SARF – FIRST REVIEWS FOR INITIAL SARF098 DRAFT REPORT

(*Extracted from PDF files to preserve anonymity*)

NOTE: Referee designations are the same as used in the second review phase, but do not necessarily equate to the designations used by the report's authors in their summarised response to first reviewer comments.

REF01

REVIEW OF SARF098: PAMP REFRESHMENT STUDY FINAL REPORT BY WILDING AND BLACK (HENCEFORWARD W&B), DATED 1ST JULY 2015.

The W&B report mainly analyses SEPA (Scottish Environment Protection Agency) data on the use of the in-feed anti-louse medicine emamectin benzoate (EMB), and its relationship to the abundance and species richness of crustaceans in sediment at three different distances from salmon farm net pens. These distances are at the cage edge (CE), at the edge of the Allowable Zone of Effect (AZE), and at Reference sites. W&B also examine data on teflubenzuron, but these data were found to be sparse and so the main focus of their report is on EMB. In this review I focus solely on W&B's analysis of EMB.

My main conclusions after reviewing W&B's report are:

1. W&B should be commended for their heroic efforts in attempting to tidy up and analyse a very noisy database. Field data are always noisy, but the organisation of the SEPA data clearly presented major analytical difficulties, some of which could not be overcome by W&B.

2. 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.

3. 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.

4. 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. 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.

Overall, I think that this is a useful addition to the literature on anti-louse medicines and their possible environmental effects. However, the main conclusions are debatable and are currently overstated.

I provide some more specific details below, by report section (line and page numbers seem to be corrupted in my printout).

Executive Summary

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).

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.

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. Introduction

Section 1.1

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. 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. Materials and Methods

Section 2.1

It is unfortunate that the sampling dates for residue analyses (contained in the FFDB database) and macroinvertebrate surveys (contained in the SMDB database) could differ by over a year, which means that residue analyses and macroinvertebrate data could not be matched. W&B identify this as a problem, but refer the reader to recommendations in Section 5.1, when I think they intend to refer to Section 5.2.

Section 2.3.5

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.

Section 2.3.6

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.

Section 2.5

W&B state that “As a modelling framework GLMMS are still in active development and where nonidentity 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 2003 , Crane et al. 2007) 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.

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.

Results

Sections 3.1.2 and 3.1.3

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?

Section 3.1.4.1

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".

- 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.

- 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. For example, 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 “distanceresponse” curve.

Section 3.2

W&B very sensibly tried to identify sites with pre- and post-EMB use so that they could undertake a before and after comparison. Unfortunately there were rather few such sites available. However, their analysis of these sites provides no evidence to suggest that that commencement of EMB use led to a decline in crustacean abundance or richness.

Discussion

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 and so 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, and that several potentially important covariables (e.g. organic enrichment) could not be included (see W&B’s statement in Section 4.3) .

Section 4.1

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.

Section 4.3

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” because Figure 2 suggests that 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.

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.

Recommendations

Section 5.1

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.

Section 5.2

I strongly support these suggestions for improving the utility of the monitoring data collated by SEPA.

REF02

There are severe problems with the analysis:

1. 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.

2. 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.

3. They used area instead of volume and did not correct for the differences between grab sizes.

4. 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.

General comments:

Is the 7.63 in line 145 correct?

Line 160…non published results should not be relied on and should not be included in such an important report

Line 191. What level of EMB. It needs to be quantified.

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).

Line 197. It would be good to know the dose levels used for the bioassay.

Section 2.1. 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.

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.

Line 290 they talk about taking 2, 3 or 5 reps. My question are these reps or subsamples. Are they taken at randomly selected positions for CE and randomly selected position for AZE and then at randomly selected positions for the REF. 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

Line 327. It seems like a lot of stations were excluded based on distance or locations without knowing the characteristics of the sites. It is a red flag to say “It is Assumed that…..” 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?

Line 358. Is it normal practice to not normalize to area (Volume) sampled?

Line 362. 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.

Sections 2.3.5. 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.

Line384. The use of the maximum EMB 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?

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.

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!!! 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.

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.

Line 462. 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. 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.

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.

Table 8. What is the difference between AZE and int:AZE in the random effects part of the table and then intercept AZE REF (but not CE) are in table for fixed effects. 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.

REF03

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, wellcontrolled 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:

1) 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.

2) If current directionality is a factor, a corollary question arises to determine how distant from the cage edge is the no-effect position. 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.

3) 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. 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.

4) 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 preexisting 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. 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).

5) 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. Likewise, the total number of grab-samples was 1259 from 6 regions and 99 sites (Table 3). The large number of sites and sampling times leads to an elevated concern of an overpowered sample. 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. 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.

6) 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. 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).

7) 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, modeling often follows a path-of-least-resistance or a random-walk, and its conclusions are not as robust as a controlled rigorous preplanned progression.

8) 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. 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.

9) 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.

10) 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.

11) 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.

12) 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. 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. 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) 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. 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.

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. 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.

14) 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.

15) Few sites had total usage of EMB above 1 kg (Figure 5, 15, 16). 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 with the rate of change decreasing as the amount of EMB increases, the noted effect may be less dosedependent 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.

REF04

We conducted an initial review of the Final Report entitled "SARF098: PAMP Refreshment Study" by Wilding and Black (2015). We did not review the underlying data or the references cited (with the exception of a cursory review of two new references). Additionally, we did not conduct a review of the statistical methods employed. Our initial observations can be summarized within the following general categories: toxicological data from the literature, sampling methodology, interpretation of potential effects at reference stations, pre-emamectin benzoate/post- emamectin benzoate analysis, comparison to findings from previous studies, and general assessment of conclusion and recommendations. Where appropriate, we have included recommendations for addressing the issues we have identified.

Reliance on non-relevant toxicological data

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.

Sampling methodology and data quality issues

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.

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.

The problem of predicted impacts at reference stations

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.

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.

The importance of Before/After treatment (or pre-EMB/post EMB analysis)

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.

Conclusions conflict with previous field studies

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.

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, sedimentdwelling 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.

References

Barnaud, A., Le Bris, H. and Aubin, J. (2002). Emamectin benzoate (SLICE) in marine farming

of the Brown Trout (Salmo trutta) in Brittany: its efficacy against the "salmon louse" (Lepeophtherius salmonis) and its environmental impact. SPAH Report 45630.

Bright, D.A., and S. Dionne. 2005. Use of emamectin benzoate in the Canadian finfish aquaculture industry: a review of environmental fate and effects. Prepared for Environment Canada. https://www.watershed-watch.org/publications/files/EnvCan-ReviewofSlice.pdf.

Chapman, P. M., Anderson, B., Carr, S., Engle, V., Green, R., Hameedi, J., ... & Windom, H. 1997. General guidelines for using the sediment quality triad. Marine Pollution Bulletin, 34(6), 368-372.

Fiori, E. (2012). Phytoplankton response to environmental variables and organic pollutants. Laboratory cultures and numerical simulation experiments. University of Bologna.

Mavraganis, T. (2012). An investigation of environmental impacts on sediments by marine cage fish farms using long term metadata analysis. University of Stirling.

Mayor, D.J., Solan, M., Martinez, I., Murray, L., McMillan, H., Paton, G.I., Killham, K. (2008). Acute toxicity of some treatments commonly used by the salmonid aquaculture industry to Corophium volutator and Hediste diversicolor: whole sediment bioassay tests. Aquaculture 285: 102-108.

Telfer, T., Baird, D.J., McHenery, J.G., Stone, J., Sutherland, I., and Wislocki, P. (2006). Environmental effects of the anti-sea lice (Copepoda: Caligidae) therapeutant emamectin benzoate under commercial use conditions in the marine environment. Aquaculture 260:163180.

Tucca, F., Diaz-Jaramillo, M., Cruz, G., Silva, J., Bay-Schmith, E., Chiang, G., and Barra, R. (2014). Toxic effects of antiparasitic pesticides used by the salmon industry in the marine amphipod Monocorophium insidiosum. Archives of Envir. Contam. Toxicol. 67: 139-148.

Veldhoen, N., Ikonomou, M.G., Buday, C., Jordan, J., Rehaume, V., Cabecinha, M., Dubetz, C., Chamberlain, J., Pittroff, S., Vallee, K., van Aggelen, and G., Helbing, C.C. (2012). Biological effects of the anti-parasitic chemotherapeutant emamectin benzoate on a non-target crustacean, the spot prawn (Pandalus platyceros Brandt, 1851), under laboratory conditions. Aquat. Toxicol. 108:94-105.

Waddy, S.L., Merritt, V.A., Hamilton-Gibson, M.N., and Aiken, D.E. (2010). Effects of emamectin benzoate on the molt cycle of ovigerous American lobsters Homarus americanus is influenced by the dosing regimen. Aquat. Biol. 11:47-52.

Wallace, C., Tauber, R., and Sandnes, O. (2004). SLICE® (emamectin benzoate) in Norwegian fish farm sediment. Veterinary Contract Research and Commercial Services. Ltd. (VESO), Oslo. Report V-1211, SPAH Report 46801.

Wilber, M., Young, S., and Wilson, L. (2012). Impact of aquaculture on commercial fisheries:

fishermen's local ecological knowledge. Hum. Ecol. 29-40.

Wilding, T.A. and Black, K.D. (2015). SARF098: PAMP Refreshment Study. Final Report. Scottish Association for Marine Sciences, Scottish Marine Institute. Oban, Argyll, PA37 1QA, United Kingdom. 01 July 2015.

REF05

Comments on SARF098 PAMP Refreshment Study (01 July 2015)

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. Then 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. 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.

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. 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

where the noise (and the random effects) are normally distributed. But if this is the case, then I do not understand how the predicted effect (and confidence intervals) of EMB stays bounded by zero in Figure 10. 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.

REF06

<table>
<tr><td colspan="3"><u>Project Number</u>: SARF098   <u>Completion Date</u>: 1 July 2015</td></tr>
<tr><td colspan="3"><u>Project Title</u>: PAMP Refreshment Study</td></tr>
<tr><td colspan="3">1. In your view have the scientific objectives been achieved. If not, does this need to be addressed by SARF?

The scientific objectives have been met.

Although the conclusions of the project are considered to be sound, in SEPA's view a more detailed interaction with SEPA regarding the extraction of data and the interpretation of the nomenclature of sample site locations in particular may have resulted in a larger data set being available for analysis.</td></tr>
<tr><td colspan="3">2. Comment on the overall results of the project, including their significance for SARF.

The results of the project raise a number of significant concerns in relation to the use of Slice as a medicinal therapeutant for treating sea-louse infestations on marine cage salmon farms and the subsequent discharge of residues to the water environment. This will result in an immediate requirement to review the frequency and scale of use of Slice by the aquaculture Industry and the licensing of the use of Slice and subsequent discharge of residues by SEPA.</td></tr>
<tr><td colspan="3">2. Is there a need for further work? If so, explain.

SEPA has received anecdotal reports of impacts on commercially harvested crustacean in the vicinity of marine cage fish farms. Further work on the effects of sea-lice medicines on these is required.

.</td></tr>
<tr><td>Overall marking</td><td colspan="2">1 - outstanding results
2 - results significantly above expectation
3 - satisfactory results
4 - results below expectation
5 - poor results</td></tr>
<tr><td>Signature</td><td>Print Name</td><td>Date 11 Sep 2015</td></tr>
</table>

REF07

Note that REF07 did not comment initially, but REF05 deputised.