

Space-Time Modelling of the Spread of Salmon Lice between and within Norwegian Marine Salmon Farms

Magne Aldrin^{1,2*}, Bård Storvik¹, Anja Bråthen Kristoffersen^{3,4}, Peder Andreas Jansen³

1 Norwegian Computing Center, Oslo, Norway, **2** Department of Mathematics, University of Oslo, Oslo, Norway, **3** Norwegian Veterinary Institute, Oslo, Norway, **4** Department of Informatics, University of Oslo, Oslo, Norway

Abstract

Parasitic salmon lice are potentially harmful to salmonid hosts and farm produced lice pose a threat to wild salmonids. To control salmon lice infections in Norwegian salmonid farming, numbers of lice are regularly counted and lice abundance is reported from all salmonid farms every month. We have developed a stochastic space-time model where monthly lice abundance is modelled simultaneously for all farms. The set of farms is regarded as a network where the degree of contact between farms depends on their seaway distance. The expected lice abundance at each farm is modelled as a function of i) lice abundance in previous months at the same farm, ii) at neighbourhood farms, and iii) other, unspecified sources. In addition, the model includes explanatory variables such as seawater temperature and farm-numbers of fish. The model gives insight into factors that affect salmon lice abundance and contributing sources of infection. New findings in this study were that 66% of the expected salmon lice abundance was attributed to infection within farms, 28% was attributed to infection from neighbourhood farms and 6% to non-specified sources of infection. Furthermore, we present the relative risk of infection between neighbourhood farms as a function of seaway distance, which can be viewed as a between farm transmission kernel for salmon lice. The present modelling framework lays the foundation for development of future scenario simulation tools for examining the spread and abundance of salmon lice on farmed salmonids under different control regimes.

Citation: Aldrin M, Storvik B, Kristoffersen AB, Jansen PA (2013) Space-Time Modelling of the Spread of Salmon Lice between and within Norwegian Marine Salmon Farms. PLoS ONE 8(5): e64039. doi:10.1371/journal.pone.0064039

Editor: Martin Krkosek, University of Toronto, Canada

Received: November 6, 2012; **Accepted:** April 10, 2013; **Published:** May 20, 2013

Copyright: © 2013 Aldrin et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was funded by the Research Council of Norway project PREVENT, "Salmon lice-prevention and treatment," project number 199778/S40. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: magne.aldrin@nr.no

Introduction

Infectious diseases constitute a constant problem in industrialised farming where there typically are both high densities of farms and high densities of animals within each farm. Disease outbreaks can have large economic consequences to farming industries, but can also have severe ecological effects if infections are spread to and impair the viability of wild animals. Several mathematical and statistical models within this field have been developed during the last decades and applied to as diverse diseases as foot-and-mouth disease, swine fever, bluetongue and infectious salmon anaemia [1–5]. In all these models, probabilities of infection relative to distance, often called transmission kernels [1,4], play an important role.

Salmon lice (*Lepeophtheirus salmonis*) are parasitic copepods that live on the skin surface of both wild and farmed salmonids. The parasite uses rasping mouthparts to feed on mucus, skin and underlying tissues of its host and thereby causes mechanical damage [6]. Possible effects of salmon farming on sea lice infections on wild stocks of salmonids, and hence the viability of such wild stocks, has evoked a large and contentious debate [7–9]. Nevertheless, the notion that salmon farming does affect local transmission of salmon lice to wild salmonids [9,10], as well as to farmed salmonids [11], seems well established. Recent studies also report that parasiticide-treatment of outwardly migrating salmon smolts significantly increases their marine survival compared to

non-treated control smolts, suggesting that salmon lice induce mortality in wild salmonid hosts [12–14]. Due to the potential impact of salmon lice of farm origin on wild stocks of salmonids, salmon lice infections on farmed fish are strictly regulated in Norway [11]. To enforce these regulations, numbers of salmon lice are counted on samples of farmed salmonids at regular intervals from all actively producing marine fish farms each month. Salmon lice abundances from these counts, i.e. the average number of salmon lice per fish, are reported to a central data base [11]. Measures to control salmon lice infections, i.e. the application of medical treatments or the use of cleaner-fish to prey on lice, are also reported to this same data base.

The large spatio-temporal dataset covering salmon lice abundance on salmon farms, and efforts to control these infections, should give insight into factors that affect farm levels of salmon lice infections and the contributing sources for such infection. The aim of the present study was to develop a modelling framework that could: i) disentangle different contributing sources of salmon lice infections; ii) estimate functional relationships between expected salmon lice infections and contributing factors, e.g. the between farm transmission kernel as a function of seaway distance between farms [1,4]; and iii) lay the foundation for a scenario simulation tool to examine the potential spread of salmon lice within and between salmon farms to assess the impact of control measures. We developed a stochastic space-time model where the monthly

salmon lice abundances at all Norwegian salmonid farms are modelled simultaneously. The model is similar in flavour to previous models developed by the same group of authors on infectious salmon anaemia and other salmonid diseases where farms appear as infected or non-infected [5,15,16]. Salmon lice, however, differ from these other diseases in that most farms are infected more or less all the time, but salmon lice abundances vary both over time and between farms.

The set of salmonid farms is regarded as a network where the degree of contact between each pair of farms depend on the seaway distance between them. The observed number of lice in a given sample of fish is assumed to be distributed according to a zero-inflated negative binomial distribution. The expected salmon lice abundance at each farm is modelled as a function of i) the observed lice abundances in previous months at the same farm, ii) the observed lice abundances in previous months at neighbouring farms, and iii) other, unspecified sources of infection. In addition, the model includes explanatory variables such as seawater temperature and numbers of fish at each farm. The model was fitted to data covering all marine salmonid farms in Norway from June 2003 until the end of 2011.

In the following sections, we first describe the data and then the model. Then, we present the fitted model and other results, and finally we draw some conclusions.

Materials and Methods

Salmonid Farming and the Problem of Salmon Lice

During a 30 year period from 1980 to 2010, annual production of farmed salmonids, i.e. Atlantic salmon (*Salmo salar*) and rainbow trout (*Oncorhynchus mykiss*), has grown from marginal to around one million tonnes in Norway [17]. For comparison, the total annual yield of wild Atlantic salmon and sea trout (*Salmo trutta*) in Norway was estimated at 850–1400 tonnes between 2000 and 2008 [18]. The production of salmonids consists of a freshwater juvenile phase, followed by a marine grow out phase, which is the focus of this study. Marine salmonid farming is regulated through a system of legal concessions authorised by the Norwegian Directorate for Fisheries (DFF; <http://www.fiskeridir.no>). All farms are registered with a geographic reference in the Aquaculture register, which is available at DFF's website. For marine farms that actively produce salmonids, it is mandatory to report key statistics on their fish stocks, fish health related statistics and seawater temperature at 3 m depth, on a monthly basis.

The production of a salmonid cohort on a farm typically initiates by stocking juvenile smolts to net-pens on the farm in spring or in the autumn. The net-pens allow free water exchange with the surroundings. After stocking, the fishes are on-grown on the farm for roughly 1.5 years and then slaughtered. The production time varies slightly according to the weight of the fish at slaughter, seawater temperature etcetera. On a given farm, only fish of a given year class are produced. After slaughtering, it is mandatory to follow the farm location for a short period before stocking a new cohort of salmonids. Occasionally fish may be moved from one farm location to another, in which case a new farm will initially report fish weights larger than juveniles.

Sea lice are naturally occurring parasitic copepods that spread by planktonic larvae. Farmed salmonids are infected by water flow through the net-pens. The parasites live on the surface of the fish and feed on mucus, skin and underlying tissues of its host [6]. Sea lice infections on farmed salmonids are dominated by the salmonid specific salmon louse, but may also be by the generalist parasite *Caligus elongatus*. However, regulations on reporting requirements regarding sea lice infections on farmed salmonids in Norway (see

below) are only directed at the salmon louse. Hence, we term the parasitic infections for salmon lice in this paper. The salmon louse has a life cycle consisting of three planktonic larval stages, of which the third copepodite-stage may attach to the surface skin of a salmonid host by a frontal filament. Upon infection, the parasite develops through four consecutive stages of attached larvae, followed by two pre-adult stages of lice that may move about on the surface of the fish (mobile stages), and finally adult males and females that also are mobile. Sea lice infections on farmed salmonids are regulated through a system of maximum thresholds of abundance of mobile stages. To enforce these regulations, farmers are instructed to count salmon lice on farmed salmonids at regular intervals. Counts are reported once a month to a database as the mean number of lice per fish within two stage categories: i) adult females, and ii) all other mobile stages (pre-adults and adult males). The farmers are instructed to report the highest mean count of adult females during a given month. In the present model, we sum the two categories of salmon lice and term this salmon lice abundance. The mean count of salmon lice from a sample of 20 fish from one net pen (before August 2009) or the mean of samples of multiples of 10 fish from half of the net-pens (from August 2009) was reported. Details of the sampling, counting and reporting procedures for salmon lice infections on farmed fish are given in Jansen et al. ([11]; and the Electronic supplementary material therein).

Data

The present data include all 1401 Norwegian marine fish farms with standing stocks of either Atlantic salmon or rainbow trout (salmonids) in any month from February 2003 to December 2011. The data is an updated version of the data presented in Jansen et al. [11], who analysed data up to 2010 in a study based on a different modelling approach than the present.

The data from the first four months were only used to construct lagged explanatory variables, whereas the model was fitted to data from June 2003 and onwards. In the following description, we therefore only summarise data from the latter period. The number of active farms was then reduced to 1361. Each salmonid farm normally had several consecutive periods of production of fish populations, interrupted by periods of fallowing (no fish on the farm). The fish population within a production period is termed a cohort and the present data consists of 4 255 cohorts with a total of 58 623 farm-months of cohort production. Farms produced between one and nine consecutive cohorts. For each salmonid farm, the salmon lice abundance on a sample of fish is obligatory to report to the responsible Norwegian Food Safety Authority. We assumed that the number of sampled fish always was 20. The reported salmon lice abundances were therefore multiplied by 20 and implemented as the response variable in the present model. Hence, we ignored that the number of fish sampled from August 2009 could be both less and more than 20 (see previous section), but note that this did not affect the mean number of lice per fish. However, lice abundance data were missing for about 3% of the total number of months with farmed fish, giving 56 836 farm-months with reported salmon lice abundance. Salmon lice abundances amount to between 0 and 7 in 95% of the farm-months, with a mean of 1.1 lice per fish. The distribution of salmon lice abundance is profoundly skewed, being exactly 0 in 32% of the farm-months, whereas the highest reported abundance is 163 lice per fish. The black curves in Figure 1 show how the observed lice abundance vary over the data period averaged over all farms and for three selected farms. The lice counts have a clear seasonal pattern driven by the seawater temperature. The farm-wise panels also show how the fish cohorts can be distributed over

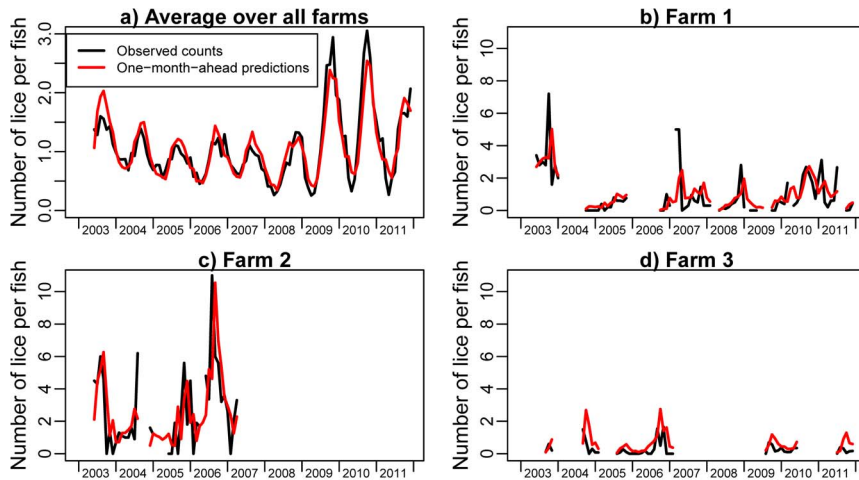


Figure 1. Time plots of observed (black) and one-month-ahead predictions (red) of salmon lice abundance averaged over all farms (panel a), and for three selected farms (panels b, c and d).
doi:10.1371/journal.pone.0064039.g001

the data period. Farms 1 and 2 in the figure have periods where the farms are active, but the salmon lice counts are missing. This is shown by predictions (red line) but missing observations (no black line) in the time plots.

The geographic location of each farm and pairwise seaway distances between all farms were compiled from Jansen et al. [11]. The seaway distance between two farms is defined as the shortest distance through water. The various salmonid farms had between zero and 31 other salmonid farms within a seaway distance of 10 km (median five). Figure 2 shows the locations of all Norwegian salmonid farms that were actively producing salmonids at some month during the study period, with a closer look at the 37 farms that were active in the Sognefjorden area. Of the latter, 19 farms were actively producing in October 2011.

Monthly figures of several other variables were also given separately for each farm. Each farm could produce either Atlantic salmon or rainbow trout, or both species at the same time. We calculated the monthly proportion of Atlantic salmon biomass for each farm, which was 91% on average. The monthly average weight of fish was between 100 g and 5.8 kg for 95% of the farm-months, with a mean of 2.0 kg. The mean number of fish per farm

was 530 000 (95% between 200 000 and 1.6 million). The mean seawater temperature at farms was 9.1°C (95% of all temperatures was between 3.4°C and 16.3°C). The seawater temperature was missing for 2.3% of the farm-months. Each missing temperature was imputed by a weighted mean of all observed temperatures the same month, with weights proportional to the inverse of the seaway distance to the current farm with the missing temperature. For about 4% of the farm-months, there were no fish at the same farm in the previous month in combination with the mean fish weight being less than 250 g. This indicates that the fish cohort was stocked for the first time on the given farm. Furthermore, for about 2% of the farm-months, there were no fish at the same farm in the previous month, but the mean fish weight was equal to or larger than 250 g. This indicates that the fish cohort had been relocated, i.e. moved from another marine farm. Medical salmon lice treatments were applied in 15% of the farm-months. Cleaner fish of the family *Labridae* were also applied at several farms [11], but we did not have sufficiently reliable data on the use of these, and we therefore ignored the use of cleaner fish in our modelling. Nor did we have sufficient data on salinity, which is known to affect the infection process of salmon lice [19].

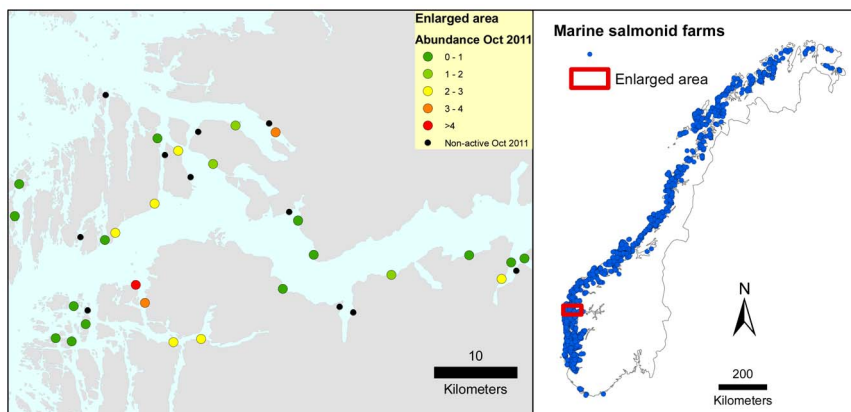


Figure 2. Location of salmonid farms. The left panel shows marine farms in the enlarged Sognefjorden area that were actively producing salmonids between June 2003 and December 2011, with observed salmon lice abundance indicated for those active in October 2011. The right panel shows marine farms that were active in the period June 2003 - December 2011 in the whole of Norway.
doi:10.1371/journal.pone.0064039.g002

Overview of the Model

Here we first describe the main features of the present salmon lice model, and then go into more detail. The model includes all actively producing Norwegian marine salmonid farms simultaneously, but the counted number of salmon lice at a given farm in a given month is modelled conditionally upon the situation at the current farm of interest and all other farms in previous months.

Now, consider farm i at time or month t , and let μ_{it} denote the expected abundance of salmon lice. This expectation is modelled as a function of i) observed lice abundances in previous months at the current farm, ii) observed lice abundances in previous months at neighbouring farms, and iii) other factors such as seaway distances to neighbouring farms, seawater temperatures etcetera.

The response variable is constructed from the reported salmon lice abundances by multiplying it by the number of fish sampled, which is always assumed to be $n=20$, see previous section and Jansen et al. [11]. The response variable is denoted by y_{it} , and modelled as a zero-inflated, negative binomially distributed variable [20] with expectation $n \cdot \mu_{it}$. This means that the values of y_{it} tend to be zero more often than can be modelled by a negative binomial distribution alone. Let p_{it}^z denote the probability of excess zero observations in this compound distribution. These excess zeroes come in addition to those expected from the negative binomial part of the distribution. Thus, y_{it} comes from a negative binomial distribution with probability $1 - p_{it}^z$, and has an excess zero value with probability p_{it}^z . Let further μ_{it}^{NB} denote the expectation in the negative binomial part of the distribution. Then, the expectation of y_{it} can be expressed as

$$E(y_{it}) = n \cdot \mu_{it} = (1 - p_{it}^z) \cdot \mu_{it}^{NB}. \tag{1}$$

Our main focus is on modelling the expected salmon lice abundance μ_{it} , and we give a detailed description of the model for μ_{it} in the next subsection. Furthermore, the probability of excess zero observations is modelled as a function of μ_{it} and a few other factors. Then, according to Eq. (1), the expectation in the negative binomial part of the compound distribution is given as $\mu_{it}^{NB} = n \cdot \mu_{it} / (1 - p_{it}^z)$. The negative binomial distribution has one parameter in addition to the expectation, here called R_{it} , and this is also modelled a function of μ_{it} . We present the sub-models for p_{it}^z and R_{it} in Appendix S1, since these only affect the shape of the distribution of y_{it} , which is not our focus here. The parameter R has often been denoted k in parasitological literature [21,22].

The motivation for using a negative binomial distribution begins with assuming that the number of lice counted on one fish, conditioned on the expected lice abundance, is Poisson distributed. However, we expect large variability from fish to fish, where typically some fish carry few salmon lice and some fish carry many. The negative binomial distribution is well suited to model such over-dispersed Poisson counts. Finally, if we sum the individual counts over $n=20$ fish, the total number of counted lice is also negative binomially distributed. However, when analysing the data, we found that there was an excess frequency of zeroes, which lead us to the zero-inflated, negative binomial distribution.

The zero-inflated, negative binomial distribution was also used by Jansen et al. [11], who modelled the same data as the present, although only updated up to December 2010. However, several other aspects of their approach differ from ours; i) they did not model all farms simultaneously in a network, ii) they ignored the lice counts at neighbouring farms, and iii) they had one model for the excess zeroes and another for the expectation in the negative

binomial distribution, yielding a complex formula for the expected lice abundance, whereas we model the latter explicitly.

The Expected Lice Abundance

The model for the expected lice abundance at farm i in month t has the following additive-multiplicative form:

$$\mu_{it} = S_{it} \cdot \kappa_{it}^{sus} \cdot (\lambda_{it}^w + \lambda_{it}^d + \lambda_{it}^o). \tag{2}$$

The two multiplicative terms in Eq. (2) are:

- S_{it} is an ‘‘at-risk’’ indicator that is 1 when farm i is active (has a positive number of farmed fish) at month t and 0 otherwise.
- κ_{it}^{sus} is a factor proportional to the susceptibility of farm i . It depends on explanatory variables that characterise the conditions for the fish at farm i at month i . Some explanatory variables are common for all farms (e.g. season), whereas others are farm-specific (e.g. seawater temperature). The farm specific term has the form

$$\kappa_{it}^{sus} = \exp\left(\sum_k \beta_k^{sus} x_{ikt}^{sus}\right), \tag{3}$$

where x_{ikt}^{sus} denotes explanatory variables for farm i at month t and β_k^{sus} denotes corresponding regression coefficients. See Table 1 for a list of variables included in the final model.

The three additive terms represent three possible sources of lice infection:

1. λ_{it}^w represents infection within the current farm of interest.
2. λ_{it}^d represents infection from neighbouring farms, depending on among others the seaway distances to these farms and on their lice abundances.
3. λ_{it}^o represents infection from other, non-specified sources, for instance a reservoir of infection on free roaming salmonids.

We assume that the infection pressure from each source can be added. Details on how each of these terms are modelled are given in the following.

Infection within a farm is modelled as

$$\lambda_{it}^w = \left(y_{i(t-1)} / n + \sum_{l=2}^{l=L} \rho_l y'_{i(t-l)} / n \right)^\alpha \cdot S_{i(t-1)}, \tag{4}$$

where

- $y'_{i(t-l)}$ is equal to $y_{i(t-l)}$ if farm i has been active in all the months from $t-1$ to $t-l$, but is zero if the farm has been inactive in any of these months, and
- α is a positive parameter that allows for a non-linear dependency of the previous months’ lice counts.
- $\rho_l, l=2, \dots, L$ are parameters that account for the effect of previous lice abundances in sequential time steps. The term $(y_{i(t-1)} / n + \sum_{l=2}^{l=L} \rho_l y'_{i(t-l)} / n)$ divided by $1 + \sum_{l=2}^{l=L} \rho_l$ is a weighted sum of the lagged observed lice abundances. Allowing for more than one lag may be useful, since the observed lice abundance is based on counts on small samples of fish, hence using several months may reduce the sampling uncertainty.

Table 1. Estimated parameters in the expected abundance μ_{it} with 95% confidence intervals for the selected model, with corresponding relative BIC values for selected parameters.

Parameter group	Variable name or parameter description	Farm specific variable	Parameter symbol	Est.	Lower	Upper	Relative BIC if deleted
misc.	Other sources		γ	0.076	0.068	0.084	
''-	Lagged lice counts		ρ_2	0.076	0.062	0.090	
''-	''-		ρ_3	0.025	0.016	0.034	
''-	''-		ρ_4	0.022	0.013	0.031	39
''-	Non-linear dependency		α	0.650	0.636	0.665	
''-	Sea distance function		ϕ_0	-1.444	-1.605	-1.283	
''-	''-		ϕ_1	-0.351	-0.272	-0.430	
''-	''-		ϕ_2	0.568	0.478	0.658	
susc.	intercept	no	β_k^{susc}	-0.385	-0.439	-0.332	
''-	$(t-103/2)$	''-	''-	$2.53 \cdot 10^{-3}$	$1.58 \cdot 10^{-3}$	$3.48 \cdot 10^{-3}$	
''-	$(t-103/2)^2$	''-	''-	$5.44 \cdot 10^{-5}$	$4.04 \cdot 10^{-5}$	$6.84 \cdot 10^{-5}$	
''-	$(t-103/2)^3$	''-	''-	$-1.50 \cdot 10^{-6}$	$-2.02 \cdot 10^{-6}$	$-0.99 \cdot 10^{-6}$	22
''-	$(temp-9)$	yes	''-	0.0979	0.0937	0.1020	
''-	$(temp-9)^2$	''-	''-	-0.0047	-0.0056	-0.0038	91
''-	$(latitude-64)$	''-	''-	0.0084	0.0042	0.0127	4
''-	$(temp-9) \times (latitude-64)$	''-	''-	0.0124	0.0112	0.0135	480
''-	$temp_t - temp_{t-1}$	''-	''-	-0.0220	-0.0278	-0.0162	45
''-	$\log(\text{weight})$	''-	''-	0.215	0.201	0.228	967
''-	Stocked	''-	''-	-0.978	-1.115	-0.841	188
''-	Relocated	''-	''-	0.218	0.101	0.336	4
''-	Salmon proportion	''-	''-	0.130	0.083	0.176	20
inf.	$\log(\text{number of fish})$	yes	β_k^{inf}	0.258	0.191	0.326	57

misc.: Miscellaneous parameters.
 susc.: Parameters related to the susceptible farm.
 inf. : Parameters related to the infectious farm.
 Est.: Estimate.
 Lower: Lower bound of 95% confidence interval.
 Lower: Upper bound of 95% confidence interval.
 doi:10.1371/journal.pone.0064039.t001

The term representing *infection from neighbouring farms* is summed over the separate contribution from all other farms and modelled as

$$\lambda_{it}^d = \sum_{j \neq i} \exp(\phi_0 + \phi_1 \cdot (d_{ij}^{\phi_2} - 1) / \phi_2) \cdot \kappa_{jt(t-1)}^{inf} \cdot \left(y_{j(t-1)} / n + \sum_{l=2}^L \rho_l y'_{j(t-1)} / n \right)^\alpha \cdot S_{j(t-1)} \tag{5}$$

Here,

- ϕ_0 quantifies the importance of neighbouring infection compared to the other two sources of infection.
- d_{ij} is the seaway distance between farms i and j .
- ϕ_1 and ϕ_2 are parameters that reflect the effect of the seaway distances to the neighbouring farms. The transformation $(d_{ij}^{\phi_2} - 1) / \phi_2$ is the well known Box-Cox transformation, which

allows for many different shapes of the distance function (Figure 3). When ϕ_2 approaches zero, the Box-Cox transformation becomes $\log(d_{ij})$, where \log here and elsewhere means the natural logarithm, and the distance function then becomes proportional to $d_{ij}^{\phi_1}$. When $\phi_2 = 1$, the distance function becomes proportional to the exponential function $\exp(\phi_1 \cdot d_{ij})$.

- κ_{jt}^{inf} is a factor proportional to the infectiousness of farm j , depending on explanatory variables that characterise the neighbouring farm j , for instance the number of fish. The infectiousness term has the form

$$\kappa_{jt}^{inf} = \exp\left(\sum_k \beta_k^{inf} x_{jkt}^{inf}\right), \tag{6}$$

where, as before, x_{jkt}^{inf} denote explanatory variables and β_k^{inf} denote corresponding regression coefficients. See Table 1 for an overview of variables included in the final model and the

Results section for other variants of the model that were investigated.

Infection from other sources, λ_{it}^o , is currently modelled as a constant γ , and as such acts as an intercept term. This can, however, be modified to a more complex form, for instance by including functions of space and time.

Relative Contribution from Each Source

For each farm and each month, the expected lice abundance can be decomposed into the relative contribution from each of the three infection sources. Summing over all farms and months gives the mean relative contribution from each source. For instance,

$$r^w = \left(\sum_t \sum_i S_{it} \cdot \kappa_{it}^{susc} \cdot \lambda_{it}^w \right) / \left(\sum_t \sum_i \mu_{it} \right) \quad (7)$$

is the mean relative contribution within a farm. The mean relative contribution from neighbouring farms, r^d , and from other sources, r^o , are defined similarly.

Estimation

The unknown parameters consist of α , γ , ϕ_0 , ϕ_1 , ϕ_2 , ρ_2, \dots, ρ_p and all β -s, joined into a parameter vector θ . The maximum likelihood estimates of θ was found by maximising the log likelihood of the data. An expression of the log likelihood is given in Appendix S1.

The log likelihood was maximised using the function *optim* in the statistical software *R*, using the method of Byrd et al. [23] for optimisation. Parameter uncertainties were based on the observed information matrix [24], and are reported as 95% Wald confidence intervals.

We investigated several variants of this model, with different explanatory variables included and with some parameters set to 0. For model selection, we used the Bayesian information criterion (BIC) [25,26]. This criterion balances the fit of the data to the number of parameters by minimising $-2\ln(\theta) + \log(m)$, where $\ln(\theta)$ is the log likelihood and m is the total number of observations and q the number of parameters. For some important explanatory variables, we considered second and third order terms as well as cross products between pairs of variables. We followed a strategy where main effects were included first, and second higher order effects or cross products were included if this improved the BIC value. Explanatory variables that were included in second or third order terms or cross products were centred to reduce correlations between the variables. We could alternatively have used Akaike's Information Criterion (AIC) [26,27] for model selection, which penalises the number of parameters less than BIC and therefore tends to give models with more parameters than when BIC is used. Choosing between BIC and AIC is primarily a matter of taste. We chose BIC because our dataset was large and hence we expected to end up with a rather complex optimal model also by using BIC.

Medical lice treatment was applied in 15% of the farm-months, which in principle should lead to reduced lice counts after treatments. It would indeed be useful if the model could be used to quantify the effect of such treatments. However, there is a dual relationship between treatment and lice counts; a high lice count will often induce a following treatment, whereas a treatment may result in lower lice counts. We don't know the actual dates for neither lice counts nor treatments, so it is not known for certain whether a treatment was applied before or after the corresponding lice count. Furthermore, we do not know what type of chemotherapeutic treatment that is used in each case. This makes

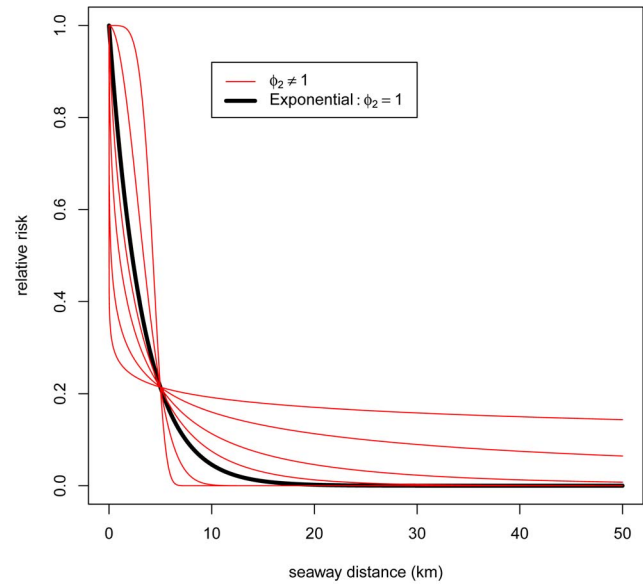


Figure 3. A selection of possible shapes for the relative effect of the seaway distance for various values of the ϕ_2 parameter. doi:10.1371/journal.pone.0064039.g003

it difficult to interpret the estimated effect of medical treatment. Therefore, we first consider a model without medical treatment as a factor, but then return to a model that includes medical treatment later on.

We estimated the model by fitting it to the observed lice counts from the 103 months between July 2003 and December 2011. We investigated models with up to $p = 4$ lags of previous lice counts, so the four months of data from February to June 2003 were only used to construct these lagged lice counts. As mentioned in the Data Section, 3% of the lice counts were missing. It is reasonable to believe that these were missing at random in the context of Little and Rubin [28], and the corresponding farm-months were therefore ignored in the likelihood. However, for the lagged lice counts, each missing value was imputed by the last observed value within the same cohort, if any. If this was impossible, the missing value was imputed by the mean of all observed lice counts in the current month.

Results

We first consider a model without medical treatment. Then, we investigated several variants of the model and selected the model with the minimum BIC value. Figure 1 shows time plots of the estimated expected number of lice per fish and the corresponding observed lice abundance averaged over all farms and for three selected farms. In our model, the expected numbers of lice per fish are the same as one-month-ahead predictions, and there is therefore a tendency for the expected values being shifted to the right compared to the observations. The selected model is given in Table 1, with estimates and 95% confidence intervals for each coefficient that enters the model for the expected salmon lice abundance. Estimates for eight additional coefficients in the sub-models for the excess zero probability and R_{it} are given in Table A in Appendix S1. For some coefficients, the table also shows how much the BIC value would increase if this coefficient was excluded (set to 0) from the model, here called the relative (to the main model) BIC values. A high value of the relative BIC indicates that the corresponding variable improves the model fit significantly.

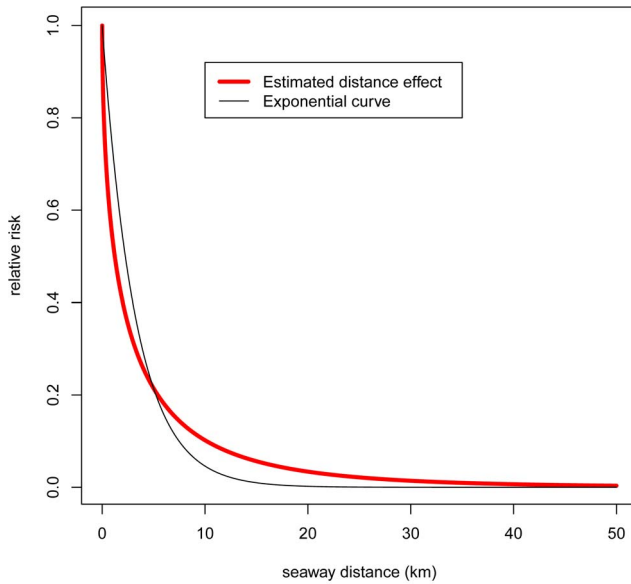


Figure 4. Estimated relative effect of the seaway distance.
doi:10.1371/journal.pone.0064039.g004

Some effects that are handled by groups of explanatory variables, for instance the seawater temperature effect, and the relative BIC values for such effects are presented in the text below. The confidence intervals are rather tight, and all parameters are highly significantly different from natural reference values of 0 or 1 (the latter is relevant for the ϕ_2 parameter in the Box-Cox transformations, where a value of 1 implies linearity on the exponential scale, i.e. an exponential distance function). We have therefore not included any p values in the table.

In the following presentation, we go through the different parts of the model, starting at the top of Table 1. The parameter $\gamma = \lambda_{it}^o$, representing infection from other, non-specified sources (Eq. (2)), has to be interpreted relative to infection within and between farms. Infection from other, non-specified sources accounted for only 6% of the lice abundance, as calculated from Eq. (7). Infection from neighbourhood farms accounted for a further 28% of the lice abundance, whereas the remaining 66% was attributed to infection within farms. The confidence intervals for these proportions are all narrower than $\pm 1.5\%$. Note that the latter

includes both within farm infection of infectious copepodite-stage salmon lice and pre-adult and adult stages of lice that survive from previous counts.

All four lags of observed lice abundance (Eq. (4) and Eq. (5)) were selected. Deleting the last three lags would increase the BIC value by 688, so including these lags improves the fit considerably. However, the estimated values of ρ_2, ρ_3 and ρ_4 sum up to 0.12, so these lags only account for 11% ($0.12/[1+0.12]$) of the weighted observed lice abundance.

The estimated value of α (Eq. (4) and Eq. (5)) was 0.65, which means that the expected lice abundance was not proportional to lagged observed lice abundance. Doubling the lagged observed lice abundance corresponds to a 57% increase in the expected lice abundance (since $2^{0.65} = 1.57$).

The infection pressure from neighbourhood farms decreased sharply by increasing seaway distance (Figure 4). A value equal to 1 for the parameter ϕ_2 , which controls the Box-Cox transformation, would correspond to an exponential distance function, which we assumed in previous work on infectious diseases with less informative data [5,16]. However, ϕ_2 was estimated to be much lower than 1, which gives a much steeper curve than the exponential.

We continue by commenting on the effect of explanatory variables included in the factor κ_{it}^{sus} (Eq. (3)). All the variables included here were measured in the current month t , except the seawater temperature difference between month t and $t-1$. The κ_{it}^{sus} term includes a function of time common to all farms, represented by a third order polynomial. Deleting the time trend (represented by three terms) from the model would increase the BIC value with only 46, so the time trend is not among the most important factors for improving the model fit.

The factor κ_{it}^{sus} is also a function of several characteristics of the susceptible farm. The most important is the seawater temperature, which is represented by a second order polynomial of temperature, a cross product of temperature and latitude and a difference between the temperatures at month t and $t-1$. Deleting these temperature terms from the model would increase the BIC value with 2395, implying that water temperature is a major predictor of salmon lice abundance. Figure 5a) shows the estimated combined effects of the seawater temperature and the latitude for farms at 60° North (close to the city of Bergen) and 68° North (in the Lofoten area). The seawater temperature dependency was much stronger in northern Norway than in southern Norway. In addition, in southern Norway, the effect flattened out when the

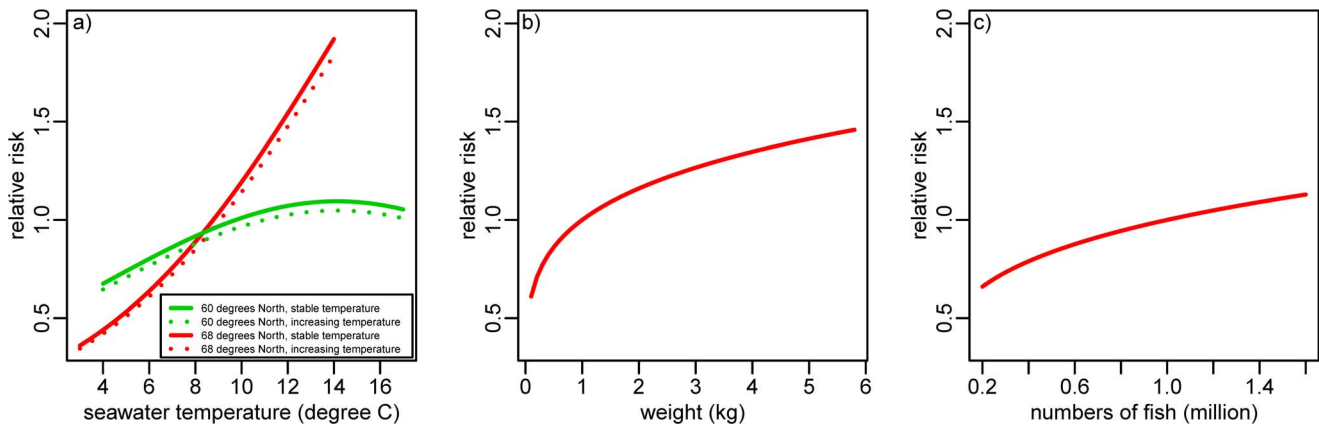


Figure 5. Estimated relative effects of the seawater temperature (a), of the mean fish weight (b) and of the number of fish at neighbouring farms (c).
doi:10.1371/journal.pone.0064039.g005

seawater temperature reached about 14°C. There was also a smaller, but statistically significant, effect of the seawater temperature increase from the previous month, which reduced the expected lice abundance slightly. In practice, this means that the effect of the nominal seawater temperature in periods with increasing seawater temperatures (typically spring and summer) differs from that in periods with decreasing seawater temperatures (autumn and winter).

The mean fish weight at susceptible farms is a very important explanatory variable, with a relative BIC value of 967. Since it enters in Eq. (3) as the logarithm of the mean fish weight, it has the form $w_{it}^{\beta^{sus}}$ after taking the exponential, where w_{it} is the mean fish weight and β^{sus} the regression coefficient, which was estimated to be 0.215. Figure 5b) illustrates the effect of the mean fish weight.

When a farm was active in month t , but not in month $t-1$, the fish cohort at the farm was considered as stocked if the mean fish weight was less than 250 g and relocated otherwise. This was modelled by including corresponding indicator variables in Eq. (3). Remember that previous lice counts in this case is 0. If a fish cohort was stocked, its expected lice abundance was reduced by 62% (since $1 - \exp(-0.978) = 1 - 0.38 = 0.62$) compared to a farm that was active in the previous month, but with 0 observed lice abundance. On the other hand, if a fish cohort was relocated, the comparable expected lice abundance increased by 24% ($\exp(0.218) - 1 = 1.24 - 1 = 0.24$). This is probably because relocated fish bring with them infection.

The expected lice abundance for a salmon cohort was 14% ($\exp(0.130) - 1 = 1.14 - 1 = 0.14$) higher than for a cohort of rainbow trout, if they had the same lagged observed lice abundance and everything else being equal.

The number of fish was the only characteristic of the infectious farms that was included in the sub-model for infectiousness (Eq. (6)). The infection pressure from neighbourhood farms increased by increasing numbers of fish at such farms since $f_{jt}^{\beta^{inf}} = \exp(\beta^{inf} \cdot \log(f_{jt}))$, where f_{jt} is the number of fish at a neighbouring farm and β^{inf} the regression coefficient, which was estimated to be 0.258. Figure 5c) illustrates the effect of the number of fish at neighbouring farms.

We finally report the results for models including medical lice treatment. Extending Eq. (3) with an indicator variable for medical treatment in the current month improved the BIC value with 352. However, the estimated coefficient was 0.28 (95% confidence interval from 0.25 to 0.31), which is the opposite sign of the expected causal relationship. If we instead include an indicator variable for medical lice treatment in the previous month, the estimated coefficient was -0.34 (95% confidence interval from -0.37 to -0.31), with an improvement in BIC value of 435. Hence medical lice treatment is potentially an important explanatory variable, but until we get more information on the timing and nature of medical treatments, this effect is not trustworthy. It does, however, make sense to include treatment in the previous month in the present data. The salmon farmers are instructed to report the highest abundance counted during a month and it is likely that this count is obtained before a potential medical treatment in a given month. Nevertheless, there is probably still some bias in the estimated effect of medical treatment. Inclusion or exclusion of medical treatment did not, however, substantially affect estimates of other parameters in the model (Table B in Appendix S1).

Discussion

The main structure of the present model is similar to models we have applied previously for viral diseases in salmonid farms

[5,15,16] and is in addition inspired by models applied for other animal diseases [1–3], in that the distances to neighbouring farms as well as conditions at these farms are important factors for the infection pressure. However, the stochastic modelling differs, which is natural since the response data are different. The viral infection data are binary, i.e. a farm is considered to be infected or non-infected, whereas the salmon lice infection data consist of counts that may take positive values and that are available for every farm-month. This results in a comparatively informative dataset, which has given us the opportunity to account for many different factors and model some of them in detail. Our model also has similarities to models used for human infections. For instance, Held and Paul [29] modelled the monthly number of laboratory cases of influenza in 140 districts of Southern Germany by a spatio-temporal negative binomial regression model, taking into account infection within and between regions (termed epidemic components) and an additional external infection source (called endemic component).

The model can be seen as a generalisation of vector autoregressive models [30], which are widely used in areas such as econometrics: Let \mathbf{y}_t denote the n^f -dimensional vector of salmon lice counts for all farms in month t , n^f being the number of farms. If the parameter α in Eq. (4) and Eq. (5) is exactly 1, the expected lice counts at month t is a weighted sum of the previous lice counts up to L lags back. This can be written as

$$E(\mathbf{y}_t) = \gamma_{0t} + \sum_{l=1}^{L} \Phi_{lt} \mathbf{y}_{t-l}, \quad (8)$$

where γ_{0t} is an n^f -dimensional vector of time-varying farm-specific intercepts representing infection from other, non-specified sources, whereas $\Phi_{lt}, l=1, \dots, L$ are time-varying ($n^f \times n^f$) matrices, where the diagonal represents within-farm infection and the off-diagonal elements represent between-farm infection.

The present modelling approach gives insight into the different contributing sources of salmon lice infection, as well as factors that affect lice abundance. Within farm infection was the dominating source of infection, accounting for 2/3 of the additive contribution to salmon lice abundance. The large within farm contribution to infection may partly be due to survival of pre-adult and adult lice over consecutive counts, since these stages are constituents of this internal source in addition to infectious copepodite-stage salmon lice produced within the farm stocks of salmonids. In contrast, the external sources of infection for a given farm constitute only infectious copepodite-stage salmon lice produced externally and transported passively with the water current. It is also important to note that the contribution from the different sources of infection will vary over time and space. As smolts are transferred from freshwater to marine waters, for example, they are free of salmon lice. Transmission of infectious copepodites will then only come from external sources. Hence, externally produced infection, which depends primarily on the density of salmon lice in neighbouring farms in the model, are important for seeding susceptible farm populations of salmonids with infection while the on-farm abundance of reproductive lice is low. The relative contribution from the within farm source of infection increases following the development of reproductive lice, and this source seems mainly responsible for amplifying and sustaining farm populations of salmon lice. The large contribution from within farm infection implies that the average farm primarily depends on itself in avoiding high levels of salmon louse infections. This, we argue, has important implications for the control and management of salmon lice. Due to the importance of self-inflicted infection, the

farmer should be motivated to control infection levels of reproductive lice to minimise the production of infectious copepodites. From a management point of view, it is reasonable to hold the farms responsible for keeping levels of reproductive lice low, to minimise their contribution to local infection pressure to wild fish or neighbourhood salmon farms.

The external unknown source of infection accounted for only 6% of the additive contribution to salmon lice abundance. This source may represent infection originating from wild salmonids [9,31], or alternatively escaped farmed salmonids [32], but which we do not have information on in the present study. The external unknown source is modelled with fewer parameters than the other sources of infection, and it is possible that it would account for a larger proportion of the infection if a more flexible function was used. However, we did not find any logical reasons to do so. The low contribution from the unknown external source also agrees with earlier suggestions that farmed salmon must be the dominating host of salmon lice in Norwegian coastal waters due to the large population size compared to wild stocks [11,32].

Infection from external neighbourhood farms accounted for the final 28% of the additive contribution to salmon lice abundance. The contribution to a given farm by a neighbourhood farm in the model is shaped by the functional relationship between the relative risk and seaway distance between farms (Figure 4). Compared to the exponential function used in previous models [5,15,16], the present relative risk decreases steeply at low but increasing distance to neighbourhood farms, but levelled off at higher relative risks for intermediately distant neighbourhood farms. We interpret this to imply that i) very close farms on average interact intensively with respect to salmon lice infection, and ii) that the number of intermediately distant neighbourhood farms and their levels of infection are important for expected salmon lice abundance. The latter probably reflects the effect of being located in areas of intensive salmon farming [11]. The function describing the relationship between infection from neighbourhood farms and distance represents an expression of a transmission kernel for salmon lice transmission between neighbourhood farms that can be used to simulate the spread of infection between farms and assess the impact of different control measures. Such transmission kernels are key elements of models designed to examine scenarios for the spread and control of a range of different livestock diseases (e.g. [4]). To our knowledge, this is the first transmission kernel estimated for metazoan parasites from an extensive system of farm populations. It is worth noting, however, that the shape of the function describing the transmission kernel for salmon lice probably varies in space and time, for example influenced by the temperature dependent duration of the planktonic phase of the louse life cycle [33], and local hydrodynamics [34].

The large body of data used to parameterise the present model yield estimates with relatively tight confidence intervals. The reports of salmon lice abundance, however, are highly skewed and zero inflated, implying that modelled expected abundance of lice will have relatively large uncertainty while predicting individual lice reports. It is also worth noting that the farmers are instructed to report the highest count during a month such that the modelled expectations should tend to be overestimates of the true lice abundance at corresponding times. In addition, there is one general characteristic of the data that we suspect may be confounded by the regulation regime authorised by the legal

authorities in Norway [11]. For the most influential factors in the model, effects seem relatively reduced for factor-values that imply expectations of high salmon lice abundance, i.e. effects are less than proportional to changes in the factors. This is apparent for example at especially high temperatures in the south, for large fish and for neighbourhood farms with large numbers of fish (Figure 5). Bearing in mind that the regulations on salmon lice infections focus on maximum legal thresholds of parasite abundances on farms, we anticipate that control efforts like the use of cleaner fish or medical treatments counter further increases in infection levels as lice abundances exceed legal thresholds. Since we cannot fully account for control efforts with the present resolution of data, this type of confounding related to legal thresholds of infection is to be expected. On top of this, we suspect that there may also be some underreporting of high salmon lice abundance. For example, the probability of finding a given salmon louse on a fish, i.e. the sensitivity of the counting procedure, probably decreases with increasing surface area to search on large fish. An implication of such potentially uncontrolled effects is that the model would be less predictive for high than for low lice abundance. Higher resolution of the data concerning timing and the nature of control measures, along with increased knowledge on the dynamic effects of different measures, are aspects that would improve the predictive capability of the present approach.

To conclude, we present a stochastic spatio-temporal model for salmon lice infection on farmed salmonids in Norway. We emphasise important insights that the model provides with respect to different sources of infection and how salmon lice abundance is affected by different environmental -, host related - and parasite related factors. In addition, we address control measures such as medical treatments and the use of cleaner fish, but point to shortcomings of the data to fully account for control measures at present. The present modelling framework lays the foundation for development of future scenario simulation tools for examining transmission and abundance of salmon lice on farmed salmonids under different control regimes. From 2012, the reporting procedures for salmon lice infections and medical treatments have changed from a monthly to a weekly frequency. With this increased time resolution we foresee that we can estimate more trustworthy effects of medical treatments and other control measures. This will benefit the prospects of applying the present model as a mathematical laboratory to investigate effects of complex and expensive actions before they are implemented in practice.

Supporting Information

Appendix S1
(PDF)

Acknowledgments

We thank Audun Stien for valuable comments to drafts of this article.

Author Contributions

Analyzed the data: MA BS. Contributed reagents/materials/analysis tools: ABK PAJ. Wrote the paper: MA BS ABK PAJ. Planned analysis methods and interpreted results: MA BS ABK PAJ.

References

- Keeling M, Woolhouse M, Shaw D, Matthews L, Chase-Topping M, et al. (2001) Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science* 294: 813–817.
- Diggle P (2006) Spatio-temporal point processes, partial likelihood, foot and mouth disease. *Stat Methods Med Res* 15: 325–336.

3. Höhle M (2009) Additive-multiplicative regression models for spatio-temporal epidemics. *Biom J* 51: 961–978.
4. Szmaraagd C, Wilson A, Carpenter S, Wood J, Mellor P, et al. (2009) A modeling framework to describe the transmission of bluetongue virus within and between farms in Great Britain. *PLoS ONE* 4: e7741.
5. Aldrin M, Lyngstad T, Kristoffersen A, Størvik B, Borgaen Ø, et al. (2011) Modelling the spread of infectious salmon anaemia (ISA) among salmon farms based on seaway distances between farms and genetic relationships between ISA virus isolates. *J R Soc Interface* 8: 1346–1356.
6. Costello M (2006) Ecology of sea lice parasitic on farmed and wild fish. *Trends Parasitol Trends Parasitol* 475–483.
7. Krkošek M, Ford J, Morton A, Lele S, Myers R, et al. (2007) Declining wild salmon populations in relation to parasites from farm salmon. *Science* 318: 1772–1775.
8. Krkošek M, Connors B, Morton A, Lewis M, Dill L, et al. (2011) Effects of parasites from salmon farms on productivity of wild salmon. *Proc Natl Acad Sci U S A* 108: 14700–14704.
9. Marty G, Saksida S, Quinn T (2010) Relationship of farm salmon, sea lice, and wild salmon populations. *Proc Natl Acad Sci U S A* 107: 22599–22604.
10. Krkošek M, Lewis M, Volpe J (2005) Transmission dynamics of parasitic sea lice from farm to wild salmon. *Proc Biol Sci* 272: 689–696.
11. Jansen P, Kristoffersen A, Viljugrein H, Jimenez D, Aldrin M, et al. (2012) Sea lice as a density dependent constraint to salmonid farming. *Proc Biol Sci* rspb20120084.
12. Krkošek M, Revic C, Gargan P, Skilbrei O, Finstad B, et al. (2012) Impact of parasites on salmon recruitment in the northeast Atlantic ocean. *Proc Biol Sci* rspb20122359.
13. Skilbrei O, Finstad B, Urdal K, Bakke G, Kroglund F, et al. (2013) Impact of early salmon louse, *Lepeophtheirus salmonis*, infestation and differences in survival and marine growth of sea-ranched Atlantic salmon, *Salmo salar* L., smolts 1997–2009. *J Fish Dis* 36: 249–260.
14. Jackson D, Cotter D, Newell J, McEvoy S, O'Donohue P, et al. (2013) Impact of *Lepeophtheirus salmonis* infestations on migrating Atlantic salmon, *Salmo salar* L., smolts at eight locations in Ireland with an analysis of lice-induced marine mortality. *J Fish Dis* 36: 273–281.
15. Scheel I, Aldrin M, Frigessi A, Jansen P (2007) A stochastic model for infectious salmon anemia (ISA) in Atlantic salmon farming. *J R Soc Interface* 4: 699–706.
16. Aldrin M, Størvik B, Frigessi A, Viljugrein H, Jansen P (2010) A stochastic model for the assessment of the transmission pathways of heart and skeleton muscle inflammation, pancreas disease and infectious salmon anaemia in marine fish farms in Norway. *Prev Vet Med* 93: 51–61.
17. Torrissen O, Olsen R, Toresen R, Hemre G, Tacon A, et al. (2011) Atlantic salmon (*Salmo salar* L.): the “super chicken” of the sea? *Reviews in Fisheries Science* 19: 257–278.
18. Anonymous (2009) Lakse- og sjøaurefiske 2008 (in Norwegian). Statistics Norway, Oslo. Noregs offisielle statistikk D 423. Available: <http://www.ssb.no/1005>.
19. Bricknell I, Dalesman S, O'Shea B, Pert C, Luntz A (2006) Effect of environmental salinity on sea lice *Lepeophtheirus salmonis* settlement success. *Dis Aquat Organ* 71: 201–202.
20. Zuur A, Ieno E, Walker N, Saveliev A, Smith G (2009) *Mixed Effects Models and Extensions in Ecology with R*. Springer, 1st edition. 574 p. doi:10.1007/978-0-387-87458-6.
21. Grenfell B, Wilson K, Isham V, Boyd H, Dietz K (1995) Modelling patterns of parasite aggregation in natural populations: trichostrongylid nematode-ruminant interactions as a case study. *Parasitology* 111 (Suppl.): S135–S151.
22. Irvine R, Stien A, Halvorsen O, Langvatn R, Albon S (2000) Life-history strategies and population dynamics of abomasal nematodes in Svalbard reindeer (*Rangifer tarandus plathyrrhynchus*). *Parasitology* 120: 297–311.
23. Byrd R, Lu P, Nocedal J, Zhu C (1995) A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing* 16: 1190–1208.
24. Aalen OO, Borgaen Ø, Gjessing H (2008) *Survival and event history analysis - A process point of view*. Springer, New York.
25. Schwarz G (1978) Estimating the dimension of a model. *The Annals of Statistics* 6: 461–464.
26. Claeskens G, Hjort N (2008) *Model selection and model averaging*. Cambridge University Press, Cambridge.
27. Akaike H (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19: 716–723.
28. Little R, Rubin D (1987) *Statistical analysis with missing data*. John Wiley & Sons, New York.
29. Held L, Paul M (2012) Modeling seasonality in space-time infectious disease surveillance data. *Biom J* 54: 824–843.
30. Lütkepohl H (2005) *New introduction to multiple time series analysis*. Springer, Berlin.
31. Bjørn P, Finstad B (2002) Salmon lice, *Lepeophtheirus salmonis* (krøyer), infestation in sympatric populations of arctic char, *Salvelinus alpinus* (L.), and sea trout, *Salmo trutta* (L.), in areas near and distant from salmon farms. *ICES J Mar Sci* 59: 131–139.
32. Heuch P, Mo T (2001) A model of salmon louse production in Norway: effects of increasing salmon production and public management measures. *Dis Aquat Organ* 45: 145–152.
33. Stien A, Bjørn P, Heuch P, Elston D (2005) Population dynamics of salmon lice *Lepeophtheirus salmonis* on Atlantic salmon and sea trout. *Mar Ecol Prog Ser* 290: 263–275.
34. Amundrud T, Murray A (2009) Modelling sea lice dispersion under varying environmental forcing in a Scottish sea loch. *J Fish Dis* 32: 27–44.
35. Johnson N, Kotz S, Kemp A (1993) *Univariate discrete distributions*. John Wiley & Sons, Inc., New York.