Bias in self-reported parasite data from the salmon farming industry

SEAN C. GODWIN [iD],[1,2,6] MARTIN KRKOŠEK,[3,4] JOHN D. REYNOLDS,[1] AND ANDREW W. BATEMAN[4,5]

[1]*Earth to Ocean Research Group, Department of Biological Sciences, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia V5A 1S6 Canada*
[2]*Department of Biology, Dalhousie University, 1355 Oxford Street, Halifax, Nova Scotia B3H 4R2 Canada*
[3]*Department of Ecology and Evolutionary Biology, University of Toronto, 25 Willcocks Street, Toronto, Ontario M5S 3B2 Canada*
[4]*Salmon Coast Field Station, General Delivery, Simoom Sound, British Columbia V0P 1S0 Canada*
[5]*Department of Geography, University of Victoria, 3800 Finnerty Road, Victoria, British Columbia V8P 5C2 Canada*

Abstract. Many industries are required to monitor themselves in meeting regulatory policies intended to protect the environment. Self-reporting of environmental performance can place the cost of monitoring on companies rather than taxpayers, but there are obvious risks of bias, often addressed through external audits or inspections. Surprisingly, there have been relatively few empirical analyses of bias in industry self-reported data. Here, we test for bias in reporting of environmental compliance data using a unique data set from Canadian salmon farms, where companies monitor the number of parasitic sea lice on fish in open sea pens, in order to minimize impacts on wild fish in surrounding waters. We fit a hierarchical population-dynamics model to these sea-louse count data using a Bayesian approach. We found that the industry's monthly counts of two sea-louse species, *Caligus clemensi* and *Lepeophtheirus salmonis*, increased by a factor of 1.95 (95% credible interval: 1.57, 2.42) and 1.18 (1.06, 1.31), respectively, in months when counts were audited by the federal fisheries department. Consequently, industry sea-louse counts are less likely to trigger costly but mandated delousing treatments intended to avoid sea-louse epidemics in wild juvenile salmon. These results highlight the potential for combining external audits of industry self-reported data with analyses of their reporting to maintain compliance with regulations, achieve intended conservation goals, and build public confidence in the process.

Key words: Caligus clemensi; environmental compliance; environmental management; environmental policy; industry data; Lepeophtheirus salmonis; Pacific salmon; policy implementation; salmon farms; salmon lice; sea lice; self-reported data.

INTRODUCTION

Environmental policies are intended to reduce human impacts on the natural world. The scope and "teeth" of an environmental policy can determine whether it meets its objectives, but implementation, or lack thereof, can have equally strong impacts to on-the-ground success (Howe 1993, Schwartz 2008, Wang et al. 2008). Analysis and evaluation of environmental policy implementation is critical for maximizing its efficacy (Potoski and Prakash 2013, Van Den Hoek et al. 2014).

For numerous environmental issues, policy implementation relies on self-reported data from industry to detect and mitigate problems (Livernois and McKenna 1999, Shimshack and Ward 2005, Barla 2007). For example, national policies commonly rely on industry to self-monitor pollution discharge or to self-report violations of pollution standards (Gamper-Rabindran and Finger 2013, Russell et al. 2013). Industry self-reporting enables

monitoring programs that are otherwise infeasible due to costs or logistics (Gunningham and Rees 1997), can shift monitoring costs from taxpayers to companies (Stoeckl 2004), and provides opportunities for companies to demonstrate cooperation with regulatory authorities (Helland 1998). On the other hand, biased self-reporting can have significant environmental consequences if inaccurate data generate management (in)actions and ecological responses that are misaligned with policy intentions. Audits, inspections, and threats of legal action or financial penalty help maintain the accuracy of industry self-reported data (Gray and Shimshack 2011), but accuracy can be costly for industry if the data indicate violations to regulations (Gunningham and Rees 1997, Heyes 2000). Although these incentives for inaccurate self-reporting can be high, the accuracy of self-reported data from industry is rarely investigated (but see references Li et al. [2017] and De Marchi and Hamilton [2006], for example).

Recent decades have seen a surge in the number of policies designed to reduce the environmental impacts of aquaculture, a set of industries whose rapid expansion, termed the blue revolution, has resulted in an

unprecedented rate of change in the relative contribution of farming to overall production (Duarte et al. 2007). Although this shift might reduce fishing pressure in some systems, it can also bring additional stressors in the form of emerging infectious diseases (Daszak et al. 2000), as did terrestrial agriculture before it (Harwood 1990). Aquaculture facilities can act as disease reservoirs that provide persistent sources of infection, even at low densities of wild hosts, and may lead to elevated extinction risk for wildlife (De Castro and Bolker 2005, Krkosek et al. 2013*a*). Effective policies are imperative to manage disease and allow aquaculture and marine wildlife to coexist, while sustaining a productive seafood supply.

Of the many forms of aquaculture, salmon farming has raised the greatest concern in terms of its impacts on wildlife. Salmon farms typically raise hundreds of thousands of Atlantic salmon (*Salmo salar*) in open-net pens (Orr 2007) and operate in the same nearshore marine waters through which wild salmon (*Oncorhynchus* spp. or *Salmo* spp.) migrate (Ford and Myers 2008). Consequently, pathogens and parasites are easily transmitted between farmed and wild salmon (Krkosek 2017). The most studied example of this phenomenon is the spread of native parasitic sea lice (primarily *Lepeophtheirus salmonis* and *Caligus* spp.) from wild adult salmon to farmed salmon, leading to amplification of infections on farms and transmission to wild juvenile salmon as they migrate past the net pens (Krkosek et al. 2007, Marty et al. 2010, Groner et al. 2016). In the absence of salmon farms, juvenile salmon typically have low infestation rates of sea lice (Costello 2009), but when they become infested they can suffer high levels of direct (Morton and Routledge 2005, Jones et al. 2008) or indirect mortality (Peacock et al. 2015, Godwin et al. 2017, Godwin et al. 2018), which is associated with reduced recruitment of wild adult salmon in Europe (Krkosek et al. 2013*b*, Vollset et al. 2016) and North America (Krkosek et al. 2011).

British Columbia (BC), Canada is the only region in the world that is a large global producer of both farmed and wild salmon (Groner et al. 2016). In this important social-ecological system, high sea-louse infestation rates on wild juvenile salmon in the early 2000s led to new regulations designed to control sea-louse numbers on farms (Peacock et al. 2013). British Columbia salmon farms are now required to self-monitor sea-louse infestation rates on their fish by performing sea-louse counts every month, which distinguish between the two main species of sea louse in BC: *L. salmonis* and *Caligus clemensi*. *Lepeophtheirus salmonis*, a salmonid specialist, is the focus of sea-louse regulations in BC (and most of the world) because until recently (Godwin et al. 2015, Godwin et al. 2017, Godwin et al. 2018) there was no evidence of harmful host effects from *C. clemensi*, a generalist that infects other nearshore marine fishes. Prior to 2017, treatments to reduce sea-louse abundances on BC salmon farms were almost exclusively conducted using an in-feed parasiticide called emamectin benzoate (trade name SLICE; Saksida et al. 2010). Current policy requires salmon farms to either harvest or conduct a delousing treatment when louse abundance exceeds three motile (i.e., adult or pre-adult) *L. salmonis* per fish. This policy drastically reduced sea-louse epizootics on BC salmon farms and seemed to halt the decline of local pink salmon populations in one region of the BC coast (Peacock et al. 2013). The implementation of this policy requires salmon farms to collect and report sea-louse count data themselves, in order to know when the three-louse threshold has been exceeded. The Canadian federal Department of Fisheries and Oceans (DFO) occasionally conducts scheduled audits of these industry counts (see the Appendix S1 for more local context).

Here, we assess whether industry self-reporting leads to underestimation bias in sea-louse monitoring data from BC salmon farms. We address this question by analyzing time series of self-reported, periodically audited parasite data from 91 farms (Fig. 1) over 6 yr, using a hierarchical model and Bayesian fitting methods. We use population-dynamic models to test for potential differences between audited and unaudited counts.
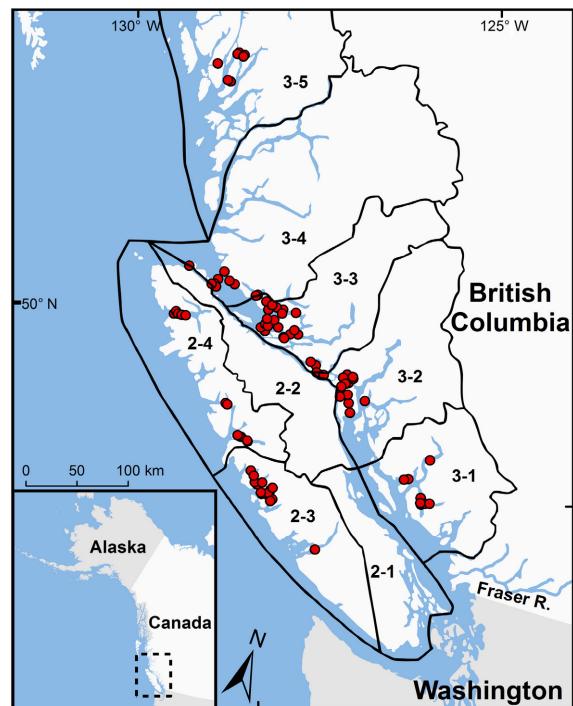

FIG. 1. Locations of the 91 British Columbia, Canada salmon farms in the industry sea-louse data set that were active (i.e., stocked with fish) for at least one month between 2011 and 2016. Solid black lines demarcate the boundaries of Department of Fisheries and Oceans (DFO) fish health surveillance zones, each of which is identified by number.

Methods

Data

We used publicly available sea-louse data collected by aquaculture industry staff on active farms in BC from 2011 through 2016 (Fisheries and Oceans Canada 2017*d*); once reported by industry, the data are managed and publicly hosted by DFO. These data are monthly averages of industry louse counts on individual salmon farms; data from individual sampling events or individual fish are not publicly available. During each sea-louse sampling event, farm staff capture stocked fish by seine net in three net pens and collect at least 20 fish by dip net from each seine (Fisheries and Oceans Canada 2016). One of the net pens is a reference pen that is assessed in every sampling event, and the other two are selected randomly (Fisheries and Oceans Canada 2016). The collected fish are then placed in an anesthetic bath of tricaine methanesulfonate (TMS, or MS-222) and assessed for lice by eye. The sea-louse counts concentrate on the more pathogenic and mobile pre-adult and adult life stages (termed "motile") rather than the attached stages earlier in development. For each farm in the data set, between zero and six sea-louse sampling events were conducted each month (mean = 1.64 events per month), and 99.1% of these months had one to three sampling events.

For every mean monthly motile louse count, the industry data set includes the month and year of the count, the number of sampling events contributing to the count, the age class of the stocked fish, a farm facility identifier, the farm's fish health surveillance zone (Fig. 1), and the previous month's mean louse count. The data set also has comments denoting planned or recently performed treatments, and from these we inferred the months in which treatments probably occurred. We excluded mean monthly counts that were missing any of this information (*n* = 440). Our final data set comprised 2,626 mean monthly louse counts over 6 yr, from seven health zones and 91 farms (Fig. 1). When visualizing the average monthly counts in a calendar year, we bootstrapped 95% confidence intervals by sampling the monthly counts in each year with replacement using the same sample size, combining the samples from the entire 6 yr span, calculating the mean of the combined sample, repeating this process 10,000 times, and calculating the 2.5th and 97.5th percentile of the 10,000 means for the confidence bounds.

In an attempt to ensure the quality of industry sea-louse data, DFO performs intermittent, prearranged audits of industry sea-louse monitoring. The farm facilities to be audited are chosen randomly each month, and the audits are then performed at the same time as an industry sampling event (Fisheries and Oceans Canada 2017*c*). During the audits, fish are selected in the same manner as for non-audited sampling events and divided equally between farm staff and DFO personnel for independent counting (Fisheries and Oceans Canada 2017*c*). Audit timing data are publicly available (Fisheries and Oceans Canada 2017*b*), and we therefore knew whether or not an audit was associated with each of the mean monthly louse counts. Audits occurred in 7.7% of the month–farm combinations in the data set.

Model

To assess the potential effect of DFO audits on industry-reported sea-louse counts, we fit a hierarchical model to the mean monthly counts from the self-reported and publicly available industry data set (Fisheries and Oceans Canada 2017*d*). While our primary goal was to clarify whether sea-louse counts decrease in months when DFO audits are not conducted, our model also accounted for the effects of month, year, and fish health surveillance zone, which incorporate variation in environmental conditions, as well as farmed salmon age class and louse density dependence. We fit the model using Markov chain Monte Carlo (MCMC) methods and uniform priors. Our model was composed of two identical parts, one for *L. salmonis* dynamics and one for *C. clemensi* dynamics, but was fit to both species' count data simultaneously to allow for the inclusion of an interspecific density dependence term in each part. For simplicity, we present only the *L. salmonis*-focused components of our model.

The model had distinct population growth and colonization components. The colonization component included covariates for year, health zone, and farmed salmon age class. The population growth component included covariates for month, year, health zone, farmed salmon age class, and inter- and intraspecific density dependence. Each component included a varying-intercept term (i.e., a random-effect level) for each farm. The effects of delousing treatment and DFO audit acted upon the entire model such that any effect would produce a proportional change in louse counts. We fit our single model and interpreted its results rather than performing model selection on smaller models because (1) parsimony was not our objective, (2) the complexity of the model was justified by the size of the data set, and (3) all the parameters had strong biological justifications (see Gelman and Rubin [1995] and Neal [2012] for discussion of this method of Bayesian model inference). Table 1 provides an overview of the model notation, and Appendix S1 presents model diagnostics, posterior plots, and the full set of parameter estimates.

In the industry data set, the mean monthly counts are rounded to the nearest 10 and include a high proportion of zeroes. Similar to Jansen et al. (2012), we model monthly louse counts per 10 fish (i.e., we multiplied the mean monthly counts by 10) and assume a negative binomial error distribution:

$$N_{\mathrm{L},t,f} \sim \text{negative binomial}\left(\mu_{\mathrm{L},t,f}, \rho_{\mathrm{L}}\right) \qquad (1)$$

TABLE 1. Overview of model notation.

Symbol	Description	Data or prior details
Response variable		
N	motile louse count per ten fish	integer
Indices for predictor variables		
t	index for time step (month)	72 months (January 2011–December 2016)
f	index for farm	91 farms (all active BC farms in 2011–2016)
treat	index for treatment status	three treatment statuses (treated in t, in $t-1$, or not treated)
zone	index for fish health surveillance zone	seven health zones (all health zones for the 91 farms)
year	index for year	six years (2011–2016)
month	index for month of the year	12 months of the year (January–December)
class	index for age class of farmed fish	two age classes (<1 yr in sea, ≥1 yr in sea)
Model predictions and parameters		
μ	predicted motile louse count per 10 fish	
λ	per-capita louse population growth rate	
γ	louse colonization rate	
τ	effect of delousing treatment	fixed, $U(-100,100)$
α	effect of DFO audit	fixed, $U(-100,100)$
ρ	negative binomial shape parameter	$U(-100,100)$
Submodel predictions and parameters		
η	linear function for λ ($\lambda = e^{\eta}$)	
η_i	coefficient in η, associated with predictor i	fixed, all $U(-100,100)$†
φ	varying-coefficients term for farms in η	random, $U(-100,100)$
β	linear function for γ ($\gamma = e^{\beta}$)	
β_i	coefficient in β, associated with predictor i	fixed, all $U(-100,100)$
ψ	varying-coefficients term for farms in β	random, $U(-100,100)$
Subscripts		
L	subscript for *Lepeophtheirus salmonis*	
C	subscript for *Caligus clemensi*	
o	subscript for reported louse count	

Note: DFO, Department of Fisheries and Oceans; BC, British Columbia, Canada.
†With the exception of $\eta_{C,zone3-1}$, which was constrained to be $U(-10,10)$ due to low sample size, as discussed in Methods.

where $N_{L,t,f}$ is the reported mean motile louse count per 10 fish in month t on farm f, $\mu_{L,t,f}$ is the predicted mean motile louse count per 10 fish for that month and farm, and ρ_L is the negative binomial shape parameter, fit as an additional free parameter. We allowed the shape parameter of the negative binomial to vary between species because one key relationship that drives overdispersion is the ratio of immigrants to births (Bolker 2008); the relative contributions of colonization and on-farm population growth to counts are likely to be very different between *C. clemensi* and *L. salmonis* due to the large difference in their mobility among hosts (Saksida et al. 2015, Atkinson et al. 2018). Using the negative binomial prevents complications associated with zero counts because, unlike the gamma distribution, the negative binomial distribution allows for zeros. A gamma hurdle model, for example, may better accommodate zero counts, but such a model that included adequate complexity (i.e., terms for density dependence, delousing treatments, temporal and spatial effects, and the necessary random effects) proved inordinately complex and challenging to fit. In this case, the negative binomial is a natural choice for a single distribution that accommodates zeros as well as extra-Poisson variability

introduced by model uncertainty. Also, parasites, including sea lice (Heuch et al. 2011), are typically overdispersed on hosts (i.e., the variance is greater than the mean) and are commonly considered to be negative binomially distributed (Crofton 1971, Anderson and May 1992, Shaw et al. 1998). We note that the standard error (and therefore the distribution) of the mean monthly louse counts will, in fact, be affected by the number of fish assessed on a farm in any given month, but this information is not available. We use the negative binomial distribution as a good empirical approximation of the true underlying distribution (see Appendix S1: Fig. S1 for diagnostics).

At their most basic, our models take the form

$$\mu_{L,t+1,f} = N_{L,t,f}\lambda_{L,t,f} + \gamma_{L,t,f} \tag{2}$$

where the mean motile count in month t ($\mu_{L,t+1,f}$) is predicted by the sum of intrinsic on-farm dynamics and external colonization pressure. The on-farm dynamics are the product of the previous month's count on that farm ($N_{L,t,f}$) and a per capita population growth rate ($\lambda_{L,t,f}$) affected by on-farm conditions, such as louse density and treatment status. The colonization rate ($\gamma_{L,t,f}$) is

a function of extrinsic factors. Both the per capita population growth rate and the colonization rate are modeled as transformed linear functions, where $\lambda_{L,t,f} = e^{\eta_{L,t,f}}$ and $\gamma_{L,t,f} = e^{\beta_{L,t,f}}$; thus $\eta_{L,t,f}$ and $\beta_{L,t,f}$ are analogous to linear predictors in generalized linear mixed models (GLMMs) with a logarithmic link function. Note that our model assumes key population processes can be captured by considering only motile lice, ignoring details of early developmental stages. But while our model does not explicitly consider larval lice, those that colonize as larvae and develop into motiles are still captured in the colonization-rate component of our model. We also implicitly model the influence of environmental conditions on development time between larval attachment and the motile stage (~27 d for *L. salmonis* at 10°C [Johnson and Albright 1991] and unknown length for *C. clemensi*) in the growth-rate component. We do this by including spatial and temporal predictors (i.e., health zone, month, and year) that are strongly correlated with temperature and salinity (Pickard and McLeod 1953, Fisheries and Oceans Canada 2017*a*) – important drivers of development timing for sea lice (Johnson and Albright 1991). While it would be more direct to use actual temperature and salinity measurements from the farms, we were unable to obtain these data.

We allow delousing treatment to influence monthly louse counts, including the motiles that developed over the month, as

$$\mu_{L,t+1,f} = e^{\tau_{L,\text{treat},t,f}} \left(N_{L,t,f}\lambda_{L,t,f} + \gamma_{L,t,f} \right) \quad (3)$$

where $e^{\tau_{L,\text{treat},t,f}}$ is a proportional mortality term for farm f in month t that results from delousing treatment. The exponent takes one of three levels: zero if treatment last occurred more than one month before t, or one of two levels to describe louse decline when treatment occurred in month t or $t - 1$, corresponding to the two-month effectiveness previously described for emamectin benzoate (Lees et al. 2008, Saksida et al. 2010).

To account for the potential effects of DFO audits on industry louse counts, we extend the model such that

$$\mu_{L,t+1,f} = e^{\tau_{L,\text{treat}t,f}} e^{\alpha_{L,t+1,f}} \left(\frac{N_{L,t,f,n}}{e^{\alpha_{L,t,f}}} \lambda_{L,t,f} + \gamma_{L,t,f} \right) \quad (4)$$

where $e^{\alpha_{L,t+1,f}}$ allows for a proportional change in a farm's counts in month t (or $t + 1$, as appropriate), if indeed they are influenced by whether the DFO audits a farm. The exponent, $\alpha_{L,t,f}$, takes the value of 0, if no audit occurs, or a fitted estimate, if an audit occurs. When an audit does occur in month t, that month's reported louse count ($N_{L,t,f,o}$) is rescaled by $e^{\alpha_{L,t,f}}$ to account for any louse count observation error associated with audits.

The linear predictor for per capita population growth rate in a given month and farm ($\ln(\lambda_{L,t,f})$ from Eq. 2) takes the form

$$\eta_{L,t,f} = \eta_{L,0} + \eta_{LL,\text{class}}N_{L,t,f} + \eta_{LC,\text{class}}N_{C,t,f}$$
$$+ \eta_{\text{class},t,f} + \eta_{L,\text{zone},f} + \eta_{L,\text{year},t} + \eta_{L,\text{month},t} + \varphi_{L,f}. \quad (5)$$

The first term is an intercept ($\eta_{L,0}$) that defines growth rate at base factor levels (class = <1 year in sea, zone = 2–3, year = 2011, month = January) and louse counts of zero for *L. salmonis* and *C. clemensi* abundance in the previous month. The next two terms represent the interspecific and intraspecific density dependence on a farm's per capita louse count in month t; they each incorporate a farm's louse counts in month $t - 1$ ($N_{L,t-1,f}$ and $N_{C,t-1,f}$) and a coefficient that describes density dependence due to either *L. salmonis* ($\eta_{LL,\text{class}}$) or *C. clemensi* ($\eta_{LC,\text{class}}$). These coefficients depend on the age class of the farm's stocked fish in month t (farm and month subscripts not shown), which can take one of two levels: fish that have spent less than one year in seawater, and fish that have spent greater than or equal to one year in seawater. Age class also directly affects the per capita growth rate of a farm in month t ($\eta_{\text{class},t,f}$) because fish surface area may influence louse survival (Tucker et al. 2002). There are three additional coefficients for categorical covariates: the age class of the fish for a given month and farm ($\eta_{\text{class},t,f}$), the health zone of the farm ($\eta_{L,\text{zone},f}$), and the year ($\eta_{L,\text{year},t}$) and month ($\eta_{L,\text{month},t}$) of the louse count. Among other things, these coefficients represent spatial and temporal variability in temperature and salinity. The last term in Eq. 5 is a varying coefficient (hereafter termed a "random effect" to continue the parallel between our submodels and GLMMs) describing how the intercept varies among farms.

We modeled the linear predictor for a farm's colonization rate in month t such that

$$\beta_{L,t,f} = \beta_{L,0} + \beta_{\text{class},t,f} + \beta_{L,\text{zone},f} + \beta_{L,\text{year},t} + \psi_{L,f} \quad (6)$$

where $\beta_{L,0}$ is an intercept term describing the colonization rate when counts were zero for *L. salmonis* and *C. clemensi* in the previous month, and at base factor levels (class = <1 yr in sea, zone = 2–3 and year = 2011). The age class of the farm's fish ($\beta_{\text{class},t,f}$), the farm's fish health surveillance zone ($\beta_{L,\text{zone},f}$), and the year of the count ($\beta_{L,\text{year}}$) affect colonization rate in the same manner that they affected per capita growth rate in Eq. 5. Finally, we include a random effect on the intercept of farm facility to account for the hierarchical nature of the data ($\psi_{L,f}$) while limiting the number of farm facility parameters in our model. We attempted to include colonization density dependence as we did for the population growth rate linear function (Eq. 5), but that model did not converge.

The full equation for the predicted mean motile louse count is as follows: where the overbraces reference the previously described equations and where the predicted

$$\mu_{L,t+1,f} = \overbrace{e^{\tau_{L,\text{treat},tf}}}^{\text{Treatment(3)}} \overbrace{e^{\alpha_{L,t+1,f}}}^{\text{Audit(4)}} \left(\overbrace{\frac{N_{L,t,f,n}}{e^{\alpha_{L,tf}}}}^{\text{Rescale count(4)}} \overbrace{e^{\eta_{L,tf}}}^{\text{Population growth rate(5)}} + \overbrace{e^{\beta_{L,tf}}}^{\text{Colonization rate(6)}} \right) \tag{7}$$

mean motile louse count is the mean of the negative binomial probability density function in Eq. 1.

We used uniform priors between -100 and 100 for all of our parameters except for a single $\eta_{\text{zone},C}$ coefficient. We constrained this parameter's uniform prior between -10 and 10 because the *C. clemensi* data for this particular fish health surveillance zone (zone 3–1) included two drastic month-to-month declines in mean louse counts that caused fitting complications. This constraint had little effect on the median parameter estimate for $\eta_{C,\text{zone}3-1}$, which was highly uncertain regardless due to low sample size in that zone (Appendix S1: Fig. S2). The uniform prior has the desirable property that the overall probability of either inflation or deflation is equal – a necessary condition to avoid biasing the audit parameters, or indeed any of the parameters. While this prior is not flat on the response scale (i.e., when exponentiated it is not distributed uniformly), the audit result was unchanged when we used a prior that was "flatter" on the response scale over the range of interest (but therefore assigned more prior weight to inflationary audit effects; see Appendix S1). We fit separate variance parameters for the two farm-facility random effects ($\varphi_{L,f}$ and $\psi_{L,f}$), in addition to a parameter describing the correlation between the two; we used the same uniform priors for these random-effects parameters. The random effects were each drawn from a multivariate normal distribution with a mean of zero and a covariance matrix determined by an inverse-Wishart distribution, which was in turn informed by the three random-effect parameters (Gelman and Hill 2007). The random effects were

needed to account for the hierarchical nature of the data while also avoiding fitting 180 separate fixed-effect parameters.

When parameter estimates are given in the text, we report them as posterior medians along with the lower and upper 95% credible intervals. Our analysis was performed using R 3.2.1 (R Core Team 2019) and JAGS 4.3.0 (Plummer 2017) with the R package R2jags 0.5-7 (Yu-Sung Su and Yajima 2015).

RESULTS

We found that industry sea-louse counts are biased downward. In months when DFO performed its prearranged audits, the industry's mean monthly *C. clemensi* counts were 1.95 (95% credible interval: 1.57, 2.42) times counts in months when DFO did not audit (Fig. 2), after accounting for all the other variables in our model.

For *L. salmonis*, counts in audit months were 1.18 (1.06, 1.31) times counts in non-audit months (Fig. 2). When this correction factor was applied to the reported counts in months that DFO did not audit, the number of mean monthly counts that exceeded the three-louse threshold rose from 376 to 437, an increase of 16%. On average, farms waited 1.2 ± 0.1 months (mean $\pm$ SE) before treating or harvesting following a month when the mean *L. salmonis* count exceeded the treatment threshold. When using counts corrected for the audit effect, this delay lengthened to 1.6 ± 0.2 months, a 30% increase. During the wild juvenile salmon migration (March–June), the mean delay before action was 0.7 ± 0.2 months, according to the reported counts, but the corrected counts reveal a 22% increase in delay time, to 0.9 ± 0.2 months.

Our model predicted that unaudited counts were reduced by a factor of 0.51 (0.41, 0.63) and 0.85 (0.76, 0.94) relative to the audited *C. clemensi* and *L. salmonis* counts, respectively (Fig. 4). In comparison, treatments in the previous month reduced counts by a factor of 0.31 (0.23, 0.1) for *C. clemensi* and by 0.38 (0.28, 0.50) for *L. salmonis*. In the absence of audits and treatments, the predicted contributions of population growth rate and colonization rate to louse counts were drastically different between the two species. For *C. clemensi*, estimated on-farm population growth contributed 25% (20%, 29%) to counts, on average, and estimated colonization contributed 65% (61%, 68%). For *L. salmonis*, these contributions were almost perfectly reversed, with population growth contributing 65% (61%, 68%) to counts and colonization contributing 28% (25%, 32%).
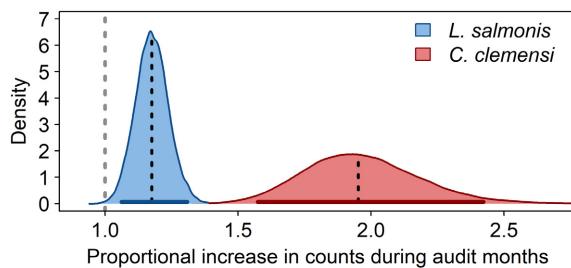
FIG. 2. Proportional increase in farm louse counts (*Caligus clemensi* and *Lepeophtheirus salmonis*) in months when DFO audited farms, as shown by the exponentiated posterior distributions for the audit parameters in our model. The colored horizontal lines give the 95% credible intervals and the dotted black vertical lines describe the exponentiated median parameter estimates. The dashed gray vertical line at 1.0 indicates no difference in counts between months with and without a DFO audit.
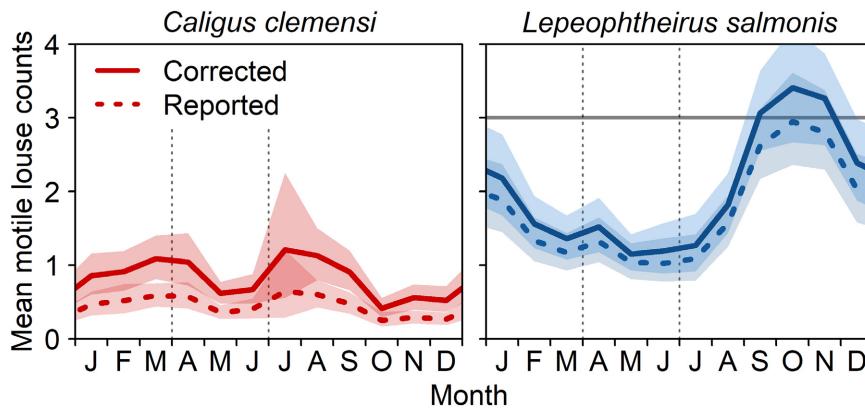
FIG. 3. Monthly variation in reported and corrected mean motile *Caligus clemensi* and *Lepeophtheirus salmonis*) counts on British Columbia salmon farms between 2011 and 2016. Shaded regions depict 95% confidence intervals for the monthly means and do not reflect uncertainty in the audit effect (Fig. 2); these confidence intervals were calculated by bootstrapping the counts 10,000 times while maintaining the hierarchical structure with respect to year (see methods). Reported counts were corrected by the median estimate of the audit effect (Fig. 2) for both species. Dashed vertical lines delineate the wild juvenile salmon migration period (April–June) and the horizontal line in the right panel represents the current threshold for delousing treatment. We show the monthly means connected by lines for visual effectiveness, but we note that these are discrete monthly data.
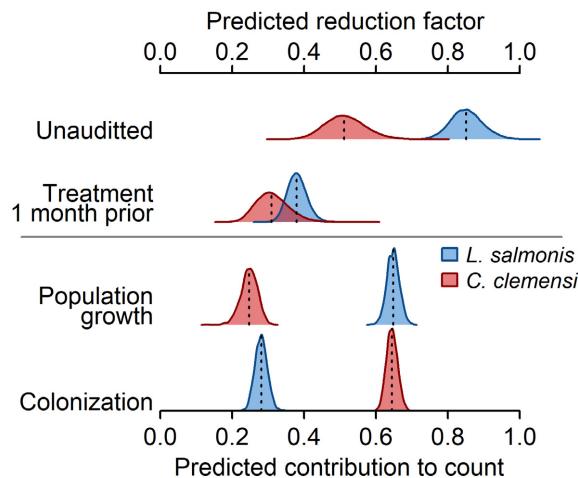

FIG. 4. Relative effects of the four main components of the model (audit, treatment, population growth, and colonization from Eq. [7]) for both *L. salmonis* (blue) and *C. clemensi* (red). The audit and treatment terms are represented as reduction factors that proportionally change louse counts. The reduction factor for the treatment term is its exponentiated posterior and the reduction factor for the audit term is the reciprocal of its exponentiated posterior. The population growth and colonization components are presented as their predicted percent contributions to unaudited and untreated counts. The dashed vertical lines denote the median values.

DISCUSSION

Our results suggest that sea-louse counts reported by the salmon farming industry are lower than the true abundance of parasites on their fish. When the federal government audited a sea-louse count, the industry's mean counts for that month increased by a factor of 1.18 for *L. salmonis* and by 1.95 for *C. clemensi*. This underestimation bias has likely led to fewer treatments and longer delays before management action is taken during sea-louse outbreaks, thus reducing the effectiveness of environmental policy designed to safeguard wild salmon populations.

When we corrected the reported sea-louse counts for the audit effect, we found considerable differences between reported and corrected counts. For *C. clemensi*, corrected counts were 1.96 times reported counts and nearly reached the same levels as *L. salmonis* during migration months for wild juvenile salmon (Fig. 3). These high *C. clemensi* levels on farms have the potential to affect wild populations given recent correlative work linking *C. clemensi* abundance with decreased foraging success, competitive ability, and growth in wild juvenile salmon (Godwin et al. 2015, Godwin et al. 2017, Godwin et al. 2018). Despite the growing evidence for effects of *C. clemensi* on wild salmon and the unexpectedly high abundances of this species on farms during the wild juvenile migration, *C. clemensi* are not currently targeted by management on salmon farms. For *L. salmonis*, the louse species targeted by management, corrected average counts were higher than the three-louse treatment threshold for one-quarter of the year (September, October, and November; Fig. 3). By contrast, there were no months in which the average reported counts exceeded the treatment threshold.

The management implication of biased sea-louse counts is that delousing treatments to control outbreaks may be delayed or may not occur at all. After correcting the reported *L. salmonis* counts for the audit effect, we found a 16% increase in the number of mean monthly counts that broke this three-louse threshold. This increase does not imply that there should have been 16% more treatments – to make this calculation we would

need to know how many *individual* counts broke the threshold, and industry counts are only reported as monthly means – but it does suggest that underestimated counts may well influence treatment decisions. Indeed, we found the average delay between a month with a threshold-breaking count and the subsequent management action increased by 30% when using corrected counts instead of reported counts, and by 22% during months in which wild juvenile salmon migrate. The corrected delay during the migration window, while shorter than the corrected average year-round delay (0.9 ± 0.2 months during migration and 1.6 ± 0.2 months year-round), is still longer than the 15 calendar days license conditions allow farms before they must "implement a plan which will reduce absolute sea lice inventory" during migration months (Fisheries and Oceans Canada 2016).

One solution to remove bias from self-reported data would be for monitoring to be conducted by an independent third party. However, regulatory transitions from self-monitoring to third-party monitoring are rare and, to our knowledge, their effect on data accuracy remains unassessed. Other options for increasing data accuracy include improving data collection training for industry staff (Dasgupta et al. 2000) and performing audits or inspections without advance notice *after* data collection takes place, so that data are always collected with the risk of subsequent review (Laplante and Rilstone 1996). The latter solution eliminates the potential for finding statistical agreement between industry and auditor counts during the audits themselves (as is the case for sea-louse monitoring (Elmoslemany et al. 2013, Fisheries and Oceans Canada 2018)), only to observe a drop in accuracy when auditors are absent (as our results show). In the case of sea lice, the counting protocol itself could be improved, for example by discontinuing the use of dip nets to avoid dislodging lice (Godwin et al. 2015). The current protocol requires high diligence to minimize or retrieve dislodged sea lice; this is especially true for the highly mobile *C. clemensi* (Saksida et al. 2015, Atkinson et al. 2018), which could explain why our results indicated that *C. clemensi* are more underestimated than *L. salmonis*.

A final option to improve the accuracy of sea-louse counts would be to systematically apply correction factors (here estimated to be 1.18 for *L. salmonis* and 1.96 for *C. clemensi*) to the reported monthly means. These correction factors could be dynamic over time and informed by empirical quantitative analyses like the one we present here; the prospect of reduced correction factors could provide incentive for industry to improve data accuracy. It should be noted that the actual bias in *individual* counts (i.e., audited vs. non-audited counts) is probably greater still than the bias reported here, because the count data in the public data set are given as monthly means. Bias in audited counts was thus diluted in the pool of each month's counts (mean = 1.64 counts per month).

CONCLUSIONS

Industry self-reported data are often thought to be unbiased because incentives exist for accurate self-reporting. For example (1) self-reporting is done under surveillance (Short and Toffel 2010), (2) audits or inspections are performed without advance notice (Russell 1992, Makofske 2019), (3) misreporting is met with administrative, legal, or financial penalties (Shimshack and Ward 2005, Gray and Shimshack 2011), (4) accurate data are easy to obtain (Gunningham and Rees 1997, Gray and Shimshack 2011), or (5) industry is not penalized when self-reporting demonstrates violations to regulations (Livernois and McKenna 1999). None of these conditions exist for sea-louse counts on salmon farms; in particular, self-reported violations (i.e., *L. salmonis* counts above three motile lice per fish) result in the farm having to perform a costly delousing treatment or harvest its fish earlier than it would otherwise. Our results suggest that when incentives for accurate self-reporting are not strong, bias can occur in industry self-reported data. By providing explicit and strong incentives for accurate self-reporting, environmental policies can reap the benefits of industry data while also increasing compliance and effectiveness.

LITERATURE CITED

Anderson, R. M., and R. M. May. 1992. Infectious diseases of humans: dynamics and control. Oxford University Press, Oxford, UK.

Atkinson, E., A. W. Bateman, L. M. Dill, M. Krkosek, J. D. Reynolds, and S. C. Godwin. 2018. Oust the louse: leaping behaviour removes sea lice from wild juvenile sockeye salmon. Journal of Fish Biology 93:263–271.

Barla, P. 2007. ISO 14001 certification and environmental performance in Quebec's pulp and paper industry. Journal of Environmental Economics and Management 53:291–306.

Bolker, B. M. 2008. Ecological models and data in R. Princeton University Press, Princeton, New Jersey, USA.

Costello, M. J. 2009. How sea lice from salmon farms may cause wild salmonid declines in Europe and North America and be a threat to fishes elsewhere. Proceedings of the Royal Society B 276:3385–3394.

Crofton, H. 1971. A quantitative approach to parasitism. Parasitology 62:179–193.

Dasgupta, S., H. Hettige, and D. Wheeler. 2000. What improves environmental compliance? Evidence from Mexican industry. Journal of Environmental Economics and Management 39:39–66.

Daszak, P., A. A. Cunningham, and A. D. Hyatt. 2000. Emerging infectious diseases of wildlife: threats to biodiversity and human health. Science 287:443–449.

De Castro, F., and B. Bolker. 2005. Mechanisms of disease-induced extinction. Ecology Letters 8:117–126.

De Marchi, S., and J. T. Hamilton. 2006. Assessing the accuracy of self-reported data: an evaluation of the toxics release inventory. Journal of Risk and Uncertainty 32:57–76.

Duarte, C. M., N. Marbá, and M. Holmer. 2007. Rapid domestication of marine species. Science 316:382.

Elmoslemany, A., S. K. Whyte, C. W. Revie, and K. L. Hammell. 2013. Sea lice monitoring on Atlantic salmon farms in New Brunswick, Canada: comparing audit and farm staff counts. Journal of Fish Diseases 36:241–247.

Fisheries and Oceans Canada. 2016. Marine finfish aquaculture licence under the Fisheries Act. http://www.pac.dfo-mpo.gc.ca/aquaculture/licence-permis/docs/licence-cond-permis-mar/licence-cond-permis-mar-eng.pdf

Fisheries and Oceans Canada. 2017a. British Columbia lighthouse sea-surface temperature and salinity data (Pacific), 1914-present. https://open.canada.ca/data/en/dataset/719955f2-bf8e-44f7-bc26-6bd623e82884

Fisheries and Oceans Canada. 2017b. DFO audits at BC marine finfish aquaculture sites. https://open.canada.ca/data/en/dataset/5cfd93bd-b3ee-4b0b-8816-33d388f6811d

Fisheries and Oceans Canada. 2017c. DFO marine finfish aquaculture audit activities in BC. http://www.pac.dfo-mpo.gc.ca/aquaculture/reporting-rapports/mer-mar-audit-verif/index-eng.html

Fisheries and Oceans Canada. 2017d. Industry sea lice counts at BC marine finfish aquaculture sites. https://open.canada.ca/data/en/dataset/3cafbe89-c98b-4b44-88f1-594e8d28838d

Fisheries and Oceans Canada. 2018. Regulating and monitoring British Columbia's marine finfish aquaculture facilities 2015-2016: monitoring and audits. https://www.dfo-mpo.gc.ca/aquaculture/management-gestion/mar-rep-rap-2015-2016/health-sante-eng.htmlsea-lice

Ford, J. S., and R. A. Myers. 2008. A global assessment of salmon aquaculture impacts on wild salmonids. PLoS Biology 6:e33.

Gamper-Rabindran, S., and S. R. Finger. 2013. Does industry self-regulation reduce pollution? Responsible Care in the chemical industry. Journal of Regulatory Economics 43:1–30.

Gelman, A., and J. Hill. 2007. Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, New York, New York, USA.

Gelman, A., and D. B. Rubin. 1995. Avoiding model selection in Bayesian social research. Sociological Methodology 25:165–173.

Godwin, S. C., L. M. Dill, M. Krkosek, M. H. H. Price, and J. D. Reynolds. 2017. Reduced growth in wild juvenile sockeye salmon infected with sea lice. Journal of Fish Biology 91:41–57.

Godwin, S. C., L. M. Dill, J. D. Reynolds, and M. Krkosek. 2015. Sea lice, sockeye salmon, and foraging competition: lousy fish are lousy competitors. Canadian Journal of Fisheries and Aquatic Sciences 72:1113–1120.

Godwin, S. C., M. Krkosek, J. D. Reynolds, L. A. Rogers, and L. M. Dill. 2018. Heavy sea louse infection is associated with decreased stomach fullness in wild juvenile sockeye salmon. Canadian Journal of Fisheries and Aquatic Sciences 75:1587–1595.

Gray, W. B., and J. P. Shimshack. 2011. The effectiveness of environmental monitoring and enforcement: a review of the empirical evidence. Review of Environmental Economics and Policy 5:3–24.

Groner, M. L., et al. 2016. Lessons from sea louse and salmon epidemiology. Philosophical Transactions of the Royal Society B 371:20150203.

Gunningham, N., and J. Rees. 1997. Industry self-regulation: an institutional perspective. Law & Policy 19:363–414.

Harwood, R. R. 1990. A history of sustainable agriculture. Pages 3–19 in C. A. Edwards, R. Lal, P. Madden, R. H. Miller, and G. House, editors. Sustainable agricultural systems. Soil and Water Conservation Society, Ankeny, Iowa, USA.

Helland, E. 1998. The enforcement of pollution control laws: inspections, violations, and self-reporting. Review of Economics and Statistics 80:141–153.

Heuch, P. A., G. Gettinby, and C. W. Revie. 2011. Counting sea lice on Atlantic salmon farms—Empirical and theoretical observations. Aquaculture 320:149–153.

Heyes, A. 2000. Implementing environmental regulation: enforcement and compliance. Journal of Regulatory Economics 17:107–129.

Howe, C. W. 1993. The US Environmental policy experience: A critique with Suggestions for the European Community. Environmental and Resource Economics 3:359–379.

Jansen, P. A., A. B. Kristoffersen, H. Viljugrein, D. Jimenez, M. Aldrin, and A. Stien. 2012. Sea lice as a density-dependent constraint to salmonid farming. Proceedings of the Royal Society B 279:2330–2338.

Johnson, S., and L. Albright. 1991. Development, growth, and survival of Lepeophtheirus salmonis (Copepoda: Caligidae) under laboratory conditions. Journal of the Marine Biological Association of the United Kingdom 71:425–436.

Jones, S. R., E. Kim, and W. Bennett. 2008. Early development of resistance to the salmon louse, Lepeophtheirus salmonis (Krøyer), in juvenile pink salmon, Oncorhynchus gorbuscha (Walbaum). Journal of Fish Diseases 31:591–600.

Krkosek, M. 2017. Population biology of infectious diseases shared by wild and farmed fish. Canadian Journal of Fisheries and Aquatic Sciences 74:620–628.

Krkosek, M., J. Ashander, L. N. Frazer, and M. A. Lewis. 2013a. Allee effect from parasite spill-back. American Naturalist 182:640–652.

Krkosek, M., B. M. Connors, A. Morton, M. A. Lewis, L. M. Dill, and R. Hilborn. 2011. Effects of parasites from salmon farms on productivity of wild salmon. Proceedings of the National Academy of Sciences USA 108:14700–14704.

Krkosek, M., A. Gottesfeld, B. Proctor, D. Rolston, C. Carr-Harris, and M. A. Lewis. 2007. Effects of host migration, diversity and aquaculture on sea lice threats to Pacific salmon populations. Proceedings of the Royal Society B 274:3141–3149.

Krkosek, M., C. W. Revie, P. G. Gargan, O. T. Skilbrei, B. Finstad, and C. D. Todd 2013b. Impact of parasites on salmon recruitment in the Northeast Atlantic Ocean. Proceedings of the Royal Society B 280:20122359.

Laplante, B., and P. Rilstone. 1996. Environmental inspections and emissions of the pulp and paper industry in Quebec. Journal of Environmental Economics and Management 31:19–36.

Lees, F., M. Baillie, G. Gettinby, and C. W. Revie. 2008. The efficacy of emamectin benzoate against infestations of *Lepeophtheirus salmonis* on farmed Atlantic salmon (*Salmo salar* L) in Scotland, 2002–2006. PLoS ONE 3:e1549.

Li, S.-M., et al. 2017. Differences between measured and reported volatile organic compound emissions from oil sands facilities in Alberta, Canada. Proceedings of the National Academy of Sciences USA 114:E3756–E3765.

Livernois, J., and C. J. McKenna. 1999. Truth or consequences: enforcing pollution standards with self-reporting. Journal of Public Economics 71:415–440.

Makofske, M. P. 2019. Inspection regimes and regulatory compliance: How important is the element of surprise? Economics Letters 177:30–34.

Marty, G. D., S. M. Saksida, and T. J. Quinn. 2010. Relationship of farm salmon, sea lice, and wild salmon populations. Proceedings of the National Academy of Sciences USA 107:22599–22604.

Morton, A., and R. Routledge. 2005. Mortality rates for juvenile pink *Oncorhynchus gorbuscha* and chum *O. keta* salmon infested with sea lice *Lepeophtheirus salmonis* in the Broughton Archipelago. Alaska Fishery Research Bulletin 11:146–152.

Neal, R. M. 2012. Bayesian learning for neural networks. Springer Science & Business Media, New York, New York, USA.

Orr, C. 2007. Estimated sea louse egg production from Marine Harvest Canada farmed Atlantic salmon in the Broughton Archipelago, British Columbia, 2003–2004. North American Journal of Fisheries Management 27:187–197.

Peacock, S. J., M. Krkosek, A. W. Bateman, and M. A. Lewis. 2015. Parasitism and food web dynamics of juvenile Pacific salmon. Ecosphere 6:1–16.

Peacock, S. J., M. Krkosek, S. Proboszcz, C. Orr, and M. A. Lewis. 2013. Cessation of a salmon decline with control of parasites. Ecological Applications 23:606–620.

Pickard, G. L., and D. McLeod. 1953. Seasonal variation of temperature and salinity of surface waters of the British Columbia coast. Journal of the Fisheries Board of Canada 10:125–145.

Plummer, M. 2017. JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. Proceedings of the 3rd international workshop on distributed statistical computing. Vol. 124, No. 125.102003, pp. 1–10.

Potoski, M., and A. Prakash. 2013. Do voluntary programs reduce pollution? Examining ISO 14001's effectiveness across countries. Policy Studies Journal 41:273–294.

R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. www.R-project.org

Russell, C. S. 1992. Monitoring and enforcement of pollution control laws in Europe and the United States. Pages 195–219 *in* R. Pethig, editor. Conflicts and cooperation in managing environmental resources. Springer, Berlin, Germany.

Russell, C. S., W. Harrington, and W. J. Vaughn. 2013. Enforcing pollution control laws. RFF Press, New York, New York, USA.

Saksida, S. M., I. Bricknell, S. Robinson, and S. Jones. 2015. Population ecology and epidemiology of sea lice in Canadian waters. Research document 2015/004. Canadian Science Advisory Secretariat, Ottawa, Ontario, Canada.

Saksida, S. M., D. Morrison, and C. W. Revie. 2010. The efficacy of emamectin benzoate against infestations of sea lice, *Lepeophtheirus salmonis*, on farmed Atlantic salmon, *Salmo salar* L., in British Columbia. Journal of Fish Diseases 33:913–917.

Schwartz, M. W. 2008. The performance of the endangered species act. Annual Review of Ecology, Evolution, and Systematics 39:279–299.

Shaw, D., B. T. Grenfell, and A. P. Dobson. 1998. Patterns of macroparasite aggregation in wildlife host populations. Parasitology 117:597–610.

Shimshack, J. P., and M. B. Ward. 2005. Regulator reputation, enforcement, and environmental compliance. Journal of Environmental Economics and Management 50:519–540.

Short, J. L., and M. W. Toffel. 2010. Making self-regulation more than merely symbolic: the critical role of the legal environment. Administrative Science Quarterly 55:361–396.

Stoeckl, N. 2004. The private costs and benefits of environmental self-regulation: which firms have most to gain? Business Strategy and the Environment 13:135–155.

Tucker, C., C. Sommerville, and R. Wootten. 2002. Does size really matter? Effects of fish surface area on the settlement and initial survival of *Lepeophtheirus salmonis*, an ectoparasite of Atlantic salmon *Salmo salar*. Diseases of Aquatic Organisms 49:145–152.

Van Den Hoek, J., M. Ozdogan, A. Burnicki, and A.-X. Zhu. 2014. Evaluating forest policy implementation effectiveness with a cross-scale remote sensing analysis in a priority conservation area of Southwest China. Applied Geography 47:177–189.

Vollset, K. W., et al. 2016. Impacts of parasites on marine survival of Atlantic salmon: a meta-analysis. Fish and Fisheries 17:714–730.

Wang, M., M. Webber, B. Finlayson, and J. Barnett. 2008. Rural industries and water pollution in China. Journal of Environmental Management 86:648–659.

Yu-Sung, S. u., and M. Yajima.2015. R2jags: using R to Run 'JAGS'. https://cran.r-project.org/web/packages/R2jags/index.html

SUPPORTING INFORMATION