-analysis. *Fish and Fisheries*, *17*, 714–730.
- Vynnycky, E., & White, R. G. (2011). *An introduction to infectious disease modelling*. Oxford, UK: Oxford University Press.
- Wickham, H., & Grolemund, G. (2017). *R for data science*. Sebastopol, CA: O'Reilly Media.
- Xie, Y. (2015). *Dynamic documents with R and knitr*. Boca Raton, FL: CRC Press.

How to cite this article: Gislason H. Statistical modelling of sea lice count data from salmon farms in the Faroe Islands. *J Fish Dis*. 2018;41:973–993. <https://doi.org/10.1111/jfd.12742>

APPENDIX A

MODEL CODE, PLOTS AND ERROR BARS

The R function `lm()` was used for the linear regression of the EWOS-development times, and `cumsum()` for the cumulative sum function applied to the mean sea lice counts.

Error bars for the mean counts were constructed as Poisson 95% confidence intervals made with `poisson.exact()` from the R package `exactci`. In this case, the Poisson distribution was used for the confidence intervals rather than the normal approximation, because the mean counts of sea lice are typically small numbers much below 10. Especially for mean counts below 1 there is a clear difference between the normal 95% confidence intervals (symmetric) and the Poisson 95% confidence intervals (asymmetric, right skewed) for low counts.

As previously stated, we construct the model estimate $m(t_i)$ for the adult lice $a(t_i)$ by a cumulative sum of the past chalimus lice $c(t_j)$ for the counting dates t_j in the interval $[t_1; t_k < t_i]$.

To estimate confidence intervals as error bars for the model value $m(t_i)$, we consider k chalimus distributions of counts $c_1(t_1), \dots, c_k(t_k)$ used to predict $m(t_i)$ at t_i . We assume these chalimus distributions to be independent and Poisson distributed with mean $\mu_1(t_1), \dots, \mu_k(t_k)$. Then, from probability theory, the sum: $c_1(t_1) + \dots + c_k(t_k)$ has a Poisson distribution with mean $\mu_1(t_1) + \dots + \mu_k(t_k)$ (Degroot & Schervish, 2002). It follows that we can construct Poisson 95% confidence intervals for the model estimate $m(t_i)$ like we do for the mean counts of adult lice. To avoid cluttering of plots, however, we do not plot the error bars of the model, but only for the lice counts, because in most cases the model values and error bars (not shown) are similar to the counts and error bars for the adult lice.

We note that the assumption of independent chalimus distributions may not strictly hold in all cases, for example, for very short counting intervals (<14 days) in which multiple counts may be taken from the same distribution, or when different chalimus distributions are correlated. This can sometimes happen for sites with strong self-infection dynamics that generate and self-infect with multiple series of chalimus distributions in one salmon-production cycle.

The modelling was performed with R/RStudio and plots made with the `ggplot2` package. The R code to find $t_k < t_i$ for each t_i and all the other modelling code from data import to model calculations and plotting were implemented in RStudio following the principles of reproducible research. This was performed using R Markdown documents that intervene marked up text (R Markdown or LaTeX),

R code and output to be knitted to HTML, pdf or Word formats (Xie, 2015). At the time of writing, the modelling code has been further developed into a parameterized R Markdown report with a user interface to set the model input parameters.

APPENDIX B

THE INHERENT VARIABILITY OF THE DEVELOPMENT TIMES

The purpose of using $t_k < t_i$ is to give all the chalimus lice at count date t_k enough time to develop into adult lice at count date t_i . Initially, we used $t_k = t_i - \mu_d$, where μ_d is the mean development time from t_k to t_i . However, in some cases this tends to underestimate the speed of the sea lice development. This is most likely because about half of the sea lice in the chalimus distribution at each t_k can develop faster than the mean development time μ_d and should also be included in the cumulative sum.

Therefore, we set $t_k = t_i - (d_f \mu_d - s_f s_d)$, where d_f is a factor that may be used to adjust the development times if they are inaccurately estimated from the linear regression formula, and/or from using inaccurate model temperatures. The s_f factor is inspired by the normal distribution, because nearly 100% of normal data are within ± 3 SD (s_d) from the mean (μ_d). Assuming that the development times are correctly estimated ($d_f = 1$) and that they are normally distributed around μ_d , the subtraction of three standard deviations: $t_k = t_i - (\mu_d - 3s_d)$ should ensure that all the chalimus lice—which have enough time to develop into adult lice—are included in the cumulative sum. This procedure moves t_k closer to t_i for each t_i and thereby allows more chalimus data in the model prediction $m(t_i)$ for the adult lice.

To find s_d , when we in fact have no data on the inherent variability of the development times, we use the standard deviation formula: $s_d = \sqrt{\mu_d}$, justified by the normal approximation to the Poisson distribution, which in practice works when the mean is greater than 10 (Peacock & Peacock, 2011). This is valid for μ_d as the mean development times are larger than 10 days.

We note that if we in one approach keep d_f fixed at $d_f = 1$ and only change s_f in discrete steps, this is equivalent to, in another approach, to only use a discrete scaling factor of $d_f = 1 - (s_f / \sqrt{\mu_d})$ as seen from solving: $\mu_{\text{eff}} = d_f \mu_d = \mu_d - s_f \sqrt{\mu_d}$, where we introduced an effective development μ_{eff} adjusted for variability.

That gives d_f : 1, 0.86, 0.72 or 0.58 for s_f : 0, 1, 2 or 3 and $\mu_d = 51$. Similarly, using $d_f = 1$ and negative s_f : -1, -2 or -3 in one approach for $\mu_d = 51$ would correspond to only using d_f : 1.14, 1.28 or 1.42 in the other approach. For the modelling from mobiles, $d_f = 1 - (s_f / \sqrt{\mu_d}) = 1 - 0.2 s_f$ for $\mu_d = 25$.

In the modelling, we either kept $d_f = 1$ fixed and changed s_f in discrete steps (-3 to 3), also allowing negative values to account for slower development than expected, or we set $s_f = 0$ and changed d_f continuously. The introduction of the standard deviation factor could

have been omitted as it is not strictly necessary, but its use may give a better intuition for the variability of the development.

APPENDIX C

THE TIME-SERIES CODE AND METHODS

For a data set *lice* from a site containing the count dates *date* and the lice counts *chalimus* and *adult*, the time-series objects *tschalimus* and *tsadult* are created in R/RStudio by `xts(lice$chalimus, lice$date)` and `xts(lice$adult, lice$date)` (Teetor, 2011). Then, the cross-correlation object *ccfobj* is created by `ccf(drop(tschalimus), drop(tsadult), plot = FALSE)`. Here, *drop* is used to get rid of the matrix dimension of the *xts* objects, and `plot = FALSE` to delay plotting. The *ccfobj* contains a list of the lagged correlations *ccfobj\$acf* computed at both negative and positive values of *ccfobj\$lag*, and the number of data points used is *ccfobj\$n.used*. The 95% confidence limits $\pm ciline$ for significant correlations are based on the normal distribution theory confidence limits $\pm \frac{1.96}{\sqrt{n}}$ (Shumway & Stoffer, 2011) for filtering a signal from white noise around mean 0: $\pm ciline \approx \pm 1.96 / \sqrt{ccfobj\$n.used}$.

For a given pair of time series at one site, *tschalimus* and *tsadult*, with the corresponding cross-correlation object *ccfobj*, we loop over the *ccfobj* object and compute the lag and lead time differences in days at each lag: `time_diff = lice$date - lag(lice$date, abs(ccfobj$lag))`, and `time_diff = lice$date - lead(lice$date, ccfobj$lag)` for negative and positive values of *ccfobj\$lag*, respectively. Negative or positive values of *ccfobj\$lag* mean, respectively, that the time series *tschalimus* is lagged or leads the time series *tsadult*.

After removing the text attribute “Time differences in days” from the *time_diff* object using `unclass(time_diff)`, a *t* test (*t.test*) is applied using `broom::tidy(t.test(time_diff))`. The `broom::tidy` call serves to produce a nicely formatted table output (one row) from the *t* test, while the purpose of using this test here is to compute the mean time difference together with the standard output, which includes the confidence limits for the mean.

When the cross-correlations are finally plotted, the mean time difference is rounded to whole days and used to label only those lags for which significant correlations exists. For this purpose, we save the output of the *t* test for each lag together with the cross-correlation object *ccfobj* and the cross-correlation confidence limit *ciline*.

Using R code, we generate all possible continuous subsets of the time series—only limited by a minimum length of 10 data points for each time series. Among the outputs from the previously described *t* test is the degree of freedom *parameter*, which for the *t* test is one less than the number of time differences used to compute the mean time difference. To ensure that the shortest time series are not used for evaluating large lags, we filter the data

to only use the cross-correlations with *parameter* at least 8, which means that at least nine time differences are used in the *t* test. For example, at lag ± 3 the minimum length of the time series used will be 12 data points, as *time_diff* will have 12 values at lag 0, but only nine values at lag ± 3 , when the three na-values are excluded.

This ensures a reasonable accuracy of the mean time differences. To give an example, one site showed cross-correlations larger than *ciline* at lag -3 in 96 subsets of the time series. For this site, the *summary* of the error margin (*conf.high* – *conf.low*)/2 in days is 2.6 (*Min.*), 4.8 (*1st Qu.*), 6.0 (*Median*), 6.2 (*Mean*), 7.3 (*3rd Qu.*) and 11 (*Max*), and the *summary* of the lengths of the 96 subsets of the time series is 12 (*Min.*), 16 (*1st Qu.*), 20 (*Median*), 20 (*Mean*), 24 (*3rd Qu.*) and 30 (*Max*).

For each subset of the time series, we also find the mean temperature from the model temperatures used in Nesvik, Faroe Islands. As the model temperatures were measured on a daily basis, we use the mean temperature between the minimum and maximum date in each subset of the time series. The length of the time-series subsets is therefore not an issue for the temperatures, and in the previous example, the *summary* of the number of days used from the temperature profile corresponding to the 96 subsets of time series are 156 (*Min.*), 244 (*1st Qu.*), 313 (*Median*), 323 (*Mean*), 398 (*3rd Qu.*) and 544 (*Max*). This shows that the 96 subsets are between 5 and 18 months in length.

Simulations of probabilities *pchalimus* and *padult* for the chalimus and adult subpopulations (Figures 1 and 12) were based on mixed normal distributions generated by *dnormix(x, obj)* from the *normix* package where *x* is a vector of time (days) to calculate the probabilities and *obj* is generated by *normix(mu, sigma)* for given vectors of means (*mu*) and standard deviations (*sigma*)—using equal mixture proportions for the distributions. Cumulative probabilities for the four simulated subpopulations (Figure 1) were computed with *pnormix(q, obj)* where *q* is a vector of quantiles (days).

Time series *tspchalimus* and *tspadult* for the 11 simulated population probabilities used to compute cross-correlations (Figure 12b) were obtained similarly as for the lice counts using *xts(pchalimus, date)* and *xts(padult, date)*. Similarly, the cross-correlation object *ccfobj* was created by *ccf(drop(tspchalimus), drop(tspadult), plot = FALSE)* and then used to plot the cross-correlations with the *ggplot2* package. The relatively low confidence limit of about ± 0.12 should not be excessively interpreted as it changes with the number of data points in the simulation and in the denominator of $\pm(1.96/\sqrt{n})$ (Figure 12b; $n = 271$).

The counts and number of lags with significant correlations (Figure 8) is not fully understood. However, one possible suggestion is that the correlations at different lags are caused by correlations of the adult lice counts with series of subpopulations of chalimus lice generated by the 10–11 egg strings laid by an adult female in its lifetime of about 191 days. Consequently, for each adult female population, a series of subpopulations of chalimus lice will self-infect the

salmon fish separated on average by about 17–19 days. Starting with an adult population created by a current born infection, this adult population will start to produce series of chalimus infections. This can explain the positive lags. Then the attached lice grow into adult lice, explaining the correlations at negative lags. Finally, correlations at both positive and negative lags can be observed, as the chalimus lice produce adult lice, and vice versa.

Let us consider a numerical example following this line of reasoning. At the mean temperature in the Faroe Islands, we assume that the chalimus lice grow into adults in 51 days and adults produce chalimus in 36 days. Starting with an adult female at time 0, the series of 11 chalimus populations will show up at 36, 53, 70, 87, 104, 121, 138, 155, 172, 189 and 206 days, respectively, if we assume they are spaced in time by 17 days (Figure 12a, bottom panel). Note that a new generation of adult females will show up at $36 + 51 = 87$ days (Figure 12a, top panel). Therefore, the adult females at day 87 generated by the first subpopulation of chalimus at day 36 will correlate for lag -3 at $87 - 36 = 51$ days, but also correlate with the next chalimus generation for lag -2 at $87 - 53 = 34$ days, to the third for lag -1 at $87 - 70 = 17$ days and to the fourth for lag 0 at $87 - 87 = 0$ days. The remaining chalimus subpopulations will lead to positive lags 1–7 with respect to the first subpopulation of adult females at day 87. In this case, we consequently expect correlations for the lags -3 to 7. This is similar to what we observed for site A, in which lags -3 to 6 were observed to have significant correlations (Figure 8).

However, complexity arises as the 11 chalimus populations will generate 11 adult populations, which then also start to generate new series of chalimus populations. To simplify things, let us only consider the first 11 chalimus subpopulations that generate 11 new adult populations, and this time, look at the next adult generation that shows up at day 104 ($87 + 17$) (Figure 12a, top panel). This adult population will correlate with the first chalimus population for lag -4 at $104 - 36 = 68$ days and with the next for lag -3 at $104 - 53 = 51$ days, and so on, leading to correlations for lags -4 to 6.

A similar analysis of the remaining subpopulations leads to the following relation for which lags should show correlations: $[-3 - g; 7 - g]$. for a given adult generation, A_g , $g = 0 \dots 10$, produced by self-infection of 11 chalimus subpopulations. This predicts correlations at lags -3 to 7 for the first generation ($g = 0$), -4 to 6 for the next generation ($g = 1$), -5 to 5 for the third ($g = 2$) and -7 to 3 for the fourth ($g = 3$), -8 to 2 for the fifth ($g = 4$) and so on until the extreme -13 to -3 ($g = 10$). Note that the distribution of lags shifts towards negative lags with the generation number and is only symmetric (-5 to 5) for the third ($g = 2$). Note that the asymmetric distributions of lags -3 to 7 and -7 to 3 are similar to the observed lags for site A and C, respectively, while the symmetric distribution (-5 to 5) is fairly similar to site B (Figure 8).

The possible correlations when considering all the 11 adult and chalimus subpopulations are from lag -13 to 7 corresponding to -221 to 119 days centred at -51 days (lag -3), which together with its closest neighbours has the strongest correlation (Figure 12b), and represents the development time of 51 days used in the

simulation (Figure 12a). The gaps of 17 days between the side-band correlations (Figure 12b) represent the gaps of 17 days between the subpopulations used in the simulation (Figure 12a). Side-band gaps of about 15–16 days are observed in this study (Figure 11).

From the presented reasoning, different distributions of significant lags are to be expected depending on which subpopulations of the chalimus and adult lice dominate the lice counts, because not all 11 copepod populations are likely to self-infect the salmon with attached chalimus lice, and not all grow into adults. A continuous distribution of significant lags could indicate strong self-infection dynamics (Figure 8, sites A and D), while gaps in lags (Figure 8, sites B and C) could indicate missing subpopulations (e.g., removed copepod populations by currents) and less strong self-infection dynamics.

APPENDIX D

THE TEMPERATURE-DEPENDENCE PARAMETERS

To find the parameters $(\alpha_1, \alpha_2, \alpha_3)$, we first note that $\log(D(T)) = \log(\alpha_1) + \alpha_3 \log(T + \alpha_2)$. Therefore, a plot of $\log(D(T))$

versus $\log(T + \alpha_2)$ should be linear for a given α_2 . Inspired by *R for Data Science* (Wickham & Grolemund, 2017), we define the model function (the logarithmic form, in our case) and a standard distance measure between our data and our model. We then applied best parameter search using `optim()` with a guess. Initially, we only guessed one of the parameters α_2 , until we got a fair fit of $\log(D(T))$ versus $\log(T + \alpha_2)$. We subsequently used the resulting linear regression parameters to assist the guessing of the two remaining parameters, which we fed into `optim()` together with our best guess of α_2 .