Final

Peer Review Summary Report for the External Peer Review of *Comparative Testing of Corexit EC9500A, Finasol OSR 52, Accell Clean DWD, and ZI 400 at Ohmsett in a Simulated Arctic Environment*

BSEE Contract No. E14PA00008

Task Order No. E15PS00052

August 17, 2015

TABLE OF CONTENTS

1. INTRODUCTION

The EnDyna Team was tasked with selecting five (5) scientific experts to evaluate the draft document "*Comparative Testing of Corexit EC9500A, Finasol OSR 52, Accell Clean DWD, and ZI 400 at Ohmsett in a Simulated Arctic Environment*." In recruiting these peer reviewers and coordinating the peer review, the EnDyna Team evaluated the qualifications of peer review candidates, conducted a thorough conflict of interest (COI) screening process, and independently selected the peer reviewers. The sections below describe the EnDyna Team's selection process for external peer reviewers of the BSEE report.

1.1 Identification of Experts

The experts that were considered for this peer review were identified by literature searches of scientific journals, professional societies, and scientific meetings, as well as searches of our internal peer review database of more than 3,000 scientists. As a result of this search, the EnDyna Team identified a total of 22 potential scientific experts with expertise/experience in oil spill response in Arctic waters and a demonstrated understanding of the methods utilized to understand the efficacy/effectiveness of chemical dispersant use. Of the 22 experts contacted, the EnDyna Team received seven (7) positive responses expressing interest and availability to participate. The remaining 15 candidates were either not available during the peer review timeframe or did not respond to our invitation. Interested candidates provided their name, contact information, and curriculum vitae (CV) and/or biographical sketch containing their education, employment history, area(s) of expertise, research activities, recent service on advisory committees, publications, and awards.

1.2 Conflict of Interest Screening Process

The EnDyna Team initiated COI screening on the seven (7) interested individuals to ensure that the experts had no COI or appearance of the lack of impartiality. The screening was conducted in accordance with the BSEE Peer Review Process Manual (dated August 2014) and involved each expert completing a COI questionnaire to determine if they were involved with any other work and/or organizations that might create a real or perceived COI for the current task. The COI questionnaire distributed to the candidates is provided in Attachment A.

The EnDyna Team received completed COI questionnaires for six of the seven candidates and evaluated each expert's professional and financial information. One of the candidates did not provide their COI information because we proposed to select another candidate from the same organization (i.e., Cedre). Thus, COI information was only completed for one of the Cedre candidates.

1.3 Selection of Candidates

In selecting the peer reviewers, the EnDyna Team evaluated each candidate's credentials to select the experts that, collectively, covered the areas of expertise needed for this peer review, had no real or apparent COI or appearance of the lack of impartiality, and were available to complete the peer review within the desired timeframe. Of the seven candidates considered, one expert was eliminated from further consideration due to potential or actual COIs identified during the screening process. Specifically, this candidate has ongoing work with BSEE related to this area of interest. Another candidate was eliminated because we had two candidates from the same organization (i.e., Cedre) and plan to only propose one to serve as a reviewer. The remaining five (5) candidates were evaluated and after careful consideration of the available information described above, the EnDyna

Team selected the five peer reviewers to participate in the peer review. The names, affiliations, and expertise of the five peer reviewers are provided below.

It should be noted that Dr. Brandvik was initially thought to have a potential COI. Based on the responses provided to the EnDyna Team, no COIs were identified that were relevant to this peer review.

Peer Reviewers Selected by the EnDyna Team:

1.	**NAME:**	Dr. Per Johan Brandvik, Senior Research Scientist/Professor
	AFFILIATION:	SINTEF, Oil Spill Research, Trondheim, Norway
	EXPERTISE:	Dr. Brandvik is a Senior Scientist in SINTEF's Department of Environmental Technology, where he leads research related to the fate of subsea release of oil and injection of dispersants. He is also an Adjunct Professor in Organic Marine Environmental Chemistry (oil spills) at the Norwegian University of Science and Technology, NTNU. Dr. Brandvik has more than 25 years of experience in assessing the fate and behavior of marine oil spills and their influence on operational oil spill contingency. Over the past 15 years, he has focused his research on oil weathering, arctic basin and field testing in Arctic areas. Dr. Brandvik has a Ph.D. from the Norwegian University of Science and Technology, NTNU.

2.	**NAME:**	Dr. Merv Fingas
	AFFILIATION:	Spill Science
	EXPERTISE:	Dr. Fingas has more than 40 years of experience working on oil and chemicals spills. Prior to working as a private consultant at Spill Science, he was Chief of the Emergencies Science Division of Environment Canada for over 30 years, where he conducted studies in oil chemistry, spill dynamics and behavior, spill treating agents, oil spill remote sensing and detection, oil spill tracking and sampling, in-situ burning, and chemical counter-terrorism. He also has experience conducting oil spill research in cold environments and was one of the founders of the Arctic and Marine Oil spill Program (AMOP). Dr. Fingas has a Ph.D. in Environmental Sciences from McGill University.

3.	**NAME:**	Julien Guyomarch
	AFFILIATION:	Cedre (France)
	EXPERTISE:	Mr. Guyomarch is a Chemical Engineer at Cedre in France, where he specializes in developing laboratory and pilot-scale tests to study the fate and behavior of marine oil spills. Specifically, he has developed oil weathering and dispersibility studies to evaluate the effectiveness of treatment agents such as dispersants and emulsion breakers. He has also been involved in identifying the origin of oil pollution incidents through specific chemical analyses. Mr. Guyomarch has a degree in Engineering from the Ecole Nationale Supérieure de Chimie de Rennes, France.

4.	NAME:	Dr. George A. Sorial
	AFFILIATION:	University of Cincinnati
	EXPERTISE:	Dr. Sorial is a Professor in the Department of Biomedical, Chemical, and Environmental Engineering at the University of Cincinnati. He has 30 years of experience in bench-scale and pilot-scale research and chemical analysis with various analytical instruments. This includes using swirling and baffled flask tests to evaluate the effectiveness of dispersants. He also has experience evaluating dispersant effectiveness in wave tank studies. Dr. Sorial has a Ph.D. in Chemical Engineering from the University of Bradford, United Kingdom.

5.	NAME:	Dr. Albert D. Venosa
	AFFILIATION:	Independent Consultant (retired from EPA)
	EXPERTISE:	Dr. Venosa is the former Director of the Land Remediation and Pollution Control Division at the U.S. Environmental Protection Agency (EPA). He worked at EPA for more than 40 years prior to retiring and becoming an independent consultant in 2014. His research interests and expertise include developing protocols to evaluate the effectiveness of chemical oil dispersants, surface-washing agents, and solidifiers. He also specializes in designing and developing pilot- and field-scale tests to determine the effectiveness of bioremediation technologies for oil spill cleanup in seawater, freshwater, beach sediments, wetlands, and soils. Dr. Venosa has a Ph.D. in Environmental Science from the University of Cincinnati.

This peer review report is comprised of Sections 2, 3, 4, 5, and 6. Section 2 provides the charge questions sent to each of the peer reviewers for comments, Section 3 provides the synthesis of their review comments, and Section 4 provides the peer reviewer comments of each reviewer by charge question. In addition, Section 5 (Appendix A) consists of the individual peer review comments and the peer review materials package in Section 6 (Appendix B) is attached separately.

2. CHARGE QUESTIONS

The purpose of this review was to obtain written comments from individual experts on the research report entitled, *Comparative Testing of Corexit EC9500A, Finasol OSR 52, Accell Clean DWD, and ZI 400 at Ohmsett in a Simulated Arctic Environment*. Each reviewer was charged with evaluating the report, providing their overall impressions of the scientific merit of the report, responding to seven charge questions, and providing any other specific comments on the report. The seven charge questions provided to the reviewers are included below.

1	*Are the objectives of the study clearly defined? If not, what are your recommendations for improving the description of objectives?*
2	*Was the methodology used to define the selection and testing conditions of the four dispersants clearly described?*
3	*Were the testing procedures used appropriately describe and properly implemented?*
4	*Were the results of the sampling and statistical testing adequately characterized and clearly described? Were there any critical results or limitations that were not discussed or addressed in the report?*
5	*Are the findings and overall discussion of the results for each product tested clearly discussed?*
6	*Are the conclusions logical and appropriate based on the results? Are there any additional conclusions that could be drawn?*
7	*Does this report present sufficient new data and knowledge, and are the findings useful for informing oil spill response planning in the Arctic regions?*

3. SUMMARY OF PEER REVIEWERS COMMENTS

The section below provides the synthesis of peer reviewers' comments, including general impressions, responses to charge questions, and specific observations.

3.1 General Impressions

The reviewers provided a range of comments on the report, varying from agreement to almost complete disagreement on whether the report was a sound analysis and provided useful information. The reviewers provided their general impressions on the accuracy and clarity of information on product selection, general methodology, statistical methodology, presentation of data or results, validity of conclusions, and applicability to the Artic.

Product Selection

One reviewer expressed disappointment that the researchers did not select two products for testing at Ohmsett that were previously tested and reported in the literature as being excellent dispersants in laboratory and wave tank studies, namely, SPC1000 and JD-2000. This reviewer acknowledged that the researchers tried with SPC1000, but they seemed unable to overcome the viscosity problems they faced with the product. The reviewer suggested that the researchers might have more successful in overcoming those problems with SPC1000 if they had communicated with the manufacturer more fully.[1]

Another reviewer suggested that a preliminary study, preferably at a small scale, would have clearly demonstrated that some of the dispersants were significantly less efficient, and should then not have been included in the large-scale tests. This reviewer stated that a reduced number of dispersants could have allowed for more different test conditions. In particular, the reviewer mentioned testing a more viscous (or weathered) oil as differences between products are more interesting when the dispersibility is reduced. The reviewer added that the control could also be lower.[2]

General Methodology

One reviewer commented that the Ohmsett facility offers unique possibilities regarding large-scale dispersant testing and stated that the experimental procedures were clearly and concisely described in the report. This reviewer believed that the amount of work performed was impressive. The reviewer concurred with the approach where all experiments were performed with three replicates.[3]

Another reviewer supported the approach of blending two crude oils into one homogeneous test oil that could be used for the full study, for 15 planned tests. The reviewer agreed with measuring the viscosity and its API gravity and comparing that to actual region oil. The reviewer supported the approach where each dispersant was tested in three separate replicates to provide detailed statistical analysis. The reviewer emphasized that even the control tests (oil without dispersant) were tested in replicates in order to determine the effect of the test conditions. This reviewer believed that the test runs could represent a real field scenario.[4]

[1] Albert Venosa
[2] Julien Guyomarch
[3] Per Johan Brandvik
[4] George Sorial

Another reviewer noted that this test program was interesting because the reviewer felt there was definitely a need to better understand the effect of various dispersants under arctic test conditions. This reviewer agreed with the approach where experiments were conducted at a large scale, with replicates and controls, and the quality of the dispersion was monitored through many measurements, in particular the droplet size distribution. The reviewer also observed it appeared that significant efforts were made to control the dispersant application, which was then realistic.[2]

Another reviewer expressed several concerns about the methodology. This reviewer stated that the 20 minutes of tests was certainly very rapid. Because the rise time of droplets would be well above this time, this reviewer suggested that the test does not count semi-stable droplets that would rise after the time of the water sweeping. The reviewer recommended that a time of about three hours might be more appropriate. This reviewer agreed that the 70 micron cutoff for droplets may be accepted, but the reviewer suggested that the origin (from Lunel, 1993) was less than scientific. The reviewer stated that this cutoff point happens to be the range of the particular instrument that Lunel was using. The reviewer suggested that this value may have "suddenly become magical." The reviewer recommended that 50 microns might be a better cut-off value. Finally, this reviewer stated that the SPC1000 should have not been diluted with water, which the reviewer believed was certainly part of the reason for the foam production.[5]

Another reviewer emphasized that the methodology was lacking in regard to allowing sufficient time for recoalescence to occur, representing a major weakness. The reviewer believed that the researchers seemed to be unfamiliar with the literature, or they simply did not review it and include their review comments in the report.[1]

Statistical Methodology

One reviewer commented that the need for and relevance of some of the advanced statistical methods applied were not clear. The reviewer recommended that simpler and more well-known statistical methods could be applied. This reviewer also commented that very few of the statistics presented in multiple tables in the report were used in the discussion and conclusions. This reviewer observed that the method used to quantify dispersed oil droplets has an upper detection limit of 500 microns, but there was a substantial fraction of larger droplets present. The reviewer emphasized that cumulative distribution curves should not be used to calculate the median droplet diameter, since they only describe a fraction of the droplets.[3]

Another reviewer stated that the statistical methodology was less than appropriate because no analyses of variance were used to establish significant differences. The reviewer believed, that without such analyses of variance, the quantitative, statistical findings, and overall discussion of results were inadequately presented. This reviewer also commented that pair-wise statistical differences were inadequately discussed or explained.[1] Another reviewer commented that the detailed statistical analyses were one of the strengths of the report.[4]

Another reviewer recommended more justification for the removal of the data point for the Accell Clean DWD test. The reviewer suggested using a statistical test to justify the need for removing that data point.[5]

[5] Merv Fingas

Presentation of Data or Results

One reviewer commented that all the data necessary to evaluate the data analysis and conclusions in the report were available in tables in the report and in the appendices.[3]

Another reviewer identified several weakness of the report. This reviewer commented that the plots provided for the data selections were labelled as Run number without specifying the dispersant type or oil control. The reviewer recommended that concentration units should be provided in those plots, not within the text. In addition, this reviewer commented that the droplet size distributions were measured at two depths, and questioned how the plot in Figure 5 was generated from these two depths. Overall, this reviewer recommended that all the tables provide data to two decimal places. The reviewer also suggested that a nomenclature for all symbols used and acronyms should be provided after the table of contents.[4]

Another reviewer questioned why the full Appendix B was needed, as now presented. The reviewer commented that size data was not particularly useful in graphical form. In addition, this reviewer stated it is standard practice to report values as the test value less the control value. The reviewer provided specific suggestions for Table 4 (see charge question #4).[5]

Another reviewer commented that the syntheses of results were well presented in synthetic tables, thus clearly showing the effect of the various factors. However, this reviewer suggested that more details on the experimental results would be useful, considering the importance of the work. In particular, this reviewer suggested providing more details on the way the concentrations were determined to provide more clarity, as well as the data concerning the processing of the two sets of droplet sizes. The reviewer felt that Appendix B would also benefit from details about the experiment rather than the test number. The reviewer commented that the statistical analysis was also unclear, and suggested that a graphical presentation of the results would be preferable, rather than a series of figures with (too) many significant results.[2]

Validity of Conclusions

One reviewer stated that the main conclusion of the report conflicted with the statistical analysis. The reviewer specifically emphasized there was no evidence in the data supporting a significant difference in performance between Corexit EC9500A and Finasol OSR 52. The reviewer believed that the main conclusions stating that Corexit EC9500A was best, followed by Finasol OSR 52, was not justified by the data. Instead the reviewer suggested that both the Corexit EC9500A and Finasol OSR 52 products were significantly better than the control and the other products tested. The reviewer noted that the other products were not significantly better than the control.[3]

Another reviewer commented that the conclusions were not logical or appropriate due to: 1) the flawed statistical analyses, and 2) the bias this reviewer observed that the researchers had toward Corexit EC9500A and Finasol OSR 52. The reviewer mentioned that Finasol OSR 52 is a product that is very similar, if not identical, to Corexit EC9500A. Overall, this reviewer felt the report was mediocre, and that the accuracy of information was subpar.[1]

Another reviewer stated it was not clear why Finasol OSR 52 provided similar dispersant effectiveness as Corexit EC9500A while the droplet size distribution curves were significantly different. This reviewer recommended providing more discussion to support this conclusion.[4]

Finally, another reviewer commented that the report addressed the terms of reference and was generally accurate.[5]

Applicability to the Artic

One reviewer recommended highlighting that dispersants had to be heated for this study, and expressed concerns that heating dispersants in order to discharge them ruled out Artic dispersion. The reviewer acknowledged that the established protocol at Ohmsett was used, but suggested that the report should comment on its applicability to this study. The reviewer mentioned that this study was to establish dispersibility of oils at low temperatures; however, the temperature and pressure of oil release were not noted in the report. The reviewer emphasized this data could be very important. This reviewer stated there is no practical way that dispersants could be heated in the Arctic and certainly they would cool to frozen droplets (snow) by the time they hit the water.[5]

Another reviewer commented that the only "data" from the report that might be useful in future decision making for Arctic oil spills were the qualitative observations on foaming of two of the four products.[1]

3.2 Responses to Charge Questions

The section below provides the synthesis of the five peer reviewers' comments, concerns, and suggestions to the charge questions.

1	*Are the objectives of the study clearly defined? If not, what are your recommendations for improving the description of objectives?*
Comments:	The reviewers agreed that the objectives of the study were clearly defined.[2,3,4,5] However, one reviewer commented that the study objectives imposed a preconceived bias toward one dispersant just because it was already the most frequently tested dispersant on the NCP Product Schedule. This reviewer identified the study objectives as: "This study compared the performance of Corexit and three other commercially available dispersant formulations, as measured by Dispersant Efficiency and the size distribution of dispersed oil droplets in the water column." Because this reviewer saw a positive bias towards Corexit, the reviewer recommended revising the study objectives to something like: "This study compared the performance of four commercially available dispersant formulations…" This reviewer emphasized that all biases should be eliminated before BSEE publishes this report.[1] Another reviewer believed that the report followed the study objectives.[5] Another reviewer believed that the objective of the study, which this reviewer identified as "to better understand and compare the effectiveness of various dispersants under simulated arctic test conditions," was very relevant for both the industry and authorities which need to improve Arctic oil spill contingency.[3]

2	*Was the methodology used to define the selection and testing conditions of the four dispersants clearly described?*
Comments:	The reviewers provided varying responses to this charge question. Two reviewers agreed overall that the report clearly described the methodology. Of those two reviewers, one not only agreed but added that he believed the selection of dispersants followed the established criteria.[5] Another reviewer felt that the methodology used to define the selection and testing conditions was clearly defined in methods, test oil, and testing procedures.[4] Discussed below are the comments of three reviewers who disagreed with the methodology used to either select the dispersants or testing conditions. <u>Selection of Dispersants</u> One reviewer commented that the methodology for selection of dispersants was simple and not fully explained. This reviewer expressed concerns that the researchers may be unfamiliar with the literature, because they selected the two most popular dispersants on EPA's Product Schedule and two other lesser known products without regard to whether any of the selected products were ever reported in peer-reviewed literature citations either in lab, mesocosm, or field studies. This reviewer suggested the researchers also should have studied the MSDS sheets of EPA's Product Schedule products that have detailed descriptions of differences in constituents, properties, and performance for those products. This reviewer commented that reviewing the literature and MSDS sheets might have led to a priori selection of more effective products instead of including a non-effective product (e.g., ZI 400) in the study.[1] Another reviewer suggested that selecting all products based on stockpiled amounts would be a better approach. This reviewer noted the discrepancy in selection of Corexit EC9500A/Finasol OSR 52 due to their large volume in U.S. stockpiles, while the only criteria listed for selecting the other two products was because they were "not as well known."[3] <u>Selection of Testing Conditions</u> Only one reviewer specifically emphasized that the methodology for selection of testing conditions remained largely unclear. This reviewer commented that a very limited number of parameters (API gravity and sulfur content) were used as the basis for relating test oil characteristics to other arctic crudes. The reviewer emphasized other factors were just as relevant in terms of behavior at sea, such as asphaltenes and wax contents, distillation curve, and viscosity. This reviewer also commented that the pour point of the oils seemed high in comparison with the test temperatures, and also it was too high when considering dispersibility studies. This reviewer suggested that a preliminary study could have better identified the most important parameters and helped to achieve lower dispersant efficiencies. Finally, this reviewer commented that the control experiment should have led to poor dispersibility.[2]

3	*Were the testing procedures used appropriately described and properly implemented?*
Comments:	The reviewers in general agreed that the testing procedures were appropriately described, but some reviewers had suggestions or concerns regarding the testing procedures. Two reviewers agreed overall with the description and implementation of the testing procedures. One reviewer commented that testing procedures were described accurately within the report and Appendix A and were properly implemented.[4] After noting that the Ohmsett facility offers unique possibilities regarding near field conditions for dispersant testing, another reviewer commented that the experimental procedures were clearly and concisely described in the report. This reviewer emphasized that all the data needed to evaluate the data analysis and conclusions were available in tables in the report and in the appendices.[3] Another reviewer commented that the testing procedures were described adequately, but not appropriately because recoalescence is an extremely important outcome of using any oil dispersant. The reviewer explained that droplet sizes <70 microns should generate droplets that were neutrally buoyant and will not likely recoalesce. However, a good scientific method should attempt to study and confirm this outcome rather than making assumptions. Halting the wave experiments after 20 minutes and then collecting whatever oil was not dispersed does not allow sufficient time for re-coalescence. Using the 20 minute interval would add to the amount of oil that was not dispersed to rise and be collected with the rest of the non-dispersed oil.[1] The reviewer acknowledged that an additional three hours passed before preparing for the next experiment, but commented there was no mention of whether any more recoalesced oil was collected at that time. This reviewer recommended that a full 24 hours should pass before confirming no more oil was surfacing from the water column and emphasized that not allowing longer time periods between wave experiments was a major weakness in the study.[1] Another reviewer also commented that the 20-minute time period after the test when sweeping was started was too short and would result in missing much of the resurfaced oil.[5] Another reviewer stated that the report followed the established procedures; however, it should have included information on the need to heat the dispersant. This reviewer believed it was unclear if the oil was heated and at what pressure it was discharged. This reviewer expressed concerns that the viscosity of the oil appears to be only calculated.[5] Another reviewer commented that the testing procedure was well explained, but some parameters of the experiments were not consistent with the methodology. In particular, this reviewer expressed concerns about using the same Dispersant to Oil Ratio for Finasol OSR 52 for any conditions of dosage (lowest, mean, and

	highest). Adding information about the location of the oil slick and laser scattering (LISST) equipment would be helpful. This reviewer recommended better description of the protocol for the oil concentrations measurements (e.g., type of measurement, calibration, uncertainty) as well as the method used to assess the remaining oil at the end of each experiment.[2]

4	*Were the results of the sampling and statistical testing adequately characterized and clearly described? Were there any critical results or limitations that were not discussed or addressed in the report?*
Comments:	The reviewers provided varying responses to this charge question. Summarized below are reviewer comments about dispersant effectiveness results, statistical methods, and data presentation. Also, summarized below are comments from reviewers about critical results or limitations that were not addressed in the report. <u>Dispersant Effectiveness Results</u> One reviewer commented that the definition of Dispersant Effectiveness (DE) and volume fraction below 70 microns were well defined and relevant. This reviewer also commented that the documentation of released oil, applied dispersant, and DOR were all relevant and well described. The reviewer stated that the documentation of droplet sizes measured was also well documented. The reviewer acknowledged that data from replicate testing of each dispersant were available both for DE and volume of droplets <70 microns. This reviewer identified the use of droplet size data as a weak point in the report. The method used to quantify oil droplet sizes was based on laser scattering (LISST 100X). This reviewer noted that this instrument has an upper detection limit of 500 microns. The reviewer emphasized that the shape of (most) cumulative distribution curves in the report shows that there were a significant amount of particles larger than 500 microns. The reviewer suggested that these cumulative curves should not be used to calculate the median droplet diameter, since they only describe a fraction of the present droplets (see Table 5). The reviewer commented this issue needs to be discussed in the report.[3] Another reviewer commented that the way the two sets of data resulting from the LISST measurements were combined was not explained in the report. This reviewer noted that the threshold value of 70 microns was not discussed in the report. The reviewer suggested it would be interesting to reassess this limit based on a large dataset obtained at a large scale.[2] Another reviewer stated it was not clear why Finasol OSR 52 provided similar DE as Corexit EC9500A while the droplet size distribution curves were significantly different. This reviewer recommended providing additional description to explain this result.[4] <u>Statistical Methods</u>

One reviewer stated that the need for some of the statistical methods applied was not clear. The reviewer summarized his understanding of how the statistical methods were used: 1) Find the significant experimental variables that influence DE—Stepwise Regression, 2) Find if there was a surfactant memory in the system affecting DE—Breusch-Godfrey test, and 3) Find if there was a significant difference between products and control (DE and vol. <70 microns)—Tukey Honestly Significant Difference (Tukey HSD) test. The reviewer suggested applying simpler and better known statistical methods. The reviewer felt that the first and second methods were appropriate, but the third seemed over complicated. The reviewer recommended that using a student T-test comparing each dispersant with the control would be simpler and sufficient for this study. The reviewer felt that a student T-test would be much easier to communicate to the reader, which was maybe the most important reason for using it instead of the Tukey HSD test.[3]

Another reviewer commented that the sampling and statistical testing methods were not fully appropriate, although the reviewer felt they were adequately described. The reviewer stated that an analysis of variance (ANOVA) should have been performed at the conclusion of all tests. The reviewer emphasized that researchers usually use ANOVA to determine if there were any significant differences in experimental results. If there were none, then there would be no need to conduct any further statistical tests. The reviewer felt that the HSD test should be used only when statistical differences were determined, and the HSD test would determine where the significant differences existed. The reviewer summarized his understanding that the authors used stepwise multiple regression analysis to establish a best-fit model to keep only those variables that explain the variability in the data. The reviewer stated that the authors then used the HSD test to identify pair-wise differences in regard to DE, % oil droplet sizes <70 microns, and oil concentration having droplet sizes <70 microns.[1]

The reviewer added that using statistical methods to study the influence of salinity was inappropriate. The reviewer summarized that the stepwise multiple regressions asked how significant salinity, interfacial tension, and viscosity were in explaining the variability. However, the reviewer felt that was not appropriate because the same seawater and tank were used in all experiments without emptying the tank and refilling it with fresh seawater. The reviewer questioned why Table 6 showed that salinity played a role for the dependent variable "% below 70 µm" and for viscosity. This reviewer believed that none of the three variables in the stepwise multiple regression significantly affected DE or concentration below 70 microns.[1] Another reviewer commented that the statistical analysis was not clearly described in the report. This reviewer suggested that figures (with error bars) were preferred over tables with too many significant numbers.[2]

<u>Data Presentation</u>

Several reviewers suggested areas where the presentation of data or results could be improved in the report. One reviewer stated it is standard scientific practice to report values as the test value less the control value, which this reviewer felt was not done correctly in Table 4. This reviewer also suggested that the arithmetic to calculate the actual effectiveness appears to be incorrect. According to this reviewer, the values should be presented as listed below.[5]

Product	Reported Value	Test Value
Corexit	22.9	72.7
Finasol	22.4	72.2
Accell	1.5	51.3
ZI 400	-4.1	45.7
Control	0	49.8

Another reviewer commented that the plots provided for the data selections were labelled as Run number without specifying the dispersant type or oil control. This reviewer recommended providing concentration units in the plot not within the text. This reviewer also stated it was not clear how the plot in Figure 5 was generated from the two depths for measuring the droplet size distributions.[4]

Another reviewer commented that the results of oil concentrations were not clearly explained in the report. This reviewer concurred that providing no units on those figures was a problem.[2]

Another reviewer questioned the blank cells in Table 6. This reviewer believed that if all three variables were tested, then there should be probability values (p-values) in all table cells.[1]

Critical Results or Limitations Not Addressed

One reviewer noted that Tables 7-9 contained the critical statistical data, but the reviewer emphasized these results were not discussed and explained in sufficient detail. This reviewer outlined some aspects of the results that need further discussion. In Tables 8 and 9, only two pairwise significances were observed. For DE in Table 8, only the pairs ZI 400-Corexit EC9500A and ZI 400-Finasol OSR 52 were statistically different from each other. For the <70 micron concentrations in Table 9, only the pairs Finasol OSR 52-Control and ZI 400-Finasol OSR 52 were statistically different from each other. Finally, Table 7 shows seven different pairs showing significant differences from each other.[1]

Another reviewer commented that the control was very high, and identified this as an important result that was not discussed in the report. The reviewer noted that the effects of dispersants were expressed as a gain in percentage compared to this control condition. The reviewer emphasized that the control shows a significant variability (from 43.0 to 59.7), and suggested that the comparison between

	dispersants should include this key point. This reviewer also recommended adding a comment or an explanation on the increase of the oil concentration over time for the lower LISST (test 8).[2]

5	*Are the findings and overall discussion of the results for each product tested clearly discussed?*
Comments:	Three reviewers agreed that the results were clearly described and two reviewers disagreed. Among those reviewers that agreed, two reviewers agreed that the results were clearly described with the exception of those issues the reviewers had raised under charge question #4.[2,4] One of those two reviewers re-emphasized that the high level of the efficiency for the control, combined with a great variability, should prompt reservations about the conclusions.[2] The third reviewer also recommended more discussion on the removal of a data point from one of the dispersants. This reviewer suggested the removal of that data point should be done on a statistical basis, not on an ad hoc basis.[5] Summarized below are the comments from reviewers that disagreed relating to inadequate discussion of statistical analysis, and selection of dispersant products for the project. <u>Inadequate Discussion of Statistical Analysis</u> The concerns of one reviewer who disagreed focused on lack of discussion about statistical analyses in the discussion and conclusion sections.[3] The reviewer commented that the report presents results from the statistical methods in Table 6, 7, 8, and 9, but was disappointed that not a single reference to those tables was found in the subsequent discussion and conclusion sections. The reviewer suggested that the discussion section should have included statistical material instead of focusing on justification of why a fifth dispersant was omitted and additional operational observations like spray patterns, challenges with high viscosity, and possible foaming. The reviewer provided an example from Table 8, which shows that DE for neither Corexit EC9500A nor Finasol OSR 52 was significantly higher than the control. The reviewer suggested this was probably due to the large deviation (±6) for the two products compared to the difference between the products and the control (22-23). The reviewer commented this seemed strict, suggested that an ordinary T-test would probably have shown that they were significant, and recommended that results like this should be included in the conclusions.[3] Another reviewer emphasized that the quantitative, statistical findings, and overall discussion of results were inadequately discussed and/or explained. The reviewer stated that the findings and overall discussion of results would not lead to a more informed decision on which dispersant product should be used in an oil spill. The reviewer expressed concerns that many unanswered questions remained for the reader.[1]

	Selection of Dispersant Products This reviewer stated that the dispersant product selection was inadequate since two dispersants that have been well tested in the literature were not tested in the project, namely, SPC1000 and JD-2000. This reviewer acknowledged that the authors had problems with SPC1000, but was not convinced that the authors tried hard enough to overcome the viscosity and application problems of SPC1000. The reviewer suggested that JD-2000 should have been chosen based on the literature. The reviewer commented that not even mentioning JD-2000 in their selection procedure may demonstrate the author's unawareness of the literature.[1]

6	*Are the conclusions logical and appropriate based on the results? Are there any additional conclusions that could be drawn?*
Comments:	The reviewers provided varying responses to this question. One reviewer agreed that conclusions were clearly provided for the four dispersants studied.[4] Another reviewer commented that the study did not explain the arithmetic for the reported values. The reviewer suggested more emphasis on discussing the practicality of the results for Arctic field applications. This reviewer noted that the oil and dispersants cannot be heated in the Artic, and questioned application of the results to the field. The reviewer recommended this should be discussed in the conclusions.[5] Three reviewers disagreed. One reviewer referred back to comments under charge question #4.[2] The other two reviewers that disagreed relating to flawed statistical analyses and product bias comments are summarized below. Flawed Statistical Analyses One reviewer stated that the conclusions were not logical and appropriate, and re-emphasized that the statistical analyses were flawed.[1] Another reviewer stated that the conclusions were not logical. The reviewer noted that although the report has a comprehensive statistical chapter, very little of this information was used in the conclusions.[3] The reviewer conducted a detailed evaluation of the report's data and found that two of the dispersants have a significantly better effectiveness than the control (Corexit EC9500A and Finasol OSR 52) based on both DE and volume of droplets <70 microns. This reviewer quoted conclusions from the report's abstract that: ▪ "Corexit EC9500A performed very well …. producing the highest average Dispersant Effectiveness (DE)" ▪ "Finasol OSR 52 demonstrated a performance close to that of Corexit" The reviewer stated the terms "highest average" and "close to" were not very comparable to the statistical ambitions earlier in the report. Based on the data

	presented in Table 3, the reviewer suggested it was not possible to see a significant difference between the DE of Corexit EC9500A (73% ±6) and Finasol OSR 52 (72 ±6). The reviewer also noted that Corexit EC9500A has also a higher DOR (5%) than Finasol OSR 52 (3%), which was not discussed in the report. The reviewer questioned this ranking as highly surprising based on the presented data and especially based on the use of advanced statistical methods in other sections in the report.[3] Product Bias One reviewer stated that the authors clearly displayed positive biases toward Corexit EC9500A and Finasol OSR 52. The reviewer acknowledged that qualitative observations about the lack of foaming by Corexit and Finasol versus the presence of foaming by ZI 400 and Accell Clean DWD, which may have negatively impacted efficient and effective applications of the latter products to the slick, were indeed helpful in allowing good decisions to be made against use of the latter two products in the Arctic and in favor of the former two. However, the reviewer emphasized that the lack of statistical significances between Finasol, Corexit, and Accell do not help in such decision making.[1]

7	*Does this report present sufficient new data and knowledge, and are the findings useful for informing oil spill response planning in the Arctic regions?*
Comments:	One reviewer mentioned that performing the testing on fresh oil was a major weakness in the report. This reviewer also suggested that the report include recommendations about testing necessary to get a dispersant on the EPA list.[3] Others commented on the usefulness of the findings for Artic regions. Summarized below are comments from reviewers on testing on fresh oil, recommendations for the EPA list, and applicability to the Artic. Testing on Fresh Oil One reviewer emphasized that in actual oil spill response operations the dispersants are not applied to fresh oils. In most cases, it will take a couple of hours to make a decision (e.g., obtain a permit), get the dispersant to the site, detect the slick, and initiate dispersant spraying. This means that dispersants are usually sprayed on slightly weathered oil, were emulsification is the most significant process, raising the viscosity due to increased water content. This gives the dispersant other challenges than dispersing water free of fresh oil. These more realistic challenges include penetrating viscous emulsions, increased oil pour point due to evaporation, breaking emulsions before dispersing the oil, etc. The reviewer stated that the reduced relevance of performing the testing on fresh oil was not addressed and this was a major weakness of the report.[3] Recommendations for the EPA List One reviewer extracted some operationally important messages from the data about dispersants that the reviewer recommended should be presented more

clearly in the report. Those observations are: 1) There were larger differences in effectiveness between the dispersants on the EPA list, 2) some of them have excellent performance and should be preferred (stockpiled) by industry and authorities, 3) some of the dispersants on the list have a very low effectiveness (since the threshold to get into the list is very low) and should NOT be used or stockpiled as marine oil spill dispersants, and 4) more realistic and challenging testing is needed to get into the EPA list.[3]

Applicability to the Artic

One reviewer noted that the applicability to the Artic was in question. The reviewer commented that the correct results (value - control), even with their weaknesses, showed that dispersant application to the Arctic would be questionable, although this does not come out in the report.[5]

Another reviewer stated that this study well addressed the possibility of applying dispersants, from a technical point of view. However, this reviewer commented that even if tests were performed at low temperatures, these temperatures were not totally representative of arctic conditions. In particular, only low pour point oils can be treated, while that was not the case for this study. This reviewer suggested it would have been interesting to compare dispersants at the laboratory scale at very low temperatures, to clearly identify the influence of this parameter, and then to conduct large-scale experiments by varying the oil condition and not only the dispersant products.[2]

Another reviewer believed the only “data” that might be useful in future decision making for Arctic oil spills were the qualitative observations on foaming of two of the four products.[1] Another reviewer commented that the report provided new data that could be representative to an actual spill scenario.[4]

4. PEER REVIEWER COMMENTS BY CHARGE QUESTIONS

4.1 General Impressions

<table>
<tr><th colspan="2">General Impressions</th></tr>
<tr><td colspan="2">Provide overall impressions (approximately 1/2 page in length) addressing the accuracy of information presented, clarity of presentation, and soundness of conclusions.</td></tr>
<tr><td>Per Johan Brandvik</td><td>The Ohmsett facility offers unique possibilities regarding large-scale dispersant testing and the description of the experimental procedures are clear and concisely described. The amount of work performed is impressive and all experiments are performed with three replicative. All the data needed to evaluate the data analysis and conclusions in the report are available in tables in the report and in the appendices.

However, there are some shortcomings in the report, the most significant are:

▪ The need for and relevance of some of the advanced statistical methods applied are not clear, simpler and more well-known statistical methods could have been applied.

▪ Very little of the statistics presented in multiple tables are used in the discussion and conclusions.

▪ The main conclusion is conflicting with the statistical analysis.

▪ There is no evidence in the data material supporting a (significant) difference in performance between C9500 and OSR-52

▪ The method used to quantify dispersed oil droplets has an upper detection limit of 500 microns, but there is a substantial fraction of larger droplets present. The cumulative distribution curves should not be used to calculate the median droplet diameter, since they only describe a fraction of the droplets.

The main conclusions stating that C9500 is best, followed by OSR-5 is not justified by the data. They are both significantly better than the control and the other product tested. The other products are not significantly better than the control.</td></tr>
<tr><td>Merv Fingas</td><td>The report addresses the terms of reference and is generally accurate, but could be improved in several ways:

Calculation of values – It is standard practice to report values as the test value less the control value, thus the values here should be:

<table>
<tr><th>Product</th><th>Reported Value</th><th>Test Value</th></tr>
<tr><td>Corexit</td><td>22.9</td><td>72.7</td></tr>
<tr><td>Finasol</td><td>22.4</td><td>72.2</td></tr>
<tr><td>Accell</td><td>1.5</td><td>51.3</td></tr>
</table>
</td></tr>
</table>

ZI 400	-4.1	45.7
Control	0	49.8

(Table 4 tries to show this, but the math is unclear or incorrect.)

Secondly, although the established protocol at Ohmsett is used, comments should be made on its applicability to this study. The study was to establish dispersibility of oils at low temperatures; however, the temperature and pressure of oil release were not noted. This could be very important. Importantly, the dispersants had to be heated to be discharged. Does this not rule out Arctic dispersion? There is no practical way that dispersants could be heated in the Arctic and certainly they would cool to frozen droplets (snow) by the time they hit the water. This should at least be highlighted. Another point about the protocol is that the 20 minutes of tests is certainly very rapid. The rise time of droplets would be well above this time, thus the test does not count semi-stable droplets that would rise after the time of the water sweeping. A time of about 3 hours might be more appropriate

The **70 µm cutoff** for droplets may be accepted, but the origin (from Lunel, 1993) is less than scientific. This cutoff point happens to be the range of the particular instrument, Lunel was using. This is a comment on a value that has suddenly become magical. A better cut-off might be 50 µm.

The **SPC 1000** should have not been diluted with water. This is certainly part of the reason for the foam production.

More **justification** is needed for the removal of the data point for Accell test. Perhaps a statistical test should be used.

Appendix B is hardly needed in full as now presented. The size data is not particularly useful in graphical form.

Julien Guyomarch

This test program, conducted in order to better understand the effect of various dispersants under arctic test conditions, is interesting as there is definitely a need in this field. In addition, experiments were conducted at a large scale, with replicates and controls, and the quality of the dispersion was monitored through many measurements, in particular the droplets size distribution. It also appears that significant efforts were made to control the dispersant application, which was then realistic. Syntheses of results were well presented in synthetic tables, thus clearly showing the effect of the various factors. However, considering the importance of the work, more details on the experimental results would be useful. In particular, the way the concentrations were determined is not clear, as well as the date concerning the processing of the two sets of droplet sizes. Appendix B would also benefit from details about the experiment rather than the test number. The statistical analysis is also unclear, and a graphical presentation of the results would be preferable (rather than a series of figures with (too) many significant results).

Finally, regarding the plan of experiment, it seems that a preliminary study, preferably at a small scale, would have clearly demonstrated that some of the

	dispersants were significantly less efficient, and should then not have been included in the large scale tests. A reduced number of dispersants could then have allowed different conditions of tests, in particular a more viscous (or weathered) oil as differences between products are more interesting when the dispersibility is reduced (and the control could also be lower...).
George A. Sorial	The research work provided a comparative study for the effectiveness of four dispersants, namely, Corexit EC9500A, Finasol OSR 52, Accell Clean DWD, and Z1 400 on Alaskan Crude Oil at ambient temperatures varying between -4.9C to 8.7 C and water temperatures varying between -1.2 and 1.3 C. The effectiveness of the dispersants was captured by applying the dispersants on a surface slick using Ohmsett's spray bar simulating boat spraying system. Droplet size distribution of the dispersed oil and the volume of oil remained on the surface after the test compared to the original volume applied was measured. The droplet size distribution was measured at two depth, one meter and two meters. Droplet size distributions smaller than 70 microns are considered to be fully dispersed. Strength of the Report: 1. The authors blended two crude oils into one homogeneous test oil that could be used for the full study – 15 planned tests. They measured the viscosity and its API gravity and compared to actual region oil. 2. Each dispersant was tested in three separate replicates to provide detailed statistical analysis. Even the control tests (Oil without dispersant) were tested in replicates in order to determine the effect of the test conditions. 3. Detailed statistical analysis was provided. 4. The test runs could represent a real field scenario. Weakness of the report: 1. Data selection: The plots provided for the data selections are labelled as Run number without specifying the dispersants type or oil control. The concentration units should be provided in the plot not within the text. 2. It is not clear why Finasol provided similar dispersant effectiveness as Corexit while the droplet size distribution curves are significantly different. More elaborations need to be provided. 3. The droplet size distributions were measured at two depths, how was the plot in Figure 5 generated from these two depths. 4. All the tables should be provided to two decimal places. 5. A nomenclature for all symbols used and acronyms should be provided after the table of contents.

Albert D. Venosa	My overall impression is mediocre. The accuracy of information, as evident in my review comments below, is subpar. The researchers seemed to be unfamiliar with the literature, or they simply did not review it and include their review comments in the report. Their methodology was lacking in regard to the allowance of sufficient time for recoalescence to occur, a major weakness. Statistical methodology was less than appropriate since no analyses of variance were used to establish significant differences. Thus, the quantitative, statistical findings and overall discussion of results were inadequately presented. Pair-wise statistical differences were inadequately discussed or explained. As for product selection, it was disappointing that the researchers did not select two products for testing at Ohmsett that had been previously tested and reported in the literature as being excellent dispersants in laboratory and wave tank studies, JD2000 and SPC1000 (they tried with SPC1000, but they seemed unable to overcome the viscosity problems they faced with the product, which they might have more successful with if they had communicated with the manufacturer more fully). Their conclusions were not logical or appropriate due to the flawed statistical analyses and the bias they had toward Corexit 9500 and Finasol (Finasol is a product that is very similar if not identical to Corexit 9500). In my opinion, the only "data" that might be useful in future decision-making for Arctic oil spills are the qualitative observations on foaming of two of the four products.

4.2 Responses to Charge Questions

1	*Are the objectives of the study clearly defined? If not, what are your recommendations for improving the description of objectives?*
Per Johan Brandvik	The objective of the study is clearly defined (*to better understand and compare the effectiveness of various dispersants under simulated arctic test conditions*) and very relevant for both the industry and authorities which need to improve Arctic oil spill contingency.
Merv Fingas	Yes, and I believe that the report follows the study objectives.
Julien Guyomarch	The objectives of the study are clearly defined.
George Sorial	The objectives of the study were clearly defined in the last paragraph of the introduction section.
Albert D. Venosa	The objectives were presented in one sentence on p. 3, i.e., "This study compared the performance of Corexit and three other commercially available dispersant formulations, as measured by Dispersant Efficiency and the size distribution of dispersed oil droplets in the water column." Alas, the sentence displays early in the report the positive bias towards Corexit. Rather, it should state something like "This study compared the performance of four commercially available dispersant formulations…" So, the objectives were indeed clear but imposed a preconceived bias toward one dispersant just because it has been the most tested one on the

	NCP Product Schedule. All biases should be eliminated before BSEE publishes this report.

2	*Was the methodology used to define the selection and testing conditions of the four dispersants clearly described?*
Per Johan Brandvik	No, C9500/OSR-52 were selected due to their large volume in U.S. stockpiles, but the only criteria listed for selecting the other two products was because they were "*not as well known*". Selecting all products based on stockpiled amounts would be a better approach.
Merv Fingas	Yes, and the selection did follow the criteria set down.
Julien Guyomarch	The methodology used to select the dispersants is clear, but the choice of the testing conditions remains largely unclear. The characteristics of the test oil were related to other arctic crudes but based on a very limited number of parameters (API gravity and sulfur content), whereas, in terms of behavior at sea, asphaltenes and wax contents, distillation curve, viscosity… are as relevant. In addition, the pour point of the oils seems high in comparison with the test temperatures (and too high when considering dispersibility studies). A preliminary study should have been conducted in order to better identify the most important parameters and to get lower dispersant efficiencies. And the control experiment should have led to poor dispersibility.
George A. Sorial	The methodology used to define the selection and testing conditions were clearly defined in methods, test oil, and testing procedures.
Albert D. Venosa	The methodology of selection was simple and not fully explained. The investigators confined their selections to the two most popular dispersants on EPA's Product Schedule and two other lesser known products without regard to whether any of the selected products have ever been reported in peer-reviewed literature citations either in lab, mesocosm, or field studies. This implies the researchers are unfamiliar with the literature. It also implies that the researchers did not study the MSDS sheets of the Product Schedule products that describe in greater detail how the products differ in constituents, properties, and performance from one another. Had they done any of this, they might have confined their selections to more effective products a priori rather than a non-effective one like ZI 400 (despite the fact that the latter product is on the Product Schedule).

3	*Were the testing procedures used appropriately describe and properly implemented?*
Per Johan Brandvik	The Ohmsett facility offers unique possibilities regarding near field conditions for dispersant testing and the description of the experimental procedures are clear and concisely described. All the data needed to evaluate the data analysis and conclusions in the report are available in tables in the report and in the appendices.

Merv Fingas	The report followed the established procedures; however, comments should have been made on the need to heat the dispersant. It is unclear if the oil was heated and at what pressure it was discharged, further the viscosity of the oil appears to be only calculated. The time after the test (20 min.) that sweeping was started is too short and would result in missing much of the resurface oil.
Julien Guyomarch	The testing procedure was well explained but some parameters of the experiments were not consistent with the methodology. In particular, for the Finasol OSR 52, the Dispersant to Oil Ratio was the same for any conditions of dosage (lowest, mean, and highest). A scheme of the test protocol would also have been added, with the location of the oil slick and LISST equipment. The protocol for the oil concentrations measurements (type of measurement, calibration, uncertainty…) should have been better described, as well as the method used to assess the remaining oil at the end of each experiment.
George A. Sorial	The testing procedures described within the report and Appendix A were described accurately and properly implemented.
Albert D. Venosa	The testing procedures were adequately described, but unfortunately not in an appropriate fashion. What I mean by this is that recoalescence is an extremely important outcome of using any oil dispersant. Droplet sizes < 70 µm should generate droplets that are neutrally buoyant and will not likely recoalesce. However, a good scientific method should attempt to confirm this outcome by studying it in greater detail than described rather than assuming it should. Halting the wave experiments after 20 min and then collecting whatever oil was not dispersed does not allow sufficient time for recoalescence to take place, which would add to the amount of oil that was not dispersed to rise and be collected with the rest of the non-dispersed oil. Granted, three more hours were allowed to pass by before preparing for the next run, but no description was presented that any more recoalesced oil was collected at that time. I would recommend that a full 24 hours should be allowed to pass before confirming no more oil surfacing from the water column. This was another major weakness in the report.

4	*Were the results of the sampling and statistical testing adequately characterized and clearly described? Were there any critical results or limitations that were not discussed or addressed in the report?*
Per Johan Brandvik	The definition of Dispersant Effectiveness (DE) and volume fraction below 70 microns are well defined and relevant. The documentation of released oil, applied dispersant and DOR are all relevant and well described. The documentation of droplet sizes measured is also well documented. The data from replicate testing of each dispersant are available both for DE and volume of droplets < 70 microns. The use of the droplet size data is a weak point in the report. The method used to quantify oil droplet sizes is based on laser scattering (LISST 100X). This instrument has an upper detection limit of 500 microns. The shape of (most) cumulative distribution curves shows that there is a significant amount of particles

<table>
<tr><td></td><td>larger than 500 microns. These cumulative curves should not be used to calculate the median droplet diameter, since they only describe a fraction of the present droplets (see table 5). This is not at all discussed in the report.

The need for some of the statistical methods applied is not clear. I think simpler and better known statistical methods could have been applied. My understanding is that the statistic methods presented; Stepwise Regression, Breusch-Godfrey test and the Tukey Honestly Significant Difference (Tukey HSD) test is used to:

1. Find the significant experimental variables that influence DE (Stepwise regressions)

2. Find if there is a surfactant memory in the system affecting DE (Breusch-Godfrey)

3. Find significant difference between products and control (DE and vol. < 70 microns) Tukey HSD.

The first and second methods are appropriate, but the third seems over complicated. A simpler student t-test comparing each dispersant with the control would be simpler, sufficient and maybe most important, and much easier to communicate to the reader.</td></tr>
<tr><td>Merv Fingas</td><td>The arithmetic to calculate the actual effectiveness appears to be incorrect. The value of the controls should be subtracted from any values presented – standard scientific procedure. The values should be:

<table><tr><th>Product</th><th>Reported Value</th><th>Test Value</th></tr><tr><td>Corexit</td><td>22.9</td><td>72.7</td></tr><tr><td>Finasol</td><td>22.4</td><td>72.2</td></tr><tr><td>Accell</td><td>1.5</td><td>51.3</td></tr><tr><td>ZI 400</td><td>-4.1</td><td>45.7</td></tr><tr><td>Control</td><td>0</td><td>49.8</td></tr></table></td></tr>
<tr><td>Julien Guyomarch</td><td>Results of oil concentrations are not clearly explained, and there are no units on figures. The way the two sets of data resulting from the LISST measurements were combined is not explained. The statistical analysis is also not clear and figure (with error bars) would have been preferred to tables with too many significant numbers.

The threshold value of 70 µ m is also not discussed whereas it would have been interesting to reassess this limit based on a big set of data obtained at a large scale.

Finally, one important result has not been discussed: the control is very high, and the effects of dispersants are expressed as a gain in percentage compared to this condition. Moreover, the control shows a significant variability (from 43.0 to 59.7), thus suggesting that the comparison between dispersants should include this</td></tr>
</table>

	key point. A comment or an explanation on the increase of the oil concentration over time for the lower LISST (test 8) should be added.
George A. Sorial	As mentioned above in the overall view of the report, the weakness in point 1 to 3 should be addressed.
Albert D. Venosa	The sampling and statistical testing were adequately described, but unfortunately the methods were not fully appropriate in my opinion. First of all, an analysis of variance (ANOVA) should have been performed at the conclusion of all tests. This is usually what researchers do to determine if there were any significant differences noted in the experiment. If there were none, then there would be no need to conduct any further statistical tests. Only when statistical differences are determined, then the HST test should be done to determine where the significances existed. Instead, the authors used stepwise multiple regression analysis to establish a best-fit model to keep only those variables that explain the variability in the data. They then used the HST to identify pair-wise differences in regard to DE, % oil droplet sizes < 70 µm, and oil concentration having droplet sizes < 70 µm. In their stepwise multiple regressions, they asked how significant salinity, interfacial tension, and viscosity were in explaining the variability. But, since the same seawater and tank were used in all experiments without emptying the tank and refilling it with fresh seawater, studying the influence of salinity is a waste of time. It turned out (Table 6) that salinity played a role for the dependent variable "% below 70 µm", which I do not understand, as did viscosity. None of those 3 variables significantly affected DE or concentration below 70 µm (unless I'm reading Table 6 incorrectly). Frankly, I don't understand the blank cells in the table. If they tested all 3 variables, then there should be probability values (p-values) in all table cells. Tables 7-9 contained the critical statistical data. In Tables 8 and 9, only 2 pairwise significances were observed. For DE (Table 8), only the pairs ZI400-Corexit and ZI400-Finasol were statistically different from each other. For the < 70 µm concentrations (Table 9), only the pairs Finasol-Control and ZI400-Finasol were statistically different from each other. Table 7 shows 7 different pairs showing significant differences from each other. These results beg to be discussed and explained in greater detail but were not.

5	*Are the findings and overall discussion of the results for each product tested clearly discussed?*
Per Johan Brandvik	No, results from the statistical methods (discussed above) are presented in table 6, 7, 8, and 9, they should have been be used in the subsequent discussion and conclusion sections. That's unfortunately not the situation. The results in these tables are presented, but I can't find a single reference to these tables in the "Discussion" or "Conclusion" sections. The discussion section is dominated by a justification of why a fifth dispersant was omitted and additional operational observations like spray patterns, challenges with high viscosity, and possible foaming. None of the statistical material is discussed.

	One example is that table 8 shows that DE for neither C9500 nor OSR-52 is significantly higher than the control. This is probably due to the deviation (±6) for the two products is large compared to the difference between the products and the control (22-23). This seems strict. An ordinary T-test would have probably shown that they were significant! However, this and several other results from the statistical analysis, are not discussed or referred to in the conclusions.
Merv Fingas	Yes, there should have been more discussion on the removal of data point from one of the dispersants. This should have been done on a statistical basis, not on an ad hoc basis.
Julien Guyomarch	Results and conclusions obtained for each product are generally well discussed, with the limitations mentioned above. Once again, the high level of the efficiency for the control, combined with a great variability, should prompt reservations to conclusions.
George A. Sorial	The discussions of the results are clearly described with exception to the three weakness points described above.
Albert D. Venosa	Based on the above responses, it should be clear that the quantitative, statistical findings and overall discussion of results were inadequately discussed and/or explained that would lead to a more informed decision on which dispersant product should be used in an oil spill. Many unanswered questions remain to the reader. The dispersant product selection was inadequate since two dispersants that have been well tested in the literature were not tested in the project, namely, SPC 1000 and JD2000. Admittedly, the authors had problems with SPC 1000, but I am not convinced, based on their reporting, that they tried hard enough to overcome the viscosity and application problems of the product. As for JD2000, this product should have been chosen based on the literature, but the authors demonstrated an unawareness of the literature by not even mentioning this product in their selection procedure.

6	*Are the conclusions logical and appropriate based on the results? Are there any additional conclusions that could be drawn?*
Per Johan Brandvik	No, the conclusions are not logical. The report has a comprehensive statistic chapter, but very little of this information is used in the conclusions. My evaluation of the presented data is that two of the dispersants have a significantly better effectiveness than the control (C9500 and OSR-52) based on both DE and volume of droplets < 70 microns. However, the report concludes (abstract) that: ▪ "Corexit EC9500A performed very well …. producing the highest average Dispersant Effectiveness (DE)" ▪ "Finasol OSR 52 demonstrated a performance close to that of Corexit"

	"Highest average" and "Close to" are terms which are not very comparable to the statistical ambitions earlier in the report. Based on the presented data, it is not possible to see a significant difference between DE of C9500 (73% ±6) and OSR-52 (72±6), see table 3. The C9500 has also a higher DOR (5%) than OSR-52 (3%), which is not discussed in the report. This ranking is highly surprising based on the presented data and especially based on the use of advanced statistical methods in other sections in the report.
Merv Fingas	Notwithstanding the arithmetic of the output values, more emphasis might have been put on discussing the practicality of this in the Arctic. One cannot heat the oil and dispersants there, so this reviewer does not see the application to field. This should have at least been discussed.
Julien Guyomarch	See section 4.
George A. Sorial	The conclusions were clearly provided for the four dispersants studied.
Albert D. Venosa	No, the conclusions are not logical and appropriate. As I stated early in this review, the statistical analyses were flawed to begin with, but even if one accepts the analyses, the authors clearly displayed positive biases toward Corexit and Finasol. Although, based on their qualitative observations about the lack of foaming by Corexit and Finasol and the presence of foaming by ZI 400 and Accell that may have negatively impacted efficient and effective applications of the latter products to the slick, such observations are indeed helpful in allowing good decisions to be made by decision-makers against the use of the latter two products in the Arctic and in favor of the former two. However, the lack of statistical significances between Finasol, Corexit, and Accell do not help in such decision making.

7	*Does this report present sufficient new data and knowledge, and are the findings useful for informing oil spill response planning in the Arctic regions?*
Per Johan Brandvik	The main operational important messages extracted by me from this data material are: 1. There are larger differences in effectiveness between the dispersant on the EPA list. 2. Some of them perform excellent and should be preferred (stockpiled) by industry and authorities. 3. Some of the dispersants on the list have a very low effectiveness (since the threshold to get into the list is very low) and should NOT be used or stockpiled as marine oil spill dispersants. 4. More realistic and challenging testing is needed to get into the EPA list. This should be presented more clearly in the report.

	Another operational relevant fact is that dispersants are in a real situation not applied to fresh oils. In most cases it takes a couple of hours to take a decision (obtain a permit?), get the dispersant to the site, detect the slick and initiate dispersant spraying. This mean that dispersant usually are sprayed on slightly weathered oil, were emulsification is the most significant process, raising the viscosity due to increased water content. This gives the dispersant other challenges than dispersing water free fresh oil (penetrating viscous emulsions, increased oil pour point due to evaporation, breaking emulsions before dispersing the oil etc.). The reduced relevance of performing the testing on fresh oil is not addressed and this is a major weakness of the report.
Merv Fingas	As noted, the applicability to the Arctic is in question. Furthermore the presentation of the correct results (value – control) would have been helpful. The correct results, even with their weaknesses, show that dispersant application to the Arctic would be questionable, although this does not come out in the report.
Julien Guyomarch	The questions raised by the Arctic conditions, in terms of oil spill response, in relationship with dispersant application, is based on environmental issues, and also on efficiency of techniques. This study well address the possibility of applying dispersants, from a technical point of view. However, even if tests are performed at low temperatures, these temperatures are not totally representative of arctic conditions. In particular, only low pour point oils can be treated, while it is not the case of this study. It should have been interesting to compare dispersants at the laboratory scale at very low temperatures, to clearly identify the influence of this parameter, and then to conduct large scale experiments by varying the oil condition and not only the dispersant products.
George A. Sorial	Yes, the report provides new data that could be representative to actual spill scenario.
Albert D. Venosa	I partially answered this question above in Question 6. I think the only “data” that might be useful in future decision-making for Arctic oil spills are the qualitative observations on foaming of two of the four products.

4.3 Specific Observations

Specific Observations			
Provide specific observations or comments on the report mentioning page and paragraph (expand table if needed).			
	Page	***Paragraph***	***Comment or Question***
Per Johan Brandvik	4	first	Droplets of 70 microns are not neutrally buoyant, but kept dispersed due to the turbulence level in the top 10-30 meters.
	7	second	Ohmsett water depth is 2.45 meter
	7	second	These wave types (Pierson-Moskowitz and JONSWAP) need a reference or some explanations.
	9	Table 1	Wax content should be included, important property for dispersant effectiveness and oil behavior
	10	Table 2	Sulfur content? Not very relevant for oil weathering or dispersant effectiveness
	15	Figure 5	This figure shows that there are significant amounts of droplets larger than 500 microns present
	17	Table 5	Two significant digits in MVD can't be justified
	17	Table 6	A very high number of significant digits can't be justified
	18	Table 7	A very high number of significant digits can't be justified
	19	Table 8-9	A very high number of significant digits can't be justified
	20	Figure 8	All the space used to justify that SPC100 was not included in the report (half of the discussion section?) looks strange.
	Page	***Paragraph***	***Comment or Question***
Merv Fingas	every	most	Values and units are placed together – this is incorrect and is correct only for
	most	most	Significant figures – for all numbers in the report should be either 2 or 3 (at most)
	many	many	Droplet size diagrams do not have units on x scale
	most	many	Arctic should always be spelled with capital A
	many	many	µ should be used on ‘u’
	Page	***Paragraph***	***Comment or Question***
Julien Guyomarch	9	Table 1	Pour point seems high (10°C) considering the test temperatures. It is also high for a slightly paraffinic oil (2.77 % wt.). Only one significant figure should be enough for contents in %.

	10	Table 2	Test oil is actually different from the other U.S. Arctic oils based on the API gravity. Additional parameters such as wax and asphaltenes content, viscosity, distillation data would give more information when dealing with oil dispersibility.
	13	Figure 4	Information should be added on the concentrations units, and also how the data are processed (collected vs used).
	14	Table 3	DOR is identical for the 3 dosing rates of Finasol.
	15	Table 4	The column "improvement over Corexit" is not very useful (already one relative difference with control).
	15	Figure 5	Cumulative distributions are not very clear as the maximum do not appear and should be replaced by "classical" distributions.
	16	2nd	The 70 µm cutoff should have been discussed as it refers to an "old" paper (which many papers also refer to…).
	17	Table 5	Too many significant figures.
	17	Table 6	Too many significant figures. Explain the difference between *%* and *concentration*.
	18	Table 7	Too many significant figures. A figure with error bars should be preferred to a table.
	19	Table 8 & 9	Too many significant figures. Two-to-two comparisons should be replaced by a general scheme.
	23	2nd	Comparisons are performed without considering any variability of the results. In addition, the highest DOR of the Corexit (0.09) was between 2 and 3 times more than for the other products (from 0.03 to 0.05).
		Appendix B	Specify test conditions for each test, rather than only the test number (Control, Corexit Low, Corexit High…).
		Appendix C	To be placed before Appendix B.
		Appendix D	Cumulative distributions are not very clear to identify the mean size.
	Page	***Paragraph***	***Comment or Question***
George A. Sorial	17	Table 6	Provide values to two decimal places
	18	Table 7	Provide values to two decimal places
	19	Table 8	Provide values to two decimal places
	19	Table 9	Provide values to two decimal places
		Appendix B	Concentration units should be provided within the plot and the run number should be clearly identified with respect to dispersant used or oil control
	Page	***Paragraph***	***Comment or Question***

Albert Venosa	1	2	Another dispersant product, SPC 1000, was pre-tested but found to be too viscous to apply via the pressure spray apparatus.
	1	3	Second sentence and onwards: this statement, plus the ones in the subsequent sentences, needs to be couched in terms of statistical significance. Otherwise, it points out a possible positive bias in favor of Corexit over the other products tested. Such bias should be avoided.
	3	1	Line 4: Minor point. It seems incongruous to discuss ice in terms of "concentration." Perhaps better terms might be "density or area of consolidated pack ice" or "ice cover formed by the packing and freezing together of pack ice floes."
	3	1	2nd last line: Need references here to support the claim that 30 years of dispersion effectiveness data in Arctic conditions are available.
	3	2	Lines 2 and 3: Is there a supportive reference for this statement? JD2000 has also generated interest for publications that have appeared in the peer-reviewed literature.
	4	1	Lines 3 and 4: Dispersants may also contain solvents that tend to decrease viscosity for ease of application.
	4	1	Line 8: Regarding droplets of size < 70 µm can be consumed by microorganisms: this statement implies that droplets > 70 µm cannot be consumed, which is not true. The bigger the droplet size, the lower is the surface area that can be biodegraded, so rates are lower. So, add “more rapidly” after the word “consumed.”
	4	1	Line 10: the Conover reference is extremely old (1971). Much mesocosm research has been done in early to mid-2000's, mostly at the wave tank at DFO-Canada and at Ohmsett. The authors seem not to have reviewed the literature.
	5	2	First sentence, actually five dispersants were selected. One (SPC 1000) was not tested due to problems with viscosity and throughput in the spray apparatus. In the absence of SPC 1000, they should have tested JD2000 because of previous published research on that product.
	5	2	About halfway down: the term “concentration” should be changed to volume concentration, ppmv.

	6	7	In the numbered bullet #7, the Baffled Flask Test is replacing the inferior Swirling Flask Test by EPA, so that regulation will soon be replaced.
	6	11	Venders must also submit to EPA toxicity data from testing mysid shrimp (*Americamysis bahia*) and silversides (*Menidia beryllina*).
	6	Last	The product ZI 400 should never have been selected for testing. It is a very poor dispersant as shown in published lab data.
	7	1	Line 3: Text should read SPC 1000 was replaced by Accell.
	7	1	Last line of 1st paragraph: Why didn't the researchers contact the manufacturer and ask for another sample? Other readers will undoubtedly ask this same question.
	7	2	Last 3 lines before bullets: dispersant concentrations < 400 ppb do not affect the outcome of tests, as referenced by SL Ross in 2000. The reference was not from a peer reviewed paper and had no statistical data to support it.
	8	1	The oil was not weathered. It was a blend of 2 unweathered oils. The freshness should be pointed out.
	8	1	Line 7: Ohmsett developed, not measured, a temperature-viscosity curve…
	10	1	Line 1: the term “replicates” is misleading. Replicate samples or replicate experiments? Big difference. Later it says the experiments were repeated, so this word should be "replicate experiments conducted in random order." Was the order randomized? It should have been.
	10	2	Tests continued for 20 min, after which they were terminated. The tank was allowed to settle for an additional 3 hours before prepping it for another test. The 20 min is not nearly long enough to allow recoalescence to occur. Nor was 3 hours. They should have waited 24 hours (or at least overnight) before prepping for the next run. This is very important, and it affects conclusions on total dispersion effectiveness.
	11	1	Last line: did they warm the dispersant product SPC 1000 like they did the others? No statement appears in the text regarding this point.
	11	2	Last 3 lines: Was the assumption tested and confirmed that any resurfaced oil would stay dispersed? This is very important and could have a

			bearing on results. The resurfaced oil should have been measured and included in the subsequent calculations.
	11	3	Lines 1 and 2 before the equation: By their own definition, surface oil that no longer remains on the surface should be included in the Vor after sufficient time is allowed for re-coalescence and surfacing. No mention of waiting a defined amount of time was made.
	11	3	In the equation, the last term should be (1 – fraction of water in the source oil), not (1 – % water in source oil). Same for both equations.
	12	Last	As mentioned above, the 20 min time period for the experiment was too short for recoalescence to have occurred. This is a major flaw in the design.
	13	1	Line 3: the concentration should be reported in units of ppmv, as it also should on the y-axis of Figure 4. Do not let the reader make assumptions.
	13	Fig. 4	What does “data used” mean? Used for what? Why did they not include the grayed out data? Also, can time also be superimposed on the x-axis? Sample number means nothing to the reader.
	13	3	First paragraph under Statistical Analysis: (1) did salinity vary that much to be a confounding variable, especially since the same water was used for all experiments. Also, why did they not use ANOVA rather than stepwise regression analysis?
	14	1	Line 1: Should be ppb, not ppm, as specified on p. 7
	14	2	Last paragraph before Results section: The Tukey HSD test is a post-hoc test, meaning that it is performed after an analysis of variance (ANOVA) test. This means that to maintain integrity, a statistician should not perform Tukey's HSD test unless he has first performed an ANOVA. It's not clear that this was done. The ANOVA determines if there are significant differences, but the HSD tells you where those differences occurred. It appears the authors used the HST to determine if there were differences between any 2 treatments independent of an overall ANOVA, which the HST was not designed to do, at least initially.
	14	Table 3	The 49.8 mean is really high for the control! This points out the disadvantage of using a batch wave tank with no flow-through because the dynamic dispersion effectivenes (DDE) cannot be calculated. See Li et al., 2009, “*Evaluating crude oil chemical*

			dispersion efficacy in a flow-through wave tank under regular non-breaking wave and breaking wave conditions." Marine Pollution Bulletin, 58(5): 735-744.
	14	Last	Corexit had the lowest DOR, which biased the test since more dispersant was used at the lower DOR. The 0.05 value is 67% higher than 0.03 for Finasol and Accell and 25% higher than for ZI 400. Not a trivial difference.
	15	1	In the statement "Finasol performed almost identically to Corexit…," the authors should add "with significantly less dispersant".
	15	1	Line 2: need to report the p-value after DE = 72.2%.
	15	1	Line 5: Regarding the Test 10 outlier, was this outlier statistically determined as an outlier using a test like the Grubb's test? Cannot eliminate the outlier without doing a statistical significance test. Or better, do a repeat of this test.
	15	1	Line 9: The poor performance of ZI 400 was not unexpected since lab studies showed this dispersant does not work!
	16	2	Line 2: Are these bar graphs composites of the 3 experiments? They should be.
	16	2	Line 4: According to the manufacturer, Finasol is identical to C9500.
	16	2	Line 7: Is this conclusion, that Accell was not as effective as Corexit or Finasol, based on statistical analysis or eyeballing? I see no p-value in support of this statement.
	16	2	Line 9: In the line "The oil that was dispersed (by Accell) consisted of a larger percentage of very small droplets over the dispersed oil from Finasol…," again, was this statistically based or eyeballed? If statistically based, and explanation is needed.
	16	2	Line 17: Data mentioned are not shown on Fig. 6. Also, what about Corexit? No mention made of droplet sizes < 60 um.
	17	Top	Fig. 7, why not show the LISST data for all 3 replicate experiments? Would it result in a too busy chart?
	17	1	The median droplet size in the control test was 457.26 µm. This and subsequent reporting of droplet sizes should have significant figures no greater than 3 at most. It's silly to use 5 or more significant figures.

	17	1	In line 7, need p-values here to justify the word "significant."
	17	Table 6	Second row, last column, I don't understand this significance. Salinity was virtually constant (same tank, same water) across all runs. Makes little sense. Also, what is the meaning of a blank cell in the table? Very strange table.
	18	2	As stated previously, the authors should have first run an ANOVA to determine if there were significant differences to begin with. THEN, they should run the HST to determine where the significances were. Also, the table has WAY too many significant figures in the data. Confine them to max 3, perhaps even 2.
	19	1	Table 8: This table says there were no significant differences anywhere except between ZI400 and Corexit and between ZI400 and Finasol. This suggests low statistical power, and more replicate experiments should be done to improve the power of the test. Otherwise, everything in the report is reduced to qualitative statements.
	19	2	Table 9: This table is also revealing (in a negative sense). No differences were observed in the < 70 um densities except for Finasol-Control and ZI400-Finasol. This must be explained.
	20	1	Line 6: Was SPC 1000 heated as the other products? I presume so, but it was never so stated anywhere in the report.
	20	1	Bottom fourth of the paragraph, regarding the phone call to SL Ross, by whom? No need to ID the caller. Just mention if it was the manufacturer.
	20	1	End of the paragraph, why didn't the researchers just ask the manufacturer to give them another sample of SPC1000? And, did they discuss these problems with the manufacturer to determine if temperature was an issue or if foaming was an issue at low temperature but not at warmer temperatures?
	21	1	Line 1, the sentence implies positive bias toward Corexit. It should be deleted.
	21	1	The last sentence of the paragraph is key. Recoalescence should have been quantified. What would have resurfaced if they had waited 24 hours after termination?
	21	3	Line 3: Droplets remaining close to the surface implies recoalescence occurred. Did this appear to increase over time? Also, it indicates bigger

			droplets. However, according to Tables 7-9, this observation is moot since there weren't any statistical differences between Accell and the other products except ZI400.
	21	3	Line 7: Regarding the froth increasing after each test, this suggests persistent residuals between treatments.
	21	3	Line 9: But the observation was made ONLY with Accell. Why do they say it cannot be definitely associated with Accell?
	22	1	Line 3: Performance of ZI400 was poor anyway, so why do they conclude that the foam did not affect the liquid product? How do they know it didn't affect performance?
	22	1	Last sentence: nozzles were clogged and the system had to be thawed. This was despite the warming of the product before testing?
	22	1	In the report, there was absolutely no discussion about how the Ohmsett results compared with previously published tests in the literature both in the lab and in mesocosm studies. This type of scholarly discussion is mandated by peer-reviewed journals. Although this is not a journal article, nonetheless discussion of the data relative to what is already known puts everything in perspective and suggests how further research should be conducted (or how even this research should have been conducted).
	23	2	Line 1: This statement is merely subjective and not quantitative because of the lack of statistical significance as demonstrated in the tables.
	23	2	The parenthetical should be eliminated and Finasol should also be part of the subject of the sentence.
	23	1	Line 4: After “46% improvement”, they should also state “but not statistically significant.”
	23	1	Line 5: After the word “Corexit”, the words “and Finasol” should be added.
	23	1	Last line: The temperature range was miniscule, so this statement applies to all products. Again, bias!
	23	2	Last line of the page: Please include the p-values, which contradict the significance of these statements.
	24	2	Line 2: This cannot be assumed without an outlier test. Easy to do with the Grubbs' test.

	24	2	Lines 5 and 6: Here is but one example where comparison with other reported values in the literature may have shed some light on the subject.
	24	3	Line 3: The p-value was not significant.
	25	1	Regarding the calling for large scale testing, why large scale? Lab tests should suffice since everything would be controlled by the analyst. The wave tank at DFO-Canada can be used more quantitatively and with flow-through so that dynamic dispersion effectiveness (DDE) can be calculated. The Ohmsett tank is too large to allow flow-through, which argues against using it for statistical comparisons of chemically based products.
	25	1	End of paragraph: Lab testing with the soon to be enacted Baffled Flask Test would be much easier and less costly. In fact, EPA will be requiring testing at 2 temperatures (5 and 25 C) and 2 oils once the test has been adopted.
	25	2	Last sentence: Please explain this last sentence. It implies residuals may have affected performance between runs.

5. APPENDIX A: INDIVIDUAL REVIEWER COMMENTS

5.1 Per Johan Brandvik

Peer Review of the BSEE Report *"Comparative Testing of Corexit EC9500A, Finasol OSR 52, Accell Clean DWD, and ZI 400 at Ohmsett in a Simulated Arctic Environment."*	
NAME:	Dr. Per Johan Brandvik, Senior Research Scientist/Professor
AFFILIATION:	SINTEF, Oil Spill Research, Trondheim, Norway
DATE:	25 June 2015
I. GENERAL IMPRESSIONS	
Provide overall impressions (approximately 1/2 page in length) addressing the accuracy of information presented, clarity of presentation, and soundness of conclusions.	
Comments:	The Ohmsett facility offers unique possibilities regarding large-scale dispersant testing and the description of the experimental procedures are clear and concisely described. The amount of work performed is impressive and all experiments are performed with three replicative. All the data needed to evaluate the data analysis and conclusions in the report are available in tables in the report and in the appendices. However, there are some shortcomings in the report, the most significant are: ▪ The need for and relevance of some of the advanced statistical methods applied are not clear, simpler and more well-known statistical methods could have been applied. ▪ Very little of the statistics presented in multiple tables are used in the discussion and conclusions. ▪ The main conclusion is conflicting with the statistical analysis. ▪ There is no evidence in the data material supporting a (significant) difference in performance between C9500 and OSR-52 ▪ The method used to quantify dispersed oil droplets has an upper detection limit of 500 microns, but there is a substantial fraction of larger droplets present. The cumulative distribution curves should not be used to calculate the median droplet diameter, since they only describe a fraction of the droplets. The main conclusions stating that C9500 is best, followed by OSR-5 is not justified by the data. They are both significantly better than the control and the other product tested. The other products are not significantly better than the control.

II. RESPONSE TO CHARGE QUESTIONS	
Provide narrative responses to each of the nine charge questions below.	
1	*Are the objectives of the study clearly defined? If not, what are your recommendations for improving the description of objectives?*

Comments:	The objective of the study is clearly defined (*to better understand and compare the effectiveness of various dispersants under simulated arctic test conditions*) and very relevant for both the industry and authorities which need to improve Arctic oil spill contingency.

2	*Was the methodology used to define the selection and testing conditions of the four dispersants clearly described?*
Comments:	No, C9500/OSR-52 were selected due to their large volume in U.S. stockpiles, but the only criteria listed for selecting the other two products was because they were "*not as well known*". Selecting all products based on stockpiled amounts would be a better approach.

3	*Were the testing procedures used appropriately describe and properly implemented?*
Comments:	The Ohmsett facility offers unique possibilities regarding near field conditions for dispersant testing and the description of the experimental procedures are clear and concisely described. All the data needed to evaluate the data analysis and conclusions in the report are available in tables in the report and in the appendices.

4	*Were the results of the sampling and statistical testing adequately characterized and clearly described? Were there any critical results or limitations that were not discussed or addressed in the report?*
Comments:	The definition of Dispersant Effectiveness (DE) and volume fraction below 70 microns are well defined and relevant. The documentation of released oil, applied dispersant and DOR are all relevant and well described. The documentation of droplet sizes measured is also well documented. The data from replicate testing of each dispersant are available both for DE and volume of droplets < 70 microns. The use of the droplet size data is a weak point in the report. The method used to quantify oil droplet sizes is based on laser scattering (LISST 100X). This instrument has an upper detection limit of 500 microns. The shape of (most) cumulative distribution curves shows that there is a significant amount of particles larger than 500 microns. These cumulative curves should not be used to calculate the median droplet diameter, since they only describe a fraction of the present droplets (see table 5). This is not at all discussed in the report. The need for some of the statistical methods applied is not clear. I think simpler and better known statistical methods could have been applied. My understanding is that the statistic methods presented; Stepwise Regression, Breusch-Godfrey test and the Tukey Honestly Significant Difference (Tukey HSD) test is used to: 4. Find the significant experimental variables that influence DE (Stepwise regressions)

	5. Find if there is a surfactant memory in the system affecting DE (Breusch-Godfrey) 6. Find significant difference between products and control (DE and vol. < 70 microns) Tukey HSD. The first and second methods are appropriate, but the third seems over complicated. A simpler student t-test comparing each dispersant with the control would be simpler, sufficient and maybe most important, and much easier to communicate to the reader.

5	*Are the findings and overall discussion of the results for each product tested clearly discussed?*
Comments:	No, results from the statistical methods (discussed above) are presented in table 6, 7, 8, and 9, they should have been be used in the subsequent discussion and conclusion sections. That's unfortunately not the situation. The results in these tables are presented, but I can't find a single reference to these tables in the "Discussion" or "Conclusion" sections. The discussion section is dominated by a justification of why a fifth dispersant was omitted and additional operational observations like spray patterns, challenges with high viscosity, and possible foaming. None of the statistical material is discussed. One example is that table 8 shows that DE for neither C9500 nor OSR-52 is significantly higher than the control. This is probably due to the deviation (±6) for the two products is large compared to the difference between the products and the control (22-23). This seems strict. An ordinary T-test would have probably shown that they were significant! However, this and several other results from the statistical analysis, are not discussed or referred to in the conclusions.

6	*Are the conclusions logical and appropriate based on the results? Are there any additional conclusions that could be drawn?*
Comments:	No, the conclusions are not logical. The report has a comprehensive statistic chapter, but very little of this information is used in the conclusions. My evaluation of the presented data is that two of the dispersants have a significantly better effectiveness than the control (C9500 and OSR-52) based on both DE and volume of droplets < 70 microns. However, the report concludes (abstract) that: ▪ "Corexit EC9500A performed very well …. producing the highest average Dispersant Effectiveness (DE)" ▪ "Finasol OSR 52 demonstrated a performance close to that of Corexit" "Highest average" and "Close to" are terms which are not very comparable to the statistical ambitions earlier in the report. Based on the presented data, it is not

	possible to see a significant difference between DE of C9500 (73% ±6) and OSR-52 (72±6), see table 3. The C9500 has also a higher DOR (5%) than OSR-52 (3%), which is not discussed in the report. This ranking is highly surprising based on the presented data and especially based on the use of advanced statistical methods in other sections in the report.

7	*Does this report present sufficient new data and knowledge, and are the findings useful for informing oil spill response planning in the Arctic regions?*
Comments:	The main operational important messages extracted by me from this data material are: 5. There are larger differences in effectiveness between the dispersant on the EPA list. 6. Some of them perform excellent and should be preferred (stockpiled) by industry and authorities. 7. Some of the dispersants on the list have a very low effectiveness (since the threshold to get into the list is very low) and should NOT be used or stockpiled as marine oil spill dispersants. 8. More realistic and challenging testing is needed to get into the EPA list. This should be presented more clearly in the report. Another operational relevant fact is that dispersants are in a real situation not applied to fresh oils. In most cases it takes a couple of hours to take a decision (obtain a permit?), get the dispersant to the site, detect the slick and initiate dispersant spraying. This mean that dispersant usually are sprayed on slightly weathered oil, were emulsification is the most significant process, raising the viscosity due to increased water content. This gives the dispersant other challenges than dispersing water free fresh oil (penetrating viscous emulsions, increased oil pour point due to evaporation, breaking emulsions before dispersing the oil etc.). The reduced relevance of performing the testing on fresh oil is not addressed and this is a major weakness of the report.

III. SPECIFIC OBSERVATIONS

Provide specific observations or comments on the report mentioning page and paragraph (expand table if needed).

	Page	***Paragraph***	***Comment or Question***
	4	first	Droplets of 70 microns are not neutrally buoyant, but kept dispersed due to the turbulence level in the top 10-30 meters.
	7	second	Ohmsett water depth is 2.45 meter

	7	second	These wave types (Pierson-Moskowitz and JONSWAP) need a reference or some explanations.
	9	Table 1	Wax content should be included, important property for dispersant effectiveness and oil behavior
	10	Table 2	Sulfur content? Not very relevant for oil weathering or dispersant effectiveness
	15	Figure 5	This figure shows that there are significant amounts of droplets larger than 500 microns present
	17	Table 5	Two significant digits in MVD can't be justified
	17	Table 6	A very high number of significant digits can't be justified
	18	Table 7	A very high number of significant digits can't be justified
	19	Table 8-9	A very high number of significant digits can't be justified
	20	Figure 8	All the space used to justify that SPC100 was not included in the report (half of the discussion section?) looks strange.

5.2 Merv Fingas

Peer Review of the BSEE Report *"Comparative Testing of Corexit EC9500A, Finasol OSR 52, Accell Clean DWD, and ZI 400 at Ohmsett in a Simulated Arctic Environment."*

NAME:	Merv Fingas
AFFILIATION:	Spill Science
DATE:	May 20, 2015

I. GENERAL IMPRESSIONS

Provide overall impressions (approximately 1/2 page in length) addressing the accuracy of information presented, clarity of presentation, and soundness of conclusions.

Comments:

The report addresses the terms of reference and is generally accurate, but could be improved in several ways:

Calculation of values – It is standard practice to report values as the test value less the control value, thus the values here should be:

Product	Reported Value	Test Value
Corexit	22.9	72.7
Finasol	22.4	72.2
Accell	1.5	51.3
ZI 400	-4.1	45.7
Control	0	49.8

(Table 4 tries to show this, but the math is unclear or incorrect.)

Secondly, although the established protocol at Ohmsett is used, comments should be made on its applicability to this study. The study was to establish dispersibility of oils at low temperatures; however, the temperature and pressure of oil release were not noted. This could be very important. Importantly, the dispersants had to be heated to be discharged. Does this not rule out Arctic dispersion? There is no practical way that dispersants could be heated in the Arctic and certainly they would cool to frozen droplets (snow) by the time they hit the water. This should at least be highlighted. Another point about the protocol is that the 20 minutes of tests is certainly very rapid. The rise time of droplets would be well above this time, thus the test does not count semi-stable droplets that would rise after the time of the water sweeping. A time of about 3 hours might be more appropriate

The **70 µm cutoff** for droplets may be accepted, but the origin (from Lunel, 1993) is less than scientific. This cutoff point happens to be the range of the particular instrument, Lunel was using. This is a comment on a value that has suddenly become magical. A better cut-off might be 50 µm.

The **SPC 1000** should have not been diluted with water. This is certainly part of the reason for the foam production.

More **justification** is needed for the removal of the data point for Accell test. Perhaps a statistical test should be used.

	Appendix B is hardly needed in full as now presented. The size data is not particularly useful in graphical form.

II. RESPONSE TO CHARGE QUESTIONS	
Provide narrative responses to each of the nine charge questions below.	
1	*Are the objectives of the study clearly defined? If not, what are your recommendations for improving the description of objectives?*
Comments:	Yes, and I believe that the report follows the study objectives.

2	*Was the methodology used to define the selection and testing conditions of the four dispersants clearly described?*
Comments:	Yes, and the selection did follow the criteria set down.

3	*Were the testing procedures used appropriately describe and properly implemented?*
Comments:	The report followed the established procedures; however, comments should have been made on the need to heat the dispersant. It is unclear if the oil was heated and at what pressure it was discharged, further the viscosity of the oil appears to be only calculated. The time after the test (20 min.) that sweeping was started is too short and would result in missing much of the resurface oil.

4	*Were the results of the sampling and statistical testing adequately characterized and clearly described? Were there any critical results or limitations that were not discussed or addressed in the report?*
Comments:	The arithmetic to calculate the actual effectiveness appears to be incorrect. The value of the controls should be subtracted from any values presented – standard scientific procedure. The values should be:

Product	Reported Value	Test Value
Corexit	22.9	72.7
Finasol	22.4	72.2
Accell	1.5	51.3
ZI 400	-4.1	45.7
Control	0	49.8

5	*Are the findings and overall discussion of the results for each product tested clearly discussed?*
Comments:	Yes, there should have been more discussion on the removal of data point from one of the dispersants. This should have been done on a statistical basis, not on an ad hoc basis.

6	*Are the conclusions logical and appropriate based on the results? Are there any additional conclusions that could be drawn?*
Comments:	Notwithstanding the arithmetic of the output values, more emphasis might have been put on discussing the practicality of this in the Arctic. One cannot heat the oil and dispersants there, so this reviewer does not see the application to field. This should have at least been discussed.

7	*Does this report present sufficient new data and knowledge, and are the findings useful for informing oil spill response planning in the Arctic regions?*
Comments:	As noted, the applicability to the Arctic is in question. Furthermore the presentation of the correct results (value – control) would have been helpful. The correct results, even with their weaknesses, show that dispersant application to the Arctic would be questionable, although this does not come out in the report.

III. SPECIFIC OBSERVATIONS

Provide specific observations or comments on the report mentioning page and paragraph (expand table if needed).

	Page	***Paragraph***	***Comment or Question***
	every	most	Values and units are placed together – this is incorrect and is correct only for
	most	most	Significant figures – for all numbers in the report should be either 2 or 3 (at most)
	many	many	Droplet size diagrams do not have units on x scale
	most	many	Arctic should always be spelled with capital A
	many	many	μ should be used on ‘u’

5.3 Julien Guyomarch

Peer Review of the BSEE Report *"Comparative Testing of Corexit EC9500A, Finasol OSR 52, Accell Clean DWD, and ZI 400 at Ohmsett in a Simulated Arctic Environment."*	
NAME:	Julien Guyomarch
AFFILIATION:	Cedre (France)
DATE:	June 19, 2015

I. GENERAL IMPRESSIONS	
Provide overall impressions (approximately 1/2 page in length) addressing the accuracy of information presented, clarity of presentation, and soundness of conclusions.	
Comments:	This test program, conducted in order to better understand the effect of various dispersants under arctic test conditions, is interesting as there is definitely a need in this field. In addition, experiments were conducted at a large scale, with replicates and controls, and the quality of the dispersion was monitored through many measurements, in particular the droplets size distribution. It also appears that significant efforts were made to control the dispersant application, which was then realistic. Syntheses of results were well presented in synthetic tables, thus clearly showing the effect of the various factors. However, considering the importance of the work, more details on the experimental results would be useful. In particular, the way the concentrations were determined is not clear, as well as the date concerning the processing of the two sets of droplet sizes. Appendix B would also benefit from details about the experiment rather than the test number. The statistical analysis is also unclear, and a graphical presentation of the results would be preferable (rather than a series of figures with (too) many significant results). Finally, regarding the plan of experiment, it seems that a preliminary study, preferably at a small scale, would have clearly demonstrated that some of the dispersants were significantly less efficient, and should then not have been included in the large scale tests. A reduced number of dispersants could then have allowed different conditions of tests, in particular a more viscous (or weathered) oil as differences between products are more interesting when the dispersibility is reduced (and the control could also be lower...).

II. RESPONSE TO CHARGE QUESTIONS	
Provide narrative responses to each of the nine charge questions below.	
1	*Are the objectives of the study clearly defined? If not, what are your recommendations for improving the description of objectives?*
Comments:	The objectives of the study are clearly defined.

2	*Was the methodology used to define the selection and testing conditions of the four dispersants clearly described?*
Comments:	The methodology used to select the dispersants is clear, but the choice of the testing conditions remains largely unclear. The characteristics of the test oil were related to other arctic crudes but based on a very limited number of parameters

	(API gravity and sulfur content), whereas, in terms of behavior at sea, asphaltenes and wax contents, distillation curve, viscosity… are as relevant. In addition, the pour point of the oils seems high in comparison with the test temperatures (and too high when considering dispersibility studies). A preliminary study should have been conducted in order to better identify the most important parameters and to get lower dispersant efficiencies. And the control experiment should have led to poor dispersibility.

3	*Were the testing procedures used appropriately describe and properly implemented?*
Comments:	The testing procedure was well explained but some parameters of the experiments were not consistent with the methodology. In particular, for the Finasol OSR 52, the Dispersant to Oil Ratio was the same for any conditions of dosage (lowest, mean, and highest). A scheme of the test protocol would also have been added, with the location of the oil slick and LISST equipment. The protocol for the oil concentrations measurements (type of measurement, calibration, uncertainty…) should have been better described, as well as the method used to assess the remaining oil at the end of each experiment.

4	*Were the results of the sampling and statistical testing adequately characterized and clearly described? Were there any critical results or limitations that were not discussed or addressed in the report?*
Comments:	Results of oil concentrations are not clearly explained, and there are no units on figures. The way the two sets of data resulting from the LISST measurements were combined is not explained. The statistical analysis is also not clear and figure (with error bars) would have been preferred to tables with too many significant numbers. The threshold value of 70 µ m is also not discussed whereas it would have been interesting to reassess this limit based on a big set of data obtained at a large scale. Finally, one important result has not been discussed: the control is very high, and the effects of dispersants are expressed as a gain in percentage compared to this condition. Moreover, the control shows a significant variability (from 43.0 to 59.7), thus suggesting that the comparison between dispersants should include this key point. A comment or an explanation on the increase of the oil concentration over time for the lower LISST (test 8) should be added.

5	*Are the findings and overall discussion of the results for each product tested clearly discussed?*
Comments:	Results and conclusions obtained for each product are generally well discussed, with the limitations mentioned above. Once again, the high level of the efficiency

	for the control, combined with a great variability, should prompt reservations to conclusions.

6	*Are the conclusions logical and appropriate based on the results? Are there any additional conclusions that could be drawn?*
Comments:	See section 4.

7	*Does this report present sufficient new data and knowledge, and are the findings useful for informing oil spill response planning in the Arctic regions?*
Comments:	The questions raised by the Arctic conditions, in terms of oil spill response, in relationship with dispersant application, is based on environmental issues, and also on efficiency of techniques. This study well address the possibility of applying dispersants, from a technical point of view. However, even if tests are performed at low temperatures, these temperatures are not totally representative of arctic conditions. In particular, only low pour point oils can be treated, while it is not the case of this study. It should have been interesting to compare dispersants at the laboratory scale at very low temperatures, to clearly identify the influence of this parameter, and then to conduct large scale experiments by varying the oil condition and not only the dispersant products.

III. SPECIFIC OBSERVATIONS

Provide specific observations or comments on the report mentioning page and paragraph (expand table if needed).

	Page	***Paragraph***	***Comment or Question***
	9	Table 1	Pour point seems high (10°C) considering the test temperatures. It is also high for a slightly paraffinic oil (2.77 % wt.). Only one significant figure should be enough for contents in %.
	10	Table 2	Test oil is actually different from the other U.S. Arctic oils based on the API gravity. Additional parameters such as wax and asphaltenes content, viscosity, distillation data would give more information when dealing with oil dispersibility.
	13	Figure 4	Information should be added on the concentrations units, and also how the data are processed (collected vs used).
	14	Table 3	DOR is identical for the 3 dosing rates of Finasol.
	15	Table 4	The column “improvement over Corexit” is not very useful (already one relative difference with control).
	15	Figure 5	Cumulative distributions are not very clear as the maximum do not appear and should be replaced by “classical” distributions.

	16	2nd	The 70 µm cutoff should have been discussed as it refers to an "old" paper (which many papers also refer to…).
	17	Table 5	Too many significant figures.
	17	Table 6	Too many significant figures. Explain the difference between *%* and *concentration.*
	18	Table 7	Too many significant figures. A figure with error bars should be preferred to a table.
	19	Table 8 & 9	Too many significant figures. Two-to-two comparisons should be replaced by a general scheme.
	23	2nd	Comparisons are performed without considering any variability of the results. In addition, the highest DOR of the Corexit (0.09) was between 2 and 3 times more than for the other products (from 0.03 to 0.05).
		Appendix B	Specify test conditions for each test, rather than only the test number (Control, Corexit Low, Corexit High…).
		Appendix C	To be placed before Appendix B.
		Appendix D	Cumulative distributions are not very clear to identify the mean size.

5.4 George A. Sorial

Peer Review of the BSEE Report *"Comparative Testing of Corexit EC9500A, Finasol OSR 52, Accell Clean DWD, and ZI 400 at Ohmsett in a Simulated Arctic Environment."*	
NAME:	George A. Sorial
AFFILIATION:	University of Cincinnati
DATE:	June 21, 2015
I. GENERAL IMPRESSIONS	
Provide overall impressions (approximately 1/2 page in length) addressing the accuracy of information presented, clarity of presentation, and soundness of conclusions.	
Comments:	The research work provided a comparative study for the effectiveness of four dispersants, namely, Corexit EC9500A, Finasol OSR 52, Accell Clean DWD, and Z1 400 on Alaskan Crude Oil at ambient temperatures varying between -4.9C to 8.7 C and water temperatures varying between -1.2 and 1.3 C. The effectiveness of the dispersants was captured by applying the dispersants on a surface slick using Ohmsett's spray bar simulating boat spraying system. Droplet size distribution of the dispersed oil and the volume of oil remained on the surface after the test compared to the original volume applied was measured. The droplet size distribution was measured at two depth, one meter and two meters. Droplet size distributions smaller than 70 microns are considered to be fully dispersed. Strength of the Report: 5. The authors blended two crude oils into one homogeneous test oil that could be used for the full study – 15 planned tests. They measured the viscosity and its API gravity and compared to actual region oil. 6. Each dispersant was tested in three separate replicates to provide detailed statistical analysis. Even the control tests (Oil without dispersant) were tested in replicates in order to determine the effect of the test conditions. 7. Detailed statistical analysis was provided. 8. The test runs could represent a real field scenario. Weakness of the report: 6. Data selection: The plots provided for the data selections are labelled as Run number without specifying the dispersants type or oil control. The concentration units should be provided in the plot not within the text. 7. It is not clear why Finasol provided similar dispersant effectiveness as Corexit while the droplet size distribution curves are significantly different. More elaborations need to be provided. 8. The droplet size distributions were measured at two depths, how was the plot in Figure 5 generated from these two depths. 9. All the tables should be provided to two decimal places.

	10. A nomenclature for all symbols used and acronyms should be provided after the table of contents.

II. RESPONSE TO CHARGE QUESTIONS	
Provide narrative responses to each of the nine charge questions below.	
1	*Are the objectives of the study clearly defined? If not, what are your recommendations for improving the description of objectives?*
Comments:	The objectives of the study were clearly defined in the last paragraph of the introduction section.

2	*Was the methodology used to define the selection and testing conditions of the four dispersants clearly described?*
Comments:	The methodology used to define the selection and testing conditions were clearly defined in methods, test oil, and testing procedures.

3	*Were the testing procedures used appropriately describe and properly implemented?*
Comments:	The testing procedures described within the report and Appendix A were described accurately and properly implemented.

4	*Were the results of the sampling and statistical testing adequately characterized and clearly described? Were there any critical results or limitations that were not discussed or addressed in the report?*
Comments:	As mentioned above in the overall view of the report, the weakness in point 1 to 3 should be addressed.

5	*Are the findings and overall discussion of the results for each product tested clearly discussed?*
Comments:	The discussions of the results are clearly described with exception to the three weakness points described above.

6	*Are the conclusions logical and appropriate based on the results? Are there any additional conclusions that could be drawn?*
Comments:	The conclusions were clearly provided for the four dispersants studied.

7	*Does this report present sufficient new data and knowledge, and are the findings useful for informing oil spill response planning in the Arctic regions?*
Comments:	Yes, the report provides new data that could be representative to actual spill scenario.

III. SPECIFIC OBSERVATIONS

Provide specific observations or comments on the report mentioning page and paragraph (expand table if needed).

	Page	***Paragraph***	***Comment or Question***
	17	Table 6	Provide values to two decimal places
	18	Table 7	Provide values to two decimal places
	19	Table 8	Provide values to two decimal places
	19	Table 9	Provide values to two decimal places
		Appendix B	Concentration units should be provided within the plot and the run number should be clearly identified with respect to dispersant used or oil control

5.5 Albert D. Venosa

Peer Review of the BSEE Report *"Comparative Testing of Corexit EC9500A, Finasol OSR 52, Accell Clean DWD, and ZI 400 at Ohmsett in a Simulated Arctic Environment."*	
NAME:	Albert D. Venosa
AFFILIATION:	Independent Consultant (retired from EPA)
DATE:	July 1, 2015
I. GENERAL IMPRESSIONS	
Provide overall impressions (approximately 1/2 page in length) addressing the accuracy of information presented, clarity of presentation, and soundness of conclusions.	
Comments:	My overall impression is mediocre. The accuracy of information, as evident in my review comments below, is subpar. The researchers seemed to be unfamiliar with the literature, or they simply did not review it and include their review comments in the report. Their methodology was lacking in regard to the allowance of sufficient time for recoalescence to occur, a major weakness. Statistical methodology was less than appropriate since no analyses of variance were used to establish significant differences. Thus, the quantitative, statistical findings and overall discussion of results were inadequately presented. Pair-wise statistical differences were inadequately discussed or explained. As for product selection, it was disappointing that the researchers did not select two products for testing at Ohmsett that had been previously tested and reported in the literature as being excellent dispersants in laboratory and wave tank studies, JD2000 and SPC1000 (they tried with SPC1000, but they seemed unable to overcome the viscosity problems they faced with the product, which they might have more successful with if they had communicated with the manufacturer more fully). Their conclusions were not logical or appropriate due to the flawed statistical analyses and the bias they had toward Corexit 9500 and Finasol (Finasol is a product that is very similar if not identical to Corexit 9500). In my opinion, the only "data" that might be useful in future decision-making for Arctic oil spills are the qualitative observations on foaming of two of the four products.
II. RESPONSE TO CHARGE QUESTIONS	
Provide narrative responses to each of the nine charge questions below.	
1	*Are the objectives of the study clearly defined? If not, what are your recommendations for improving the description of objectives?*
Comments:	The objectives were presented in one sentence on p. 3, i.e., "This study compared the performance of Corexit and three other commercially available dispersant formulations, as measured by Dispersant Efficiency and the size distribution of dispersed oil droplets in the water column." Alas, the sentence displays early in the report the positive bias towards Corexit. Rather, it should state something like "This study compared the performance of four commercially available dispersant formulations…" So, the objectives were indeed clear but imposed a preconceived bias toward one dispersant just because it has been the most tested one on the NCP Product Schedule. All biases should be eliminated before BSEE publishes this report.

2	*Was the methodology used to define the selection and testing conditions of the four dispersants clearly described?*
Comments:	The methodology of selection was simple and not fully explained. The investigators confined their selections to the two most popular dispersants on EPA's Product Schedule and two other lesser known products without regard to whether any of the selected products have ever been reported in peer-reviewed literature citations either in lab, mesocosm, or field studies. This implies the researchers are unfamiliar with the literature. It also implies that the researchers did not study the MSDS sheets of the Product Schedule products that describe in greater detail how the products differ in constituents, properties, and performance from one another. Had they done any of this, they might have confined their selections to more effective products a priori rather than a non-effective one like ZI 400 (despite the fact that the latter product is on the Product Schedule).

3	*Were the testing procedures used appropriately describe and properly implemented?*
Comments:	The testing procedures were adequately described, but unfortunately not in an appropriate fashion. What I mean by this is that recoalescence is an extremely important outcome of using any oil dispersant. Droplet sizes < 70 µm should generate droplets that are neutrally buoyant and will not likely recoalesce. However, a good scientific method should attempt to confirm this outcome by studying it in greater detail than described rather than assuming it should. Halting the wave experiments after 20 min and then collecting whatever oil was not dispersed does not allow sufficient time for recoalescence to take place, which would add to the amount of oil that was not dispersed to rise and be collected with the rest of the non-dispersed oil. Granted, three more hours were allowed to pass by before preparing for the next run, but no description was presented that any more recoalesced oil was collected at that time. I would recommend that a full 24 hours should be allowed to pass before confirming no more oil surfacing from the water column. This was another major weakness in the report.

4	*Were the results of the sampling and statistical testing adequately characterized and clearly described? Were there any critical results or limitations that were not discussed or addressed in the report?*
Comments:	The sampling and statistical testing were adequately described, but unfortunately the methods were not fully appropriate in my opinion. First of all, an analysis of variance (ANOVA) should have been performed at the conclusion of all tests. This is usually what researchers do to determine if there were any significant differences noted in the experiment. If there were none, then there would be no need to conduct any further statistical tests. Only when statistical differences are determined, then the HST test should be done to determine where the significances existed. Instead, the authors used stepwise multiple regression analysis to establish a best-fit model to keep only those variables that explain the variability in the data. They then used the HST to identify pair-wise differences in

	regard to DE, % oil droplet sizes < 70 µm, and oil concentration having droplet sizes < 70 µm. In their stepwise multiple regressions, they asked how significant salinity, interfacial tension, and viscosity were in explaining the variability. But, since the same seawater and tank were used in all experiments without emptying the tank and refilling it with fresh seawater, studying the influence of salinity is a waste of time. It turned out (Table 6) that salinity played a role for the dependent variable "% below 70 µm", which I do not understand, as did viscosity. None of those 3 variables significantly affected DE or concentration below 70 µm (unless I'm reading Table 6 incorrectly). Frankly, I don't understand the blank cells in the table. If they tested all 3 variables, then there should be probability values (p-values) in all table cells. Tables 7-9 contained the critical statistical data. In Tables 8 and 9, only 2 pairwise significances were observed. For DE (Table 8), only the pairs ZI400-Corexit and ZI400-Finasol were statistically different from each other. For the < 70 µm concentrations (Table 9), only the pairs Finasol-Control and ZI400-Finasol were statistically different from each other. Table 7 shows 7 different pairs showing significant differences from each other. These results beg to be discussed and explained in greater detail but were not.

5	*Are the findings and overall discussion of the results for each product tested clearly discussed?*
Comments:	Based on the above responses, it should be clear that the quantitative, statistical findings and overall discussion of results were inadequately discussed and/or explained that would lead to a more informed decision on which dispersant product should be used in an oil spill. Many unanswered questions remain to the reader. The dispersant product selection was inadequate since two dispersants that have been well tested in the literature were not tested in the project, namely, SPC 1000 and JD2000. Admittedly, the authors had problems with SPC 1000, but I am not convinced, based on their reporting, that they tried hard enough to overcome the viscosity and application problems of the product. As for JD2000, this product should have been chosen based on the literature, but the authors demonstrated an unawareness of the literature by not even mentioning this product in their selection procedure.

6	*Are the conclusions logical and appropriate based on the results? Are there any additional conclusions that could be drawn?*
Comments:	No, the conclusions are not logical and appropriate. As I stated early in this review, the statistical analyses were flawed to begin with, but even if one accepts the analyses, the authors clearly displayed positive biases toward Corexit and Finasol. Although, based on their qualitative observations about the lack of foaming by Corexit and Finasol and the presence of foaming by ZI 400 and Accell that may have negatively impacted efficient and effective applications of the latter products to the slick, such observations are indeed helpful in allowing good decisions to be made by decision-makers against the use of the latter two products in the Arctic and in favor of the former two. However, the lack of

	statistical significances between Finasol, Corexit, and Accell do not help in such decision making.

7	*Does this report present sufficient new data and knowledge, and are the findings useful for informing oil spill response planning in the Arctic regions?*
Comments:	I partially answered this question above in Question 6. I think the only "data" that might be useful in future decision-making for Arctic oil spills are the qualitative observations on foaming of two of the four products.

III. SPECIFIC OBSERVATIONS

Provide specific observations or comments on the report mentioning page and paragraph (expand table if needed).

	Page	***Paragraph***	***Comment or Question***
	1	2	Another dispersant product, SPC 1000, was pre-tested but found to be too viscous to apply via the pressure spray apparatus.
	1	3	Second sentence and onwards: this statement, plus the ones in the subsequent sentences, needs to be couched in terms of statistical significance. Otherwise, it points out a possible positive bias in favor of Corexit over the other products tested. Such bias should be avoided.
	3	1	Line 4: Minor point. It seems incongruous to discuss ice in terms of "concentration." Perhaps better terms might be "density or area of consolidated pack ice" or "ice cover formed by the packing and freezing together of pack ice floes."
	3	1	2nd last line: Need references here to support the claim that 30 years of dispersion effectiveness data in Arctic conditions are available.
	3	2	Lines 2 and 3: Is there a supportive reference for this statement? JD2000 has also generated interest for publications that have appeared in the peer-reviewed literature.
	4	1	Lines 3 and 4: Dispersants may also contain solvents that tend to decrease viscosity for ease of application.
	4	1	Line 8: Regarding droplets of size < 70 µm can be consumed by microorganisms: this statement implies that droplets > 70 µm cannot be consumed, which is not true. The bigger the droplet size, the lower is the surface area that can be biodegraded, so rates are lower. So, add "more rapidly" after the word "consumed."

	4	1	Line 10: the Conover reference is extremely old (1971). Much mesocosm research has been done in early to mid-2000's, mostly at the wave tank at DFO-Canada and at Ohmsett. The authors seem not to have reviewed the literature.
	5	2	First sentence, actually five dispersants were selected. One (SPC 1000) was not tested due to problems with viscosity and throughput in the spray apparatus. In the absence of SPC 1000, they should have tested JD2000 because of previous published research on that product.
	5	2	About halfway down: the term “concentration” should be changed to volume concentration, ppmv.
	6	7	In the numbered bullet #7, the Baffled Flask Test is replacing the inferior Swirling Flask Test by EPA, so that regulation will soon be replaced.
	6	11	Venders must also submit to EPA toxicity data from testing mysid shrimp (*Americamysis bahia*) and silversides (*Menidia beryllina*).
	6	Last	The product ZI 400 should never have been selected for testing. It is a very poor dispersant as shown in published lab data.
	7	1	Line 3: Text should read SPC 1000 was replaced by Accell.
	7	1	Last line of 1st paragraph: Why didn't the researchers contact the manufacturer and ask for another sample? Other readers will undoubtedly ask this same question.
	7	2	Last 3 lines before bullets: dispersant concentrations < 400 ppb do not affect the outcome of tests, as referenced by SL Ross in 2000. The reference was not from a peer reviewed paper and had no statistical data to support it.
	8	1	The oil was not weathered. It was a blend of 2 unweathered oils. The freshness should be pointed out.
	8	1	Line 7: Ohmsett developed, not measured, a temperature-viscosity curve…
	10	1	Line 1: the term “replicates” is misleading. Replicate samples or replicate experiments? Big difference. Later it says the experiments were repeated, so this word should be "replicate experiments conducted in random order." Was the order randomized? It should have been.
	10	2	Tests continued for 20 min, after which they were terminated. The tank was allowed to settle for an

			additional 3 hours before prepping it for another test. The 20 min is not nearly long enough to allow recoalescence to occur. Nor was 3 hours. They should have waited 24 hours (or at least overnight) before prepping for the next run. This is very important, and it affects conclusions on total dispersion effectiveness.
	11	1	Last line: did they warm the dispersant product SPC 1000 like they did the others? No statement appears in the text regarding this point.
	11	2	Last 3 lines: Was the assumption tested and confirmed that any resurfaced oil would stay dispersed? This is very important and could have a bearing on results. The resurfaced oil should have been measured and included in the subsequent calculations.
	11	3	Lines 1 and 2 before the equation: By their own definition, surface oil that no longer remains on the surface should be included in the Vor after sufficient time is allowed for re-coalescence and surfacing. No mention of waiting a defined amount of time was made.
	11	3	In the equation, the last term should be (1 – fraction of water in the source oil), not (1 – % water in source oil). Same for both equations.
	12	Last	As mentioned above, the 20 min time period for the experiment was too short for recoalescence to have occurred. This is a major flaw in the design.
	13	1	Line 3: the concentration should be reported in units of ppmv, as it also should on the y-axis of Figure 4. Do not let the reader make assumptions.
	13	Fig. 4	What does “data used” mean? Used for what? Why did they not include the grayed out data? Also, can time also be superimposed on the x-axis? Sample number means nothing to the reader.
	13	3	First paragraph under Statistical Analysis: (1) did salinity vary that much to be a confounding variable, especially since the same water was used for all experiments. Also, why did they not use ANOVA rather than stepwise regression analysis?
	14	1	Line 1: Should be ppb, not ppm, as specified on p. 7
	14	2	Last paragraph before Results section: The Tukey HSD test is a post-hoc test, meaning that it is performed after an analysis of variance (ANOVA) test. This means that to maintain integrity, a statistician should not perform Tukey's HSD test

			unless he has first performed an ANOVA. It's not clear that this was done. The ANOVA determines if there are significant differences, but the HSD tells you where those differences occurred. It appears the authors used the HST to determine if there were differences between any 2 treatments independent of an overall ANOVA, which the HST was not designed to do, at least initially.
	14	Table 3	The 49.8 mean is really high for the control! This points out the disadvantage of using a batch wave tank with no flow-through because the dynamic dispersion effectivenes (DDE) cannot be calculated. See Li et al., 2009, "*Evaluating crude oil chemical dispersion efficacy in a flow-through wave tank under regular non-breaking wave and breaking wave conditions*." Marine Pollution Bulletin, 58(5): 735-744.
	14	Last	Corexit had the lowest DOR, which biased the test since more dispersant was used at the lower DOR. The 0.05 value is 67% higher than 0.03 for Finasol and Accell and 25% higher than for ZI 400. Not a trivial difference.
	15	1	In the statement "Finasol performed almost identically to Corexit…," the authors should add "with significantly less dispersant".
	15	1	Line 2: need to report the p-value after DE = 72.2%.
	15	1	Line 5: Regarding the Test 10 outlier, was this outlier statistically determined as an outlier using a test like the Grubb's test? Cannot eliminate the outlier without doing a statistical significance test. Or better, do a repeat of this test.
	15	1	Line 9: The poor performance of ZI 400 was not unexpected since lab studies showed this dispersant does not work!
	16	2	Line 2: Are these bar graphs composites of the 3 experiments? They should be.
	16	2	Line 4: According to the manufacturer, Finasol is identical to C9500.
	16	2	Line 7: Is this conclusion, that Accell was not as effective as Corexit or Finasol, based on statistical analysis or eyeballing? I see no p-value in support of this statement.
	16	2	Line 9: In the line "The oil that was dispersed (by Accell) consisted of a larger percentage of very small droplets over the dispersed oil from Finasol…," again, was this statistically based or

			eyeballed? If statistically based, and explanation is needed.
	16	2	Line 17: Data mentioned are not shown on Fig. 6. Also, what about Corexit? No mention made of droplet sizes < 60 um.
	17	Top	Fig. 7, why not show the LISST data for all 3 replicate experiments? Would it result in a too busy chart?
	17	1	The median droplet size in the control test was 457.26 µm. This and subsequent reporting of droplet sizes should have significant figures no greater than 3 at most. It's silly to use 5 or more significant figures.
	17	1	In line 7, need p-values here to justify the word "significant."
	17	Table 6	Second row, last column, I don't understand this significance. Salinity was virtually constant (same tank, same water) across all runs. Makes little sense. Also, what is the meaning of a blank cell in the table? Very strange table.
	18	2	As stated previously, the authors should have first run an ANOVA to determine if there were significant differences to begin with. THEN, they should run the HST to determine where the significances were. Also, the table has WAY too many significant figures in the data. Confine them to max 3, perhaps even 2.
	19	1	Table 8: This table says there were no significant differences anywhere except between ZI400 and Corexit and between ZI400 and Finasol. This suggests low statistical power, and more replicate experiments should be done to improve the power of the test. Otherwise, everything in the report is reduced to qualitative statements.
	19	2	Table 9: This table is also revealing (in a negative sense). No differences were observed in the < 70 um densities except for Finasol-Control and ZI400-Finasol. This must be explained.
	20	1	Line 6: Was SPC 1000 heated as the other products? I presume so, but it was never so stated anywhere in the report.
	20	1	Bottom fourth of the paragraph, regarding the phone call to SL Ross, by whom? No need to ID the caller. Just mention if it was the manufacturer.
	20	1	End of the paragraph, why didn't the researchers just ask the manufacturer to give them another

			sample of SPC1000? And, did they discuss these problems with the manufacturer to determine if temperature was an issue or if foaming was an issue at low temperature but not at warmer temperatures?
	21	1	Line 1, the sentence implies positive bias toward Corexit. It should be deleted.
	21	1	The last sentence of the paragraph is key. Recoalescence should have been quantified. What would have resurfaced if they had waited 24 hours after termination?
	21	3	Line 3: Droplets remaining close to the surface implies recoalescence occurred. Did this appear to increase over time? Also, it indicates bigger droplets. However, according to Tables 7-9, this observation is moot since there weren't any statistical differences between Accell and the other products except ZI400.
	21	3	Line 7: Regarding the froth increasing after each test, this suggests persistent residuals between treatments.
	21	3	Line 9: But the observation was made ONLY with Accell. Why do they say it cannot be definitely associated with Accell?
	22	1	Line 3: Performance of ZI400 was poor anyway, so why do they conclude that the foam did not affect the liquid product? How do they know it didn't affect performance?
	22	1	Last sentence: nozzles were clogged and the system had to be thawed. This was despite the warming of the product before testing?
	22	1	In the report, there was absolutely no discussion about how the Ohmsett results compared with previously published tests in the literature both in the lab and in mesocosm studies. This type of scholarly discussion is mandated by peer-reviewed journals. Although this is not a journal article, nonetheless discussion of the data relative to what is already known puts everything in perspective and suggests how further research should be conducted (or how even this research should have been conducted).
	23	2	Line 1: This statement is merely subjective and not quantitative because of the lack of statistical significance as demonstrated in the tables.
	23	2	The parenthetical should be eliminated and Finasol should also be part of the subject of the sentence.

	23	1	Line 4: After “46% improvement”, they should also state “but not statistically significant.”
	23	1	Line 5: After the word “Corexit”, the words “and Finasol” should be added.
	23	1	Last line: The temperature range was miniscule, so this statement applies to all products. Again, bias!
	23	2	Last line of the page: Please include the p-values, which contradict the significance of these statements.
	24	2	Line 2: This cannot be assumed without an outlier test. Easy to do with the Grubbs' test.
	24	2	Lines 5 and 6: Here is but one example where comparison with other reported values in the literature may have shed some light on the subject.
	24	3	Line 3: The p-value was not significant.
	25	1	Regarding the calling for large scale testing, why large scale? Lab tests should suffice since everything would be controlled by the analyst. The wave tank at DFO-Canada can be used more quantitatively and with flow-through so that dynamic dispersion effectiveness (DDE) can be calculated. The Ohmsett tank is too large to allow flow-through, which argues against using it for statistical comparisons of chemically based products.
	25	1	End of paragraph: Lab testing with the soon to be enacted Baffled Flask Test would be much easier and less costly. In fact, EPA will be requiring testing at 2 temperatures (5 and 25 C) and 2 oils once the test has been adopted.
	25	2	Last sentence: Please explain this last sentence. It implies residuals may have affected performance between runs.

6. APPENDIX B: PEER REVIEW MATERIALS PACKAGES

The peer review materials packages are attached separately.