A statistical analysis of sea-lice medicine use and benthic monitoring at Scottish marine salmon farms (2002 – 2014).

Reply to reviewers' comments

17 Dec 2015

Thomas A Wilding and Kenneth D Black

Scottish Association for Marine Science

Scottish Marine Institute

OBAN

Argyll. PA37 1QA

United Kingdom.

Background

The research reported in 'A statistical analysis of sea-lice medicine use and benthic monitoring at Scottish marine salmon farms (2002 – 2014)' is the culmination of three linked research programmes. Phase I identified that EMB-treated sites tended to host fewer crustacean compared with those where EMB was not used. Phase II extended Phase I by quantifying the amount of EMB used, during each production cycle, and using this as a predictor in the models. Phase II concluded that there was an association between EMB treatment rate and decline in non-target crustacea in models where the Distance Classes (CE, AZE and Ref) were included in the same model. This report , dated 01 July 2015, was reviewed, by the SARF steering group and external reviewers (four from MERCK, one SARF appointed). This document constitutes our detailed response to the anonymised reviewers' comments.

Reviewer 1	
Comment 1.	…it is quite possible that the estimated effect in the Reference Distance Class will not differ significantly from the estimated effects in the CE and AZE Distance Classes, a common trend will be fitted, which will then suggest, incorrectly, that Richness and Abundance decline with EMB in the Reference Distance Class. I am not suggesting that there are few data with which to estimate the EMB effect in each Distance Class. The most parsimonious models (Tables 8 and 9) show that there is no strong evidence that the effect of EMB differs between Distance Class, and that the common trend is of decreasing Crustacean Richness and Abundance with increasing EMB. Searching for parsimonious models is usually a good thing to do. However, the most parsimonious model can give a misleading picture about specific effects of interest – here, the effect of EMB in the Reference Distance Class. For examples, suppose that EMB has no effect on Richness and Abundance in the Reference Distance Class. Further, suppose there are few data with which to estimate the effect
Response	The data are well balanced with approximately equal distribution between CE, AZE and Ref distance classes. The reviewer acknowledges this.
Action	See below
Comment 2.	However, given the potential implications of inferring that EMB has an effect in the Reference Distance Class, I think it is important to correctly quantify the evidence for such an effect – i.e. without 'bolstering up the evidence by assuming' that the effect is identical to the effects in the other Distance Classes. One way of doing this is to repeat the modelling using only the data from the Reference Distance Class. However, I think it would be simpler to just fit the models in Tables 8 and 9 with EMBSR replaced by EMBSR:Distance Class (i.e. a common trend replaced by a Distance Class specific trend) and to give the estimates of the slopes for each Distance Class with standard errors and credible intervals.
Response	Fitting an interaction term is one option, but the model selection process, which is recommended numerous texts/papers, is to think carefully about including terms which offer no substantive improvement in the model fit. The models presented in the Final Report do not require the interaction term and including it makes the model fit worse. Including unnecessary terms reduces the precision of the effect-estimates in the model.
Action	Despite necessitating greater effort additional reference-only analyses were conducted and are reported.

Comment 3.	More of a comment. I am reassured that the Bayesian results closely match the frequentist results from lme4. Personally, I would use the parametric resampling that is available in lme4 to check on the adequacy of the frequentist results. However, when presenting Bayesian results, it is good practice to give the priors used with justification.
Response	The priors were 'uninformative' (flat), so the results are based on the data only
Action	Methods section amended.
Comment 4.	Results can be sensitive to the choice of priors, and default priors are not always good priors, particularly when applied to variance components. It is not clear what model is fitted to Crustacean Abundance. I think it is Cube-root Abundance = fixed effects + random effects + noise
Response	The model as specified above is correct, the cube-root of abundance is a response variable. However, the response is not bounded by zero. I have checked this by extending the model to make predictions at EMB treatment rates that are much greater than those actually used. As expected, where this is done, negative abundance counts are predicted i.e. the model is correctly specified). Negative abundances are predicted by the model, but not within the actual EMB treatment range used in the data analysed (taking the predicted Reference site abundances as a baseline). This has nothing to do with the Bayesian model
Action	No action required.
Comment 5.	There is nothing in the model that precludes negative Abundances, either observed or fitted. Perhaps it is an effect of the Bayesian implementation – perhaps all simulated Abundances are rounded to the nearest non-negative integer but if so, then there are important details of the modelling that have been omitted. If there are lots of zero counts in the data, then assuming normally distributed data (even on the cube-root scale) is wishful (and I would feel more comfortable with modelling the Abundances as Poisson data with a (large) over-dispersion term (as used for the Richness data), supplemented by either parametric resampling or a Bayesian analysis that properly reflects the discrete nature of the data.
Response	The zero counts pose challenges but the normality of the residuals were checked (and looked reasonable) as part of routine model checking. In addition, initial testing indicates that the suggested modelling approach does not deal well with the over-dispersion. This is because there is considerable variance between 'reps' taken at the same time/place. Using a quasi-Poisson approach instead is not recommended where the over-dispersion is as extreme as that observed in the current data
Action	None, the abundance data are best dealt with using a transformation (to reduce heteroscedasticity) and modelled using a normal model.
Reviewer 2	
Note	W&B = the Wilding and Black report being reviewed.
Comment 6.	Foremost amongst these difficulties is the inability to match EMB sediment residue data with crustacean richness and abundance because of a lack of temporal and spatial coincidence between residue and macroinvertebrate sampling. This means that there is no direct evidence of an association between EMB exposure and effect. Instead, EMB use data (either per production cycle [PPC] or the cumulative per site total over a period of years [PST] are used by W&B as a surrogate for exposure.
Response	The points above are comments about the challenges with the data. We agree with the points made.
Action	None required, the final report details these challenges.
Comment 7.	EMB use is highly correlated with maximum fish biomass and therefore, presumably, farm

	size and other operational factors. It is not therefore possible to make any strong inferences about associations between individual factors such as EMB use and potential effects on crustacean communities. I do not believe that W&B's rather firm conclusions about this association are justified by the data.
Response	Yes, but there are numerous large farms that do not use EMB. The only limitation is that low-biomass farms tend not to use large amounts of EMB (as would be expected). MaxBio was included in the modelling protocol (see Comment 18).
Action	None required, this issue is acknowledged in the report.
Comment 8.	The modelling approach used by W&B is interesting, but should be augmented by other modelling approaches, such as quantile regression, to determine the extent of any model-dependence in the results and conclusions that they present.
Response	Two distinct modelling approaches were trialled – one based on Frequentist inference, the other on Bayesian. The model selection procedure is widely recognised (see cited references). Both modelling approaches (frequentist and Bayesian) resulted in similar parameter estimates and precision (1.96* standard errors v. credible intervals). Whilst other modelling approaches can always be adopted, given unlimited resources, the approach adopted is appropriate.
Action	See below
Comment 9.	A visual scan of the data on EMB use and crustacean richness/abundance presented in Figures in the report suggests that some of the apparent effects on crustaceans may be over-stated by the modelling approach used by W&B. It also seems that at least some of the "headline figures" presented in the Executive Summary refer to EMB use rates that are uncommon.
Response	The model projections did not extend beyond the data range but did represent the approximate maximum used at farms (PPC or PST).
Action	The basis of the projections is made very clearly in the Executive summary and throughout the document. The basis of the calculations has been added as an additional table (Table 23).
Comment 10.	The "headline figures" reported here do not adequately reflect the results presented in the main report. At an application rate of 3 kg PPC the model suggests most likely results for declines in crustacean richness and abundance of 40% (see W&B page 31) and 66% (see W&B page 35) respectively. These most likely results should be stated in the summary. The same criticism applies to the presentation of data on PST EMB use: most likely results should be presented, as well as the credible intervals (I believe that these should be "most likely" declines of 66% (estimated from the factor of 3 value in W&B page 38) and 90% (see W&B page 41) respectively).
Response	The point estimates will be included.
Action	Point estimates (called 'expected…') are now included in the Executive summary.
Comment 11.	I remain confused about the distance from cages to which the above effects levels are meant to refer. For example, the report states that: "When compared to total EMB use, per Site, the effect on crustacea was more severe with 95% certainty that the decline in richness and abundance was between 21-82% and 66-98% respectively at all Distances." However, Section 3.1.4.3 of the report states that these data refer only to the Reference stations and that the "…pattern of reduction applied to all Distance Classes". This needs to be clarified.
Response	The patterns were present at all distances, but the most noteworthy are those in relation to the References stations. This is the most interesting aspect of this data analysis hence

	the focus (see Comment 2) on the Reference stations. The cubic back-transformation is non-linear and this means that the proportionate changes (when back-transformed) differ between the Distance Classes. Given the importance of the Ref-distance class, this proportionate change is the one quoted in the Executive Summary and throughout.
Action	The challenge with the data is acknowledged in the extended and revised section 8.6.
Comment 12.	The Executive Summary states that "The changes observed at Reference stations indicate the impacts are occurring at large scales. Whilst it is possible that the observed reductions in crustacea are attributable to factor(s) that are associated with EMB use, rather than directly caused by EMB use, this is unlikely given the range and nature of the variables included in the models and the systematic reductions in crustacea as a function of EMB use." I encourage W&B to reconsider these conclusions in the light of my comments later in this review, because the apparently large effects at Reference stations seem both counter-intuitive and inconsistent with the plotted data, and I am not convinced that other plausible factors which correlate with EMB use have been adequately considered
Response	The possibility of the active transportation of EMB-enriched material to the Reference stations is acknowledged by SEPA. There is a plausible mechanism - we state "The 115 day half-life of EMB means that it is likely to be dispersed very widely around fish-farms before fully degrading, particularly where sediment resuspension occurs (e.g. during unpredictable storm events or as part of unusual tidal conditions)." See also Comment 96.
Action	We have focussed our analyses on the Reference sites. We have acknowledged the deficiencies in observational research, but we maintain (and justify in the report) that EMB could, logically, be having detrimental effects at the Ref stations. The Exec summary reads "Whilst it is possible that the modelled reductions in crustacea were attributable to factor(s) that were associated with EMB use…" - i.e. we acknowledge the potential confounding of the results.
Comment 13.	W&B cite the "only significant field study published on EMB", by Telfer et al. (2006), which concluded that "there was no evidence that the occurrence of EMB, or its desmethylamino metabolite, in sediments around fish farm cages after treatment had any toxic impacts on organisms in either water column or sediments." It is puzzling that the results from W&B's desk study disagree so fundamentally with Telfer et al.'s experimental study.
Response	Telfer's study was based at one farm only. The presence of rocks prevented sampling on some stations suggesting this was an atypical, high energy, site. We also acknowledge that the data were highly variable (e.g. "Statistical modelling showed that crustacean richness was highly variable and that that this variance occurred between Sites and between Distance classes within the same survey (relatively high Obsid, Intercept and SiteID standard deviations, **Error! Reference source not found.**)" – it is entirely plausible that the results from one study, at one farm, wouldn't match our modelled mean response (from 99 sites), at mean conditions of current speed, sediment texture etc. The Telfer study did not separate the crustacea from other macrobenthos and the treatment was only 33 g EMB (208 g is the mean per-treatment EMB use in the data analysed here). See also Comment 95 and Comment 96.
Action	Results (3.1.1) and Discussion section (4.3, para 3) augmented.
Comment 14.	Not only do W&B find apparent benthic effects around fish farms; they also find effects quite some distance from those farms at Reference sites which are supposed to be beyond the limit of any effects caused by fish farms (which is why they are called Reference sites). This discrepancy between desk study and experimental data should be fully discussed later in the report.

Response	See below
Action	We have added a sentence "It also indicates that the current recommended distance-to-farm of 'Reference stations' should be re-evaluated".
Comment 15.	W&B were unable to include a measure of organic enrichment in their model because the reported results in the database are from two methods which cannot be interconverted. This means that if EMB use is correlated with organic enrichment (which seems likely) then the latter could be the cause of any observed effects on crustacean ecology.
Response	Agreed, a possibility at CE and AZE sites but substantial organic enrichment seems unlikely to occur at Reference stations. The challenges posed by the current lack of methodological standardisation are discussed. We also suggest using a count of Capitella as a proxy for organic enrichment.
Action	Recommendation in 5.2.2(b) that organic matter assessments should be standardised.
Comment 16.	No justification is provided for using the maximum recorded EMB concentration from three replicate grab samples, rather than a measure of central tendency, such as the mean, which is more likely to represent exposure. In any case, insufficient matched pairs of residue and macroinvertebrate data were available in the databases, so W&B were unable to test whether there was a relationship between measured sediment EMB concentrations and macroinvertebrate community structure. As a result of this, there is no direct evidence to suggest that EMB concentrations in sediment are associated with declines in crustacean richness or abundance.
Response	There are numerous limitations to the data and these are acknowledged in the report. In the report we fully acknowledge the deficiencies of observational-based research and recommend manipulative studies. The reviewer acknowledges that the residue data were insufficient to be used in modelling. The choice of using max EMB residue was agreed with SEPA.
Action.	None.
Comment 17.	W&B state that "As a modelling framework GLMMS are still in active development and where non-identity link functions are used (e.g. Poisson GLMM) there are questions regards (sic) the determination of standard errors/confidence intervals…" The development of novel analytical methodologies is welcome, but I question whether the results of such methods should be used to reach conclusions of societal importance without the support of additional evidence. For example, if other modelling or experimental approaches produce broadly similar results then the weight of evidence would tend to favour W&B's conclusions. However, as I mentioned earlier, the authors themselves cite the "only significant field study published on EMB", by Telfer et al. (2006) as showing no effect of EMB near fish cages. I would like to see the use of at least one other modelling approach on the cleaned-up dataset used by W&B. Quantile regression (Cade & Noon 20031, Crane et al. 20072) may be an appropriate additional technique. W&B state that "…the 50% credible interval is the best estimate of the parameter's value, whilst the 95% credible interval gives a range of values where there is 95% certainty that the value is correct." I think that one or two typos may have crept into this statement. It needs to be reworded so that it makes sense.
Response	The modelling framework (with both frequentist and Bayesian inference) is appropriate for these data, see Reviewer 1's comments. The 50% credible interval should read '50th quantile'.
Action	Report amended.

Comment 18.	W&B consider maximum fish biomass in Table 1, but it does not seem to reappear in subsequent models. Is this because it correlates so well with EMB use that one or the other (either MaxBio or EMB) becomes redundant as a predictor within the model? If this is the case, then it would be interesting to see the effect of replacing EMB with MaxBio in all models. If the results turn out to be the same then this would suggest that attributing effects on crustaceans solely to EMB use is overstated.
Response	MaxBio was included in the modelling process and, where indicated, removed from the model using the procedures outlines in 2.6.
Action	MaxBio was tried in Model development and included, where it made a meaningful contribution to the model fit. Also, see Comment 19.
Comment 19.	W&B state that "There was a positive relationship between the maximum biomass and EMB usage during any production cycle (Figure 5)". Figure 5 shows that was a strong relationship, so it is unclear to me how it is possible to tease out the relative effect on crustaceans of EMB use from other factors which may be associated with maximum biomass (e.g. farm size or level of organic enrichment). For example, would we see a very similar result if EMB was replaced by MaxBio in Tables 6 and 7?
Response	As above – MaxBio was eliminated where it was not a useful predictor in the models (as determined by Chi-square tests and AIC, see Methods). For any farm the MaxBio consented is determined by modelling carbon deposition and, consequently, larger farms should not be associated with greater impact (in relation to carbon deposition). The model fitting process consistently indicated that EMB should be retained in the model.
Action	As above MaxBio was included in the modelling process
Comment 20.	W&B state that "Statistical modelling showed that crustacean richness was highly variable and that this variance occurred between Sites and between Distance classes within the same survey." In other words, data for the response variable were very noisy, which is quite common in field surveys. Under these circumstances one has to be very careful to avoid overfitting models, or using modelling approaches which are sensitive to outliers. LOESS is a technique which is sensitive to outliers, as illustrated in the bottom left panel of Figure 7. This is why I believe that it is very important to analyse the same data with different models to see if the results are consistent. This brings me to the plots of the data which are presented in Figures 7, 9, 11, and 13. These present the data on EMB use versus crustacean richness/abundance on which the report's main conclusions are based. I would like to make several points about these plots: The data are, unsurprisingly, very noisy, and the LOESS smoothing line fits the data very poorly and is clearly influenced by outliers. I therefore disagree with the statement in the figure legends that this line "aids interpretation".
Response	The LOESS smoother does aid interpretation, particularly where numerous points overlap on the figures. The LOESS smoothers are not used in any inferential sense. Outliers were identified (by residual analysis) in the modelling phase and removed where necessary.
Action	Methods section 2.6 amended viz residual removal.
Comment 21.	Quantile regression would most likely detect a threshold inflection point on many of these plots, below which there are negligible effects and above which there are increasing effects. This is because the plots show that there were many sites with no EMB use at which there were either no or very few crustaceans. They also show that there were many sites with EMB use up to 1 or 2 kg PPC, or about 4-6 kg PST, at which there was quite high crustacean abundance or richness. Quantile regression is designed to deal with these types of datasets.
Response	The mixed modelling approach allows for sites that are inherently variable with regards

	their 'natural' background crustacean abundance (random intercept) and models Site-dependent differences in the effect of Distance (CE, AZE or Ref - random slope). The modelling process included particle size (<63 µm and >2 mm) and this is likely to be correlated to the main factor related to crustacean abundance.
Action	None required.
Comment 22.	A comparison of Figures 7, 9, 11, and 13 with, respectively, Figures 8, 10, 12, and 14 suggests that the latter do not adequately represent the former.
Response	The 'raw data' plots do not include/account for the site effects or the covariables (e.g. particle size) included in the models. The raw data were included because they help the reader understand the variability in the data. Including raw data is recommended practice (e.g. see Korner-Nievergelt et al, 2015).
Action	None required
Comment 23.	The two plots of data from reference sites in Figure 11 provide little evidence for any association between EMB PST and crustacean richness at doses up to about 6 kg for "other sites" and doses up to about 2 kg for the smaller Orkney/Shetland dataset. It is therefore surprising to see these data represented in Figure 12 by a predictive curve which drops very sharply from 0 to 2 kg EMB PST. I remain perplexed by the apparent similarity in crustacean response at all three Distances in most of the plots. For example, in Figure 7, an apparent association between EMB PPC and crustacean richness for "Other" sites occurs at around 1.5 kg EMB use, at all Distances. Surely we would expect the effect threshold to move to the right of the plot as we move from CE to AZE, and then on to Ref sites >400 m from the cages. It seems very odd that there isn't a "distance response" curve.
Response	The crustacean communities at the CE and AZE are likely to be heavily modified (e.g. via organic enrichment) but the evidence suggests that further modification occurs as a function of EMB use. The crustacean communities at the Ref, non-EMB sites, are likely to be relatively unmodified and will include sensitive species. A plausible explanation for these results is that these sensitive species decline, in response to EMB application, in the same way (i.e. same regression slope), as those more resistant communities nearer to the farm.
Action	An additional sentence has been added (Section 4.3, para 9).
Comment 24.	In the first paragraph of the Discussion W&B make a very strong statement which I believe needs to be toned down considerably in the light of the comments I have made above. I do not think that this study has effectively ruled out EMB covariables such as MaxBio
Response	MaxBio was including/excluded from the model during the model fitting process. The independent effect of MaxBio is included where necessary. See Comment 7
Action	None required.
Comment 25.	…it is not true to state that the observed effect "…was not attributable to site-specific difference and/or covariables…" It is only true to say that it did not appear to be attributable to those covariables which could be included in the model
Response	This is true, and a point that is made in the report (in relation to the deficiencies of observational research). However, given that EMB is specifically designed to kill crustacea it seems most plausible to relate the crustacean impact around farms using EMB to EMB, given that relevant covariables (e.g. particle size/current exposure/depth) were included in the model
Action	The sentence reads 'was not attributable to site-specific differences and/or covariables **including depth, sediment texture, current exposure or sampling methodology'.** (Section

	4, para 1).
Comment 26.	W&B quite rightly draw attention again to the problem that "…the SEPA data could not support an analysis of the relationship between residue concentration and macrobenthic response because of the scarcity of spatial and temporal overlap between the sampling events." This means that only weak inferences can be drawn about the relationship between EMB exposure and effects on crustaceans.
Response	We concur that correlation does not necessarily imply causation. However, given that other factors known to influence crustacean populations (e.g. particle size) were included in the model, and that EMB is specifically designed to be toxic to crustacea, we believe that the EMB linkage is the most plausible explanation for the effects. These effects were not seen in other benthic communities.
Action	The Discussion makes it clear that these results indicate, but do not prove, causation. The final sentence in the discussion reads "The evidence suggests that benthic crustacea **may not** be adequately protected by the current regulation of EMB use by Scottish salmon farms" [emphasis on 'may not'].
Comment 27.	W&B state that: "…the models predict that at an average Site, treated with 10 kg of EMB over an extended period (up to 12 years), a reduction in crustacean richness and abundance, by a factor of 3 and 10 respectively, could be expected." It is not clear to me what W&B mean here by "average". They certainly do not seem to mean "at a Site with average EMB use"
Response	At an average site (in terms of background crustacean communities and at average environmental conditions) were 10 kg to be applied over the preceding period.
Action	Additional model interpretation detail is given in 8.6
Comment 28.	Figure 2 suggests that 5 around 75% of sites used less than 4 kg EMB and that the median value is about 3 kg. Less than 5% of sites used >10 kg. I would like to see more discussion of predicted effects at different use levels, including those which are more representative or normal use.
Response	The predicted relationships with EMB are across the entire range of usages (e.g. 0 -10 kg per site, in the model figures in Section 3.1.4). The distribution of EMB, on the square-root scale, was even across the range, 10 kg was the approximate maximum used. It is fair to base predictions on this maximum.
Action	The interpretation of the models is made clear in the Executive Summary, and in the results sections. Additional information on interpretation is given in Section 8.6.
Comment 29.	I have a problem with the logic used in the final paragraph of the Discussion, in which W&B suggest that non-detection of EMB at Reference stations indicates that toxicity is occurring at below the analytical detection limit. There is at least one other interpretation of why EMB was not detected at Reference sites, although there are effects on crustaceans: perhaps EMB really wasn't there and therefore wasn't the cause of the effects.
Response	We concur that these analyses of very noisy data do not prove a causal association but we believe, for reasons given above, that the most plausible explanation is EMB. See Comment 26.
Action	None required.
Comment 30.	W&B suggest some interesting lines of enquiry, but I believe that they have omitted the most important item of further research: analyse the data with different models to see if similar results can be obtained.
Response	The quantile-regression (QR) model suggested could, potentially, give interesting results.

	However, we make the following comments about quantile regression: 1. The apparent variability in these data, as ID by the reviewer, were down to inherent differences in Site or ID_Survey and these were accounted for in the mixed modelling approach adopted (as indicated by the model diagnostics). 2. The advantage of quantile regression is really in dealing with unaccounted increased variance in the response with environmental driver (e.g. EMB) and that doesn't apply here (see point 1), there are limited data at the Site and particularly ID_Survey level, to support QR. 3. There is no mixed-model (i.e. one that accounts for inherent Site differences) QR routine available for count data (e.g. crustacean richness).
Action	None required, mixed models are the most appropriate model for the analyses of these complex data.
Reviewer 3. Note: Line references made by the reviewer are retained, and refer to the report dated 01 July 2015.	
Comment 31.	The most severe problem is with their modeling where they did not include interactions with applied EMB and other variables they used in the models. Specifically, the slope for EMB is most likely not the same for all sites, CE, AZE, and REF. They used a main effects model that assumes the slope with EMB and say crustacean richness is the same for all three sites. It also makes sense that the slopes might be a function of current speed, depth particle size and any other variables they measured. By not including interactions of effects with EMB they are most likely making big mistakes in their predictions. Not using interactions does not make sense to me as a scientist.
Response	With all the parameters being estimated the number of potential interactions was enormous: in relation to the fixed effects the total number of parameters was 9, so the number of possible interaction was 2^9-10 = 502, the number of random effects is 2^11-12=~2000. Some interactions were trialled during the model development phase but interaction terms are difficult to estimate using the modelling approach and introduce model convergence issues (in both the fixed and random effects). The objective of the modelling was to develop a logical and plausible relationship between the response and the predictors. EMB was consistently retained as an important factor in the models.
Action	The modelling rationale is justified in Section 2.6
Comment 32.	By using the very few extreme applied EMB sites they are extrapolating out past where the normal applied levels of EMB are occurring which means they are predicting out past where there is sufficient data and are most likely making mistakes in the slopes for the models tend to go through the small number of extreme points and then through the glob of the rest of the data.
Response	The spread of square-root transformed EMB application was fairly even. I did not extend the predictions beyond the data range
Action	None, the models predications are reasonably based.
Comment 33.	They used area instead of volume.
Response	We think area is more relevant, given that most infauna live in the surface of the sediment. Grab-size was tested in the models but we agree that the inconsistency in grabbing technique was one of several issues present in the data. The inclusion of grab, as an area (factor), not an offset, was agreed with SEPA.
Action	None required.
Comment 34.	They did not give a complete list of variables and or transformed variables they started

	with in order to end up with the variables they used.
Response	Table 1 lists and describes the variables used, and their transformations.
Action	None required.
Comment 35.	Is the 7.63 in line 145 correct?
Response	yes, 7.63 µg/kg wet weight is given in SEPAs 'Regulation and monitoring of marine cage fish-farming in Scotland' Annex A – Standards [see 1.1, para 1].
Action	None required.
Comment 36.	Line 160…non published results should not be relied on and should not be included in such an important report
Response	We disagree; some elements of this report are worth noting. A review including grey literature was part of the contract.
Action	No action required.
Comment 37.	Line 191. What level of EMB. It needs to be quantified.
Response	The amounts are given in the next sentence in the report (18 ng/l).
Action	Report modified.
Comment 38.	They need to be consistent in the use of the units throughout the report and not use quantities like in line 192 concentration is 18 ng/L. How does that compare to other units (note this is a volume where most others are area).
Response	Volume was appropriate given that the species under consideration was pelagic. It is entirely reasonable to refer to such literature particularly, as here, where there is little published on the ecotoxicity associated with EMB
Action	None required.
Comment 39.	Line 197. It would be good to know the dose levels used for the bioassay.
Response	Values added to report
Action	See above.
Comment 40.	Were the two data bases merged by gps coordinates? They did attempt to use ID_Survey to distinguish sites and dates sampled. What one needs to know is WHEN were the samples taken in relationship to TIME EMB was applied to the site.
Response	The databases were linked by time (date) and space (Site ID)
Action	Section 2.2 augmented with "However, the FFDB Sampling Date did not match the date in the SMDB database and these could differ by over a year (see **Error! Reference source not found.**) so these databases were joined via their Production Cycle number which was derived from the Monthly Records of biomass and EMB treatment (available in the FFDB)".
Comment 41.	Line 292. The grabs have units of area instead of volume. There has to be some depth associated with the sample. A 0.2 m2 by 0.01 m deep provides a volume of 0.002m3 while a 0.1 m2 by 0.01 m deep provides a volume of 0.001m3 I do not understand why the numbers of whatever is measured are not expressed in terms of volume as x.x µg m-3 , using identical units for the grab sizes. Since they do not normalize the grabs to amount per unit volume, they cannot combine data from different grab samples sizes.
Response	We agree that differences in grab type are a limitation present in the database. Initial analyses were based on normalised grab-areas (not volumes) or analyses where grab-type was used as an offset. However, following discussion with SEPA this approach was dropped and grab-type was included in the model and excluded (where appropriate)

	through the model selection process. Given that grab-type was the same, per ID_Survey, including ID_Survey as a random effect (if necessary) would account for the inherent grab-effects.
Action	None, the adopted protocol was agreed with SEPA.
Comment 42.	Line 290 they talk about taking 2, 3 or 5 reps. My question are these reps or subsamples.
Response	These are pseudoreplicates hence the data were combined across them. They were, appropriately, not treated as independent samples
Action	None required, the analysis is valid. The protocol is described in 2.4.4.
Comment 43.	Are they taken at randomly selected positions for CE and randomly selected position for AZE and then at randomly selected positions for the REF.
Response	Almost definitely not. They would be taken along the axis of the main current.
Action	None - these results apply to the sampling stations and this is acknowledged e.g. in the Exec. Summary the word 'indicating' is used "At Reference stations, on a PST basis, the EMB –crustacean association was more significant with expected reductions in richness and abundance being 64% (21 - 82%) and 96% (74 - 100%) **indicating** wide-scale, cumulative impacts and incomplete recovery between successive EMB…" We do not infer outside the data range.
Comment 44.	My experience is they take samples at the same approximate site which are subsamples and not replications. Also they make not mention of using the depth of the samples –
Response	Depth was tested as a covariable, see Table 1, and included in some of the models.
Action	None, Depth, as a factor, was including in the modelling process.
Comment 45.	It seems like a lot of stations were excluded based on distance or locations without knowing the characteristics of the sites.
Response	This is true, but there is no reason to think that the excluded Sites introduced bias i.e. that the excluded sites, where EMB was used, hosted more diverse or abundant crustacean communities.
Action	None necessary.
Comment 46.	Also there are a lot of cage edges and *they did not indicate they documented which was used or did they just chose the easily place to reach?*
Response	Cage-edge samples are well represented because they were usually better coded as such in the database. We do not understand the point being made in italics?
Action	None.
Comment 47.	Is it normal practice to not normalize to area (Volume) sampled?
Response	The modelling protocol was agreed with SEPA and Grab included in the model.
Action	None required.
Comment 48.	The authors do not seem to understand that when grab area is included in the model that it means the response is linearly related to grab area, not that the model is removing the effect of grab area.
Response	Agreed, the effect of grab-type is accounted for in the model.
Action	None, the modelling protocol WRT grab was agreed with SEPA.
Comment 49.	It seems they measured depth and current speed were determined, but they are not stated as to exactly where they are measured---fastest current area, deepest place etc..

	Also slope characteristics under the cage should be important.
Response	Agreed, depth was measured at every grab (so good data) but the current speed was only measured relatively close to the cage. We acknowledge that the data deficiency in this regard but note that particle size, a good proxy for current speed, was measured at each grab.
Action	None, the modelling protocol was agreed with SEPA, the deficiencies of the data are well known, other data were not available/accessible.
Comment 50.	The use of the maximum EMB [residues] of the three sub-samples is biasing the results upward. Did they state in their protocol that they would use the maximum or did they just discover that the maximum was the only metric that show any relationship with their other characteristics?
Response	The choice to use Max EMB was pre-cautionary (and agreed with SEPA), but the residue concentrations did not form a significant part of the reported work because of a lack of spatio-temporal overlap with grab-sampling.
Action	None.
Comment 51.	Section 2.3.6. They say that the LOD differed from site to site and data set to data set. If one would include the methods LOQ along with the data's LOQ there are statistical methods that have investigated good ways to handle these data. A proper protocol would have a well-established method that would provide a sensitive LOQ that below which there is no effect.
Response	See comment above, there was virtually no spatio-temporal overlap between residue and macrobenthic sampling. We agree that the LOQ should be well below the NOEL – our analysis indicates that the LOQ is much greater than the NOEL which is an important conclusion (and comment is made to this effect in the Discussion).
Action	None.
Comment 52.	Line 398 Replace "broad interpretation of results" to "broad as possible interpretation of results". Their sites are not a random sample of all possible sites is Scotland, so they cannot represent all sites!!!
Response	There was no systematic exclusions of sites so, effectively, they were random. They include sites from the entirety of the fish-farming area along the west-coast.
Action	None – the observations are effectively a random sample of fish-farms.
Comment 53.	Line 421. Bayesian is not in general the more intuitive method among the class of statisticians. There are many more Frequentists than Bayesians in the world. In fact the authors use a mixture of Frequentist and Bayesian language in their report. They use confidence intervals for ED50's and then use creditable intervals and try to explain them by kind of relating to confidence intervals. Then they talk about testing hypotheses about the fixed effects and in their models and they provide significance levels that are Frequentists. Then they use creditable intervals to get predictions from their models.
Response	The confidence intervals for the ED50's were reported in the literature, I cannot generate Credible intervals for them. Mixed models with Bayesian inference is a modern, published (see Korner- Nievergelt et al in the Reference section) approach to analysing what is a complicated dataset. A review of the relative merits of the Bayesian and Frequentist approaches is beyond the scope of the SARF098 report. The rationale for using a Bayesian model to derive credible intervals is given in the report and the meaning of the credible intervals concept is much easier to convey compared with confidence intervals.

Action	None.
Comment 54.	Line 459. The interpretation of the fixed effect estimate is what should happen across the population of sites. One can get a predicted variable to see what is happing at a single site by using a predictable function. It is a matter of broad, intermediate or narrow inference. They have the incorrect description. I do not understand their discussion.
Response	Mixed models are not marginal models (e.g. as Generalised estimating equations are – GEE were trialled but currently they cannot model 'random-slope' models). The fixed effects parts of the model relate to the effect on the 'average site', this is the 'population model' (see Zuur, 2009, page 108), the effect at individual sites can be added to give the predictions at individual sites but this was not done because we are not interested in particular sites (hence they are included as a random term – and the normality of the random-effect residuals was checked).
Action	The description of the models is extended in 8.6.
Comment 55.	They do not achieve this task as they do not look at the relationship between EMB and community metrics. Their biggest error is that they do not look at the interaction between EMB and other covariates included in the model.
Response	This is the same comment as above Comment 31 – we could not test all conceivable interactions in the model design phase.
Action	The modelling protocol is explained in Section 2.6.
Comment 56.	In table 6 they use the same slope for EMB for each site, CE, AZE and REF. It does not make sense for the slopes with say crustacean richness and applied EMB at the AZE site to be the same as the slopes for the CE and REF sites as the actual concentration of EMB is much different at the three sites. Assuming the relationship between the applied EMB and say crustacean richness is the same at the three sites does not make science sense.
Response	WRT Table 6, it is a summary of the raw data, I'm not sure what the question is here. With regard the common slope, this was also made by another reviewer, see Comment 23.
Action	See Comment 23.
Comment 57.	Figure 7 is an example of allowing extreme values to influence the model. There are about 9 values at about 3 kg/production cycle (2 or three per graph) If those 9 observations were removed from the samples the loess curve would be approximately level.
Response	Fig 7 shows raw data. The EMB were square root transformed prior to inclusion in the model (so as to spread out the data across the range). Model predictions are based on back-transformations of the square-root transformations. Even not allowing for the covariables (as the modelling does) the downward association between EMB (per PPC) and Crustacean richness is apparent in the Fig 7 and 9.
Action	None, the data were appropriately transformed prior to analysis.
Comment 58.	Table 8. What is the difference between AZE and int:AZE in the random effects part of the table
Response	In Table 8 AZE is the standard deviation of the response at the AZE in the random effects (because this is a random slope model we have both random intercepts and correlations between the categorical factors in the random term, in this case the intercept is the CE so int:AZE is the estimated correlation between observations, at the same Site, between the CE and AZE) and then intercept AZE REF (but not CE) are in table for fixed effects. The intercept is the base model and, in this case, that means it is the prediction at the CE.
Action	Additional description of the models in given in 8.6.

Comment 59.	They need to write out the exact mathematical model with model assumptions being used…with the random effects and the fixed effects and all interactions used
Response	Model assumptions were checked, extensively, with residual plots. This is detailed in Section 2.6. For each model the random and fixed effects are detailed in the table caption, the transformations applied to each factor are given in Table 1.
Action	The model is given in each Table caption e.g. Table 8 - Crustacean richness (per production cycle) Model: Richness ~Distance Class + EMBSR + Depth + LT63+Speed+Area +(ObsID) + (Distance Class|Survey) + Site. Poisson GLM with log-link function.

Reviewer 4.

Comment 60.	The report presents the general conclusion of a reduction of crustacean abundance and crustacean richness in areas treated with emamectin benzoate (EMB). The product Sponsor recognizes that a retrospective analysis across multiple databases is challenging to conduct, and appreciates its ability to generate correlations and hypotheses. This modeling may be limited in its inference on the contribution of EMB, especially with regard to the estimated magnitude and extent on the outcome variables. Using this report's premise, with clarification of select model assumptions, can lead to prospective, well controlled studies to more definitively enumerate the magnitude of effect related to use of the approved pharmaceutical product. In that regard, a set of topics arise as the key to understanding the data and retrospective model:
Response	We agree, manipulative studies are a logical next step, see Recommendations (5.1 -point 4)
Action	See recommendation (as above)
Comment 61.	Grab Samples were collected from cage edge (CE), a variable distance at the allowable zone of effects (AZE) and a Reference (REF; >400 m from CE). A consistent finding was a dose-response reduction in crustacean abundance and crustacean richness at the REF, even though detectable levels of EMB were rarely found after 25m (Figure 3). The conclusion of the report was an effect at lower concentrations than originally considered (section 4.1). Before such a conclusion can be confirmed, this reviewer would like to consider the directionality of the REF. This was likely placed along the prevailing current to maximize any residual EMB reaching the REF location, even though there was no found relationship between EMB and current speed (section 3.1.1). To understand if non-detectable but non-zero EMB actually reached this distant REF location and attributed to the decline in crustacean abundance and crustacean richness, the other 3 major compass points would ideally also be grab-sampled to show a lack of decline; ie, confirm lack of effect in crustacean richness and abundance in directions not in the line of the current.
Response	Agreed, but other compass points really reflect different distances. We believe the sampling programme should be conducted in relation to model predicted carbon/particulate deposition in addition to linear distance.
Action	None necessary
Comment 62.	2) If current directionality is a factor, a corollary question arises to determine how distant from the cage edge is the no-effect position.
Response	We believe that a no-effect position doesn't exist, impacts will occur on a continuum.
Action	We question the location of the Reference stations in light of our findings (penultimate

	paragraph, section 4.3).
Comment 63.	However, if dispersion is not related to the current, then factors included in the model (Tables 8-11) may need to be re-evaluated, especially for REF and AZE.
Response	Not sure of the point being raised here – dispersion will be related to the current and this was included in the model.
Action	n/a
Comment 64.	The conclusion of EMB as the cause of the directional decline is based on the dose response relationships shown in report Figures 8 and 10. While the Sponsor concurs that EMB could be a likely contributor, other fish husbandry by-products including non-consumed feed and body wastes, can also be contributors to the noted declines
Response	Agreed, they will, and this was found i.e. declines in crustacean metrics at the cage edge cf AZE cf Ref. However, independently of these, additional EMB-associated effects were also found.
Action	None, we maintain that EMB is the *likely* cause (but we accept the limitations of observational research).
Comment 65.	Parsing confounded variables is challenging without demonstration of the lack of decline in crustacean abundance and crustacean richness at nearby fish farms that had similar feed and husbandry, without the use of EMB.
Response	No confounding factors were identified i.e. there was nothing inherently different about farms using EMB and those that were not. We gave this issue considerable attention and discussed with SEPA at length.
Action	None, there is no basis for thinking that the data is confounded in this way.
Comment 66.	The SEPA Fish Farm database (FFDB) and Self-Monitoring database (SMDB) were merged to the extent possible by the Report Authors (section 2.1). These databases were pre-existing for different monitoring goals, and repurposed for this analysis. As such they are useful for detecting correlations, though conclusions of cause-and-effect are not possible.
Response	Agreed, definitive conclusions are never possible when sampling from populations but we believe the most plausible explanation for the observed decline in crustacea around farms using EMB is EMB. Note that the final sentence in the Exec summary is "The evidence suggests that benthic crustacea **may not be** adequately protected by the current regulation of EMB use in Scottish salmon farms".
Action	None. Our wording is appropriate, inappropriate inference is not made.
Comment 67.	A properly designed study could be used to evaluate cause-and-effect which cannot be achieved with observation studies. This comment is aligned with the Authors' recommendations numbers 4 and 6 (Section 5.1)
Response	Agreed.
Action	n/a
Comment 68.	The Authors demonstrated similar patterns for Per Production Cycle (PPC) analyses and Site Total EMB. Reviewer comments will equally apply to both. A total of 1235 residue measurements from 271 sites were extracted from the database (line 486), which constitutes a sample so large that small findings can result in statistical significance even in the absence of biologically important differences.
Response	Not so, our credible intervals indicate the likely magnitude as well as the 'significance' of the effect
Action	None, the inference is appropriate.

Comment 69.	Likewise, the total number of grab-samples was 1259 from 6 regions and 99 sites (Table 3). This does not diminish the biological significance with the reduction of crustacean abundance and crustacean richness; however, the statistical models may be so sensitive that variables with minimal contribution are displayed as significant
Response	Not so, we are not null hypothesis significance testing – effect sizes and credible intervals form the basis of our interpretation and conclusions.
Action	None, the inference is appropriate.
Comment 70.	This was especially noted where the p-value shown for EMB was significant at $p<0.05$, but was the least significant of factors included in the model. For example, Table 8 (crustacean richness per production cycle) had 7 of the 8 fixed factors as statistically significant, ranging from EMB (minimally significant at p=0.04) to REF (highly significant at p=1.65e-44). From this outcome, EMB cannot be dismissed as a component, but is deemed to be the least important contributor.
Response	We don't interpret P values, we interpret credible intervals, see Nuzzo, R., 2014. Scientific method: Statistical errors: P values, the 'gold standard' of statistical validity, are not as reliable as many scientists assume. Nature 506, 150 – 152. We used chi-square tests to assess whether EMB should be included in the model.
Action	None, the inference is appropriate.
Comment 71.	It is unclear to the Reviewer why the statistical model used for Crustacean Richness (Tables 8 and 10) was not equivalently used for Crustacean Abundance (Tables 9 and 11). Conceptually, two continuous observation variables (richness and abundance) would be affected by the same environmental factors, including the potential effect of EMB.
Response	Richness is clearly a discrete variable, not a continuous one. This choice of model (generalised or general) and justification for transformations is clearly given in Report Section 8.6 – Statistical models.
Action	n/a
Comment 72.	The abundance model, however, dropped the effects of grab area which were shown to be minimally statistically significant in the richness analysis. Additionally, the abundance model switched to a t-value as compared to richness using a z-value, with the abundance model not displaying the p-value (which would demonstrate the level or lack of statistical significance).
Response	Our models were interpreted on the basis of their credible intervals
Action	None – the modelling approach was correct.
Comment 73.	Statistical model building, like clinical trials, are most fairly conducted when an *a priori* protocol is defined and followed. The report mentions the pathway followed (section 3.1.4 and 8.6), and discussion of approach with SEPA (lines 110-112), but it not clear if this was a definition of formal models or a general framework with updates after each model iteration. Without a predefined roadmap, modelling often follows a path-of-least-resistance or a random-walk, and its conclusions are not as robust as a controlled rigorous pre-planned progression.
Response	Agreed, this study was based on model optimisation and these results are not definitive. We agree that manipulative studies would be required to better infer from the cause (EMB) and the effect. However, in our opinion the most likely reason for the association between EMB use and the decline in crustacean abundance /richness is EMB, given that EMB is specifically designed to be toxic to crustacea and is known to remain for extended periods in

	sediments around farms using it
Action	None, the inference is appropriate.
Comment 74.	General linear modeling (GLM) can be designed retaining many unique variables in a model. The Authors in Table 8 and 10 used separate variables for AZE and REF (with exclusion of CE from the model as derivable from the other levels). Same with their inclusion of the substrate sieve size and area. This treats each presented level of each parameter as an independent continuous predictor variable in the model.
Response	Distance Class is a categorical predictor, CE is represented in the base level (i.e. the intercept). By substratum sieve size I think the reviewer means proportion passing the 63µm sieve, a variable which was treated as a continuous predictor.
Action	The variables used in the model are clearly laid-out in Table 1.
Comment 75.	However, levels within a parameter are not independent and a class effect would seem more appropriate to yield one parameter estimate per variable. Each parameter can then compare each level of the predictor with a reference level, typically the last level in sorted order. The order enables determination of a linear effect with increasing levels of the variable.
Response	We need clarification from the review to address this point.
Action	n/a
Comment 76.	Section 2.5 discussed the Author's preference for a Bayesian inference. The discussion section, however, did not explore an estimate or basis for an informative prior and/or loss functions, or show how the models would have differed with a Frequentist approach.
Response	Uninformative priors were used. Results from Frequentist and Bayesian models, both of which are given, were nearly identical and this is because the sample size was reasonably large. Interpretation/reporting of Bayesian credible intervals is more intuitive than Frequentist confidence intervals and, as stated in the report, there is ongoing debate about estimating standard errors in mixed models. This was one reason to use Bayesian inference (see reference in report).
Action	Methods section modified to include details of priors (See 2.6).
Comment 77.	Variables <63 µm and >2 mm are not independent, as assumed in a GLMM. All sediment passing through the <63 µm sieve must have already passed through the >2mm sieve, and constitutes a nested subset. The effect on the GLMM is not clear.
Response	Point accepted, but >2 mm was never included in any model. See recommendation 5.2.2(point c).
Action	None, the modelling procedure was appropriate.
Comment 78.	Throughout the report, the Authors noted a decline in crustacean abundance and crustacean richness. While not absolutely related as the sieved sediment, the 2 variables are highly correlated and not fully independent. The Sponsor concurs that separate analyses of these variables were warranted.
Response	In fact the high correlation between the richness and abundance (as is to be expected) yielded similar patterns even though different models were developed (Poisson and General linear models) and this supports the overall findings
Action	None, the analysis is appropriate.
Comment 79.	Data transformations are common in statistics to control for presumed underlying distributions and/or to normalize data to better fit modeling assumptions. The Sponsor concurs that transformations can ease interpretation, including the Authors' centring

	continuous variables by mean-subtracting. The distribution of actual meter depth and current speed were not bellshaped but right-skewed (Figure 5), though no figure was provided after log transformation to determine its effect on normalizing.
Response	Fair point, for the Draftman's plots I wanted to present the raw data so the reader could see the distributions in the native measurement scales. There are no distributional assumptions made about predictor variables (they can be any shape), the log transformation was to 'spread-out' the data (so it was less dominated by extreme values). This is routine practice.
Action	None.
Comment 80.	Centring shifts the mean depth or speed to "zero" and express other samples as positive or negative differences from the mean, but would not further stabilize the model. EMB levels, however, were square-root transformed (section 3.1.4.1). It is not mentioned why the logarithmic relationship *a priori* chosen for depth and current speed was not applied for EMB.
Response	This is explained in Section 8.5 Transformation. Log zero is undefined and there were, of course, numerous instances where EMB was not used (precluding a log-transformation).
Action	None.
Comment 81.	Square-root is a common transformation for area (Nicholas J. Cox, Durham University, http://fmwww.bc.edu/repec/bocode/t/transint.html) or counts (Handbook of Biological Statistics, John H. McDonald, http://www.biostathandbook.com/transformation.html). EMB levels were also right-skewed (Figure 5). EMB is shown square-root transformed in Figure 5) though not as extensively as depth or current speed. The cited references consider log transformations, squareroot and cuberoot as controlling for right-skew, with the primary advantage of square and cube root over logarithm when the data can have a negative value or zero which does not apply to the raw data for any of these variables though roots may have been chosen to offset the negative values resulting after centring.
Response	The EMB concentrations and crustacean richness and abundance counts were frequently zero meaning a log transformation would need to include adding a constant. Regards centring, the protocol was to transform then centre.
Action	None, the data transformations are given in Table 1 and explained in 8.5
Comment 82.	McDonald states an advantage to logs since independent factors multiplied together have a resulting product of lognormal, and the log-transformation normalizes for statistical analysis. Investigating the consistent use of a log-transformation for all continuous fixed factors may be a model consideration.
Response	The influence and interpretation of transformed variables in models is a live issue – the main problem I had with using a log-transformation on the crustacean abundance is the arbitrary nature of the added constant (e.g. Log+x, what value of x?)
Action	None, cube-root transformation is justified.
Comment 83.	13) Cube root was used for the 2 outcome variables of interest: crustacean richness and crustacean abundance. Similar to squareroot, cube root linearizes volume (Cox), which does not directly seem to apply to these variables
Response	Poisson regression was used to model the mean richness, with a standard log-link function. The cube-root transformation was required to stabilise the residuals in the abundance-models. See Comment 5.
Action	None.
Comment 84.	The effects of a third type of transformation within the same model will not likely change overall outcome of the model, but can result in questions on the accuracy of magnitude

	estimated or the contribution of a variable that minimally crosses the p=0.05 threshold.
Response	Yes, it will down-weight the outliers and reduce the back-transformed mean, this is shown in Table 17- Examples of transformations. The P value is not used in interpretation here
Action	None.
Comment 85.	Draftsman plots (Figure 5, 6, 7, 9, 11) show relationships between variables with best fit association lines, mixed with respect to transformed and non-transformed data. The relationships with original scale data, and then transformed data, should be separately displayed, to enable interpretation on the value of the transformation as well as relationships between variables. Many of these fit lines are extremely non-linear and it is unclear on how they were used to define the final statistical model presented.
Response	The transformation/non-transformation selection was based on displaying the core attributes of the associations whilst maintaining a reasonable length to the report. Only the Response metrics (Figure 6) were extremely non-linearly related and this is not relevant as response metrics were not simultaneously modelled. An explanation/justification of back transformations is given in 8.5.
Action	None.
Comment 86.	Few sites had total usage of EMB above 1 kg (Figure 5, 15, 16).
Response	Numerous sites had treatments of >1 kg (see Figure 15 and e.g. Figure 9 and this is tabulated in Table 6 and 7).
Action	None
Comment 87.	Drawing a dose-response relationship as evidenced in Figures 8, 10 and 12, is dependent on sufficient observations at the higher EMB administration level. If only a couple sites out of the 99 actually used a high quantity, then the statistical model would converge on those few sites instead of being reflective of general effects. Without those sites, it would be worth investigating if the noted pattern still appears since the Authors' claim (section 4.3) that the effect is non-linear
Response	Ok, but as shown in Table 6 and 7, there are numerous observations/sites where >1 kg EMB was used. The dose/response is non-linear on the original scale, but linear on the square-root scale which is why I transformed the EMB dose.
Action	None, the interpretation is valid.
Comment 88.	with the rate of change decreasing as the amount of EMB increases, the noted effect may be less dose dependent and more in line with the Author's conclusion that low-dose (and higher dose) rates would result in a similar level of reduction, if EMB is confirmed to be a causative agent.
Response	An alternative explanation is that at low doses EMB eliminates sensitive species leaving less sensitive species or species which are able to recruit e.g. on an annual basis from the water column. Hence the species-specific line of enquiry – the next stage of this analysis and as suggested in the recommendations
Action	None.

Reviewer 5

Comment 89.	Many of the studies cited within the introduction as evidence for toxicity of emamectin benzoate (EMB) were conducted using methodology or species that are likely poor surrogates for benthic infaunal crustaceans. For instance, data cited as Fiori (2012) are from a non-peer reviewed study and conducted with pelagic copepods, not benthic crustaceans. Further, the aqueous exposure employed in the Fiori (2012) study requires

	the SARF report authors extrapolate data to a sediment exposure: "EMB concentrations of 0.018 µg/l, between five and fifty times lower than the detection limit (per kg of sediment) as indicated in the SEPA database, have been shown to have a serious effect on pelagic copepods (Fiori, 2012, not peer-reviewed), over an 8 day exposure period." (pg 50). In contrast, data from Tucca (2014) show toxicity to a marine amphipod (*M. insidiosum*) of 890 ug/kg (sediment exposure) as a 10-day LC50. This is on the same order of magnitude as other marine sediment-dwelling crustacea and intermediate between polychaete data (see Table 4 in Tucca 2014), and hence, is more relevant than Fiori (2012) for the assessment. Several studies cited within the report focus on the sensitivity of large, mobile epibenthic crustacea, such as lobsters and shrimp (i.e., Waddy et al. 2010, Veldhoen et al 2012). Key differences in feeding strategies, home and/or feeding range, and routes of exposure render these species poor surrogates for small sediment-dwelling crustaceans. The SARF Report authors also proposed that injury to important shrimp and lobster fisheries could be a possible social impact of the use of anti-louse chemicals (including EMB). However, this is based on reports of observations from fisherman in Canada (Wiber et al. 2012), which are likely to be biased and alone cannot be considered reliable evidence of degradation. In fact, Waddy et al. (2010) reported that lobsters given a choice between "clean" and EMB-medicated feed strongly preferred the clean food. Hence, there is little evidence from the literature (as presented in the report) to support the concerns of impact to crustacean fisheries.
Response	1. the sensitivity of crustaceans, in general, to EMB is very relevant to this study regardless of whether they are benthic or pelagic. The review of 'grey' literature is warranted given the general lack of information regard EMB use (and such a review also formed part of the original proposal). Our noting that crustacean fisheries species may be negatively impacted by EMB is fair and particularly relevant given that such fisheries frequently overlap, spatially, fish-farming. The comment that American lobsters prefer un-medicated food is not particularly relevant unless consumption of feed, lost through the nets, is considered the most likely method that EMB enters the environment (and this is not the case, EMB is considered to enter the environment via excretion in faeces). We think our comments are fair.
Action	We think our review is appropriate.
Comment 90.	A number of sampling methods were employed during SEPA monitoring, which complicates analysis of the resulting field data. Grab samples of different sizes were used, which can result in significantly different measures of community health. Further, certain benthic community metrics may not be easily "scalable" when different areas of sediment are sampled. In addition, the depth of benthic samples is not stated in the report (although it may be in the underlying data). If this is also variable between sites and/or samples, it could result in the sampling of notably different components of the benthic infaunal community. As identified by the SARF Report authors, there is a strong correlation between sampling station and the specific sediment sampling equipment employed, which is a significant confounding factor for any analysis conducted with this data. Since different sampling equipment was employed at different sites, it may be difficult separating site effects and effects of sampling equipment. In fact the report notes "There were apparent differences between the diversity metrics determined from the data as a function of grabbing protocol (total grab area) and sampling region" (pg 28). It should also be noted that there was a significant imbalance of the number of samples collected within different SEPA regions, which could also adversely impact the statistical analysis.
Response	We agree that the lack of standardisation in the collection of the data disadvantages the

	analysis; one of our recommendations is that this be rectified. The grab-effect was included, following consultation with SEPA, into the model as a factor and was, therefore, accounted for. Depth was included in the model, it is shown in Figure 5 and detailed in Table 1. A standardised grab-size would have made the analysis more robust but given that EMB was used, in varying amounts, in all regions and using all grab-types, the models were able to account for this non-standardisation. With the exception of North Aryshire, there were numerous samples from each Local Authority Region.
Action	Grab was included in the modelling process with SEPAs agreement. Region was designated to model (if necessary) regional differences that were not accounted for by the environmental parameters.
Comment 91.	Most critically, the report states that "the SEPA data could not support an analysis of the relationship between residue concentration and macrobenthic response because of the scarcity of spatial and temporal overlap between the sampling events". Given that currents and sedimentation can result in mixing and movement of EMB residues, concurrent chemical biological sampling is necessary to understand the relationship between sediment EMB concentrations and field benthic community changes. The Sediment Quality Triad approach (Chapman et al., 1997) is a an example of a comprehensive sampling and analysis methodology commonly used to assess injury and identify key drivers of injury using synoptic measurements of sediment chemistry, laboratory toxicity, and benthic community structure and health. The lack of concurrent chemical and biological sampling is a key deficiency of the underlying data, which severely restricts the power of statistical assessment and limits the ability to determine the primary driver of observed changes in benthic community structure.
Response	We agree that these data do not support the analysis of the relationship between EMB residue and crustacean metrics. Our analysis indicates an association between EMB use (not EMB residue) and declines in crustacean metrics. We make it clear that this association requires further investigiation (e.g. via manipulative studies) but, given the toxicity of EMB to crustacea, it is fair to speculate that the association is caused by the toxicity of EMB to crustacea.
Action	None, our inference is appropriate.
Comment 92.	The authors reported that the results of the Bayesian analysis indicated significant changes in benthic community health at reference stations in regions with significant EMB use. Based on this, it was hypothesized that use of EMB was resulting in widespread (i.e., sea-loch scale) effects on benthic crustacea. However, this conclusion does not take into account several key sources of uncertainty that significantly weaken this line of reasoning. First, there is little detail on the appropriateness of reference stations; as such it is unclear if they were merely selected based on distance or if they are true Reference Stations (e.g., matched in terms of sediment characteristics and hydrodynamic regime to the treated sites). Although this information may be in the underlying data or specified in the requirements of the monitoring program, it is not mentioned in the SARF Report. High variability of crustacean richness and abundance was noted to be ". . .exacerbated by the range of sampling strategies…and unknown covariables such as the degree of organic enrichment/oxygen depletion in the sediment." Given this, it is critical that reference station appropriateness is clearly assessed and reported. Further, the SARF Report authors report that "the main focus of the interpretation of the models is on the effect of EMB, not the covariables" (pg 17), but it is unclear to what degree covariables were assessed and

	accounted for within the analysis.
Response	Reference stations should match the farm-site in terms of sediment characteristics (as per SEPA guidelines). The effect of sediment texture (e.g. %<63µm) was tested in all the models. The variability attributable to organic enrichment is only really relevant to the CE and AZE stations, during the model selection process all the covariates were considered and dropped according to the protocol outlined in the methods (section 2.6). The main focus of the interpretation is on EMB because that is the primary focus of the study, we were not particularly interested in the relationship between, for example, crustacea and %<63µm but we need to account for this relationship in the model.
Action	None.
Comment 93.	Also of concern is the fact that the analysis found no relationship between current velocity and EMB concentration or detection. This seems very unlikely to be real, as deposition of feces and uneaten feed is the primary means of input of EMB to the environment, and this is highly influenced by current velocity and direction. Further, EMB was not present at detectable concentrations in a large number of samples (approximately 30 to 40% of total samples) and "EMB residues were mostly not detected at 25 m distance from the cage-edge and, by extrapolation, are highly unlikely to be detectable at the Reference stations". The fact that EMB concentrations were not measured at reference sites is an additional source of uncertainty in assessing appropriateness of "reference' designation. The lack of quantifiable EMB sediment concentrations, coupled with the apparent adverse effects predicted at reference stations, led to the following conclusion: "It is entirely plausible, therefore, that chronic exposure to EMB, even at currently undetectable concentrations, may have serious consequences for crustacean communities" (pg 50). While technically plausible, it is not shown that this explanation is more plausible than other potential causes. Further, 'plausibility' is insufficient for identifying the key chemical drivers of ecological degradation. Other plausible explanations might include other, non-point source pollution, altered physical environment (i.e., changes in water temperature, salinity, sediment deposition or erosion, organic enrichment, etc.), or the use of other chemotherapeutants.
Response	The data regards EMB residue concentration was insufficient for any robust assessment. We agree that an extended monitoring programme, for EMB residues, would help assess the relationship between EMB residue and benthic crustacea (see Recommendations, 5.1). To our knowledge, there is no confounding in relation to e.g. point-sources of pollution, temperature or salinity, for example, there is no basis for thinking that Reference stations, that are associated with farms using EMB, will be commensurately impacted by any other source of impact. We are not aware of other chemotheraputants that are used in conjunction with EMB.
Action	None, our inference is appropriate.
Comment 94.	Given the identified issues with data quality, reference station responses, and chemical detection levels, the SARF Report authors were correct to include an analysis of individual sites that compared conditions prior to and following initiation of EMB use. These types of analyses are common tools used to quantify the effects of environmental stressors, and are commonly referred to as Before-After-Control-Impact (BACI) studies. However, due to aforementioned issues with data quality, only a subset of stations had data prior to and following the introduction of EMB at said site; the authors reported that only three could support a statistical analysis (FFMC47, LINB1, TAI1). The results of the pre-EMB/post-EMB analyses indicated that there was no obvious pattern to crustacean response following initiation of EMB use. While the SARF Report authors stated no conclusions could be

	drawn for these sites due to limited number, the information from these Before/After sites constitutes an important line of evidence that calls into question the plausibility of widespread chronic effects following exposure to EMB concentrations below the detection limit. At a minimum, it indicates the need for careful consideration of alternate stressors and contaminants as drivers of benthic community changes.
Response	Alternative sources of contaminants and confounding were discussed at length with SEPA. None were identified. We acknowledge the deficiencies of observational-based data analysis.
Action	None, our inference is appropriate.
Comment 95.	In the SARF Report conclusion, the authors outline the proposed mechanism of EMB effects as follows: "Given that EMB is toxic to crustacea, is found around farms using it and the relationship between EMB use and crustacean response reported here, we believe that the most likely explanation for the association between EMB treatment and crustacea is because of a direct toxic effect." (pg 52). "These data indicate that, even at low dose rates, EMB will cause a reduction in crustacean richness and abundance; there was no evidence of a threshold beneath which change did not occur." (pg 53). However, these conclusions are not consistent with findings from previous field studies. Telfer et al (2006) reported that there was no evidence that EMB use adversely affected sediment communities in the vicinity of treated fish farm cages. In an unpublished thesis, Mavraganis (2012) noted impacts on sediment infaunal communities, but only at sites with "significant levels of SLICE". Other field studies conducted by Intervet at sites in France (Barnaud et al. 2002) and Norway (Wallace et al. 2004) further support the lack of impact. Together, with laboratory studies on the sediment toxicity of EMB (i.e., Mayor et al. 2008, Tucca 2014), these field studies indicate that environmental impacts occur at sediment exposures that far exceed the analytical detection limits.
Response	Regards the Telfer study see Comment 13. Given the high degree of variability within benthic communities, and that long-term exposure may be responsible for the changes we modelled, it is not surprising that some authors have not found 'significant' results. The Telfer study was based on a 33 g treatment, much smaller than most of the treatments reported in the current study (where the mean is 208 g per treatment and there can be multiple treatments per production cycle). Our contention is that there is evidence that EMB has effects on reproduction at very low concentrations and that this might impact brooding species where immigration (e.g. from the planktonic larvae) will be limited. This is a logical and reasonable working hypothesis, based on the data analysis and literature review. Regards Bright, this is a review, from 2002, (i.e. does not offer new data), and the main conclusion, in respect of emamectin, is The data are limited to that derived primarily from studies on acute toxicity, rather than chronic or sub-chronic effects on mortality, fecundity, reproductive success, growth, or other sub-lethal responses. Figure 7-1 shows the distribution of ecotoxicity data for EB This supports our contention that the impact of chronic exposure of crustacea, to EMB, is unknown. Also, on page 38, it is noted that meiobenthic crustacea (including juveniles of macrobenthic species) are at risk from EMB. The Barnaud et al (2002) and Wallace et al (2004) research is not available to us.
Action	See Comment 13 and Comment 95
Comment 96.	Also of importance is the finding that crustacean species are not always the most sensitive to EMB. Telfer et al (2006) noted that "annelids were the most sensitive to the presence of

	emamectin benzoate". Also, the SEPA EQS of 0.73 ug/kg is based on the observed high sensitivity of the polychaete, *Arenicola marina* (Bright and Dionne 2005). In general, sediment dwelling crustaceans are sensitive to a number of toxicants and stressors, including metals, PAHs and organic enrichment. Therefore, in absence of environmental data collected under carefully designed sampling plans, it can be difficult to ascribe declines in crustacean infauna to particular stressors.
Response	Crustaceans assemblages are responsive to a broad range of stressors. We, with SEPA, carefully considered sources of confounding i.e. where, for some reason, EMB-treated farms might also be subjected to other sources of impact. We, with SEPA, could not identify any sources of confounding. Given that EMB has been designed to be toxic to crustacea, we believe that the pattern of crustacea metrics observed around EMB-using farms is mostly likely to be attributable to EMB. Telfer found EMB residues in mussels located 100 m from the farm, one week following a small (33 g) treatment (this compares with the average treatment rate of 208 g over the period 2002 – 2014). This indicates, contrary to the interpretation of the reviewer, that EMB is dispersing widely around the farm.
Action	Report amended 3.1.1 "The amount of EMB used, per treatment (of which there could be several in any production cycle), ranged between 0.378 and 1474 g. The mean EMB treatment was 208 g".