GOLDER

# TECHNICAL MEMORANDUM

**DATE** 7 October 2019

**Reference No.** 1663724-117-TM-Rev1

**TO** Lou Kamermans, Director Corporate Sustainability
Baffinland Iron Mines Corporation

**FROM** Sima Usvyatsov

**EMAIL** Sima_Usvyatsov@golder.com

## POWER ANALYSIS FOR BAFFINLAND'S MARINE ENVIRONMENTAL EFFECTS MMONITORING PROGRAM (MEEMP)

This document contains supplemental information related to power analysis results for the Marine Environmental Effects Monitoring Program (MEEMP) in response to Technical Comments provided by Fisheries and Oceans Canada (DFO) and commitments made by Baffinland during the last two Technical Review Meetings held in Iqaluit, Nunavut (April 08-10 and June 17-19), as outlined in Table 1.

**Table 1: DFO Technical Comments - Power Analysis for Marine Environmental Effects Monitoring Program (MEEMP)**

| ID# | Technical Comment/ Recommendation | Commitment |
|---|---|---|
| DFO 3.9.2 | DFO recommends Baffinland account for all species of fish and their habitat, including marine mammals, potentially impacted by the project as part of their offsetting accounting. | Baffinland commits to accounting for all species of fish, marine mammal and their food/ habitat potentially impacted by the project.<br>Baffinland will provide the results of a power analysis for its marine environmental effects monitoring sampling program, including for fish and fish habitat by May 13. DFO requests that Baffinland commit to provide the results of power analyses for all non-significant statistical tests in future monitoring reports; power analyses are recognized as being a critical component of ongoing adaptive management to ensure sample sizes are appropriate for the level variation observed. |

| | | |
|---|---|---|
| TM2-DFO-NEW7 | DFO recommends that power analysis of benthic invertebrate sampling (left out of power analysis memo document) be conducted and the memo revised to include it.<br><br>DFO recommends a revision of the MEEMP to include increased sample sizes and adaptation of sampling design to improve power and abilities to detect project effects.<br><br>DFO recommends that sampling and analyses be designed to control for confounding natural sources of variability (particularly depth and habitat for benthic species). DFO notes the use of a reference site approach (such as BACI, Before-After-Control-Impact approach) will help to control for temporal effects of interannual variability that may confound abilities to detect project impacts. | Baffinland will update the previous power analysis memo to respond to the requests provided by DFO.<br><br>Based on results of the power analysis, Baffinland has revised its MEEMP program design to include additional sampling stations and one additional transect to improve power and abilities to detect Project effects.<br><br>The design of the MEEMP benthic infaunal sampling is based on a radial gradient (RG) where the same replicates (stations) located along a distance gradient are re-sampled at specific time intervals (years). This design is recommended by Environment Canada (2012) and advocated by Ellis and Schneider (1997) as an alternative to the BACI design. Environment Canada recommends the RG design, in which a broader geographic area is sampled, as particularly useful in non-homogenous environments, such as in a marine environment which often has complex current and circulation patterns or a variety of equally important habitat classes or gradients. The gradient design enables physical, chemical, and biological changes to be assessed as a function of distance from a point source. This design is very effective at elucidating the spatial scale of impacts and therefore can provide considerable insights into potential mitigations and/or alterations to Project activities to address any observed negative environmental effects. Radial gradient designs are effective at addressing threshold of effects as a function of distance and/or quantification of effect (e.g., contaminant level). Influence of other confounding factors, such as depth, can also be addressed using the sampling design. In general, the RG design doesn't necessitate use of a reference site because transects cover a large spatial extent and sampling stations located at far distances from the source point ultimately serve as reference sites. For instance, one of the transects — the Coastal Transect (CT) — extends north along the eastern shore of Milne Inlet, outside of the predicted ZOI of project activities. The Coastal Transect overlaps with Reference Site 1, established in 2013, thereby maximizing the use of existing baseline data. Therefore, the Coastal Transect serves as a reference transect. From a modeling perspective, the inclusion of a standalone reference area in a gradient analysis is not meaningful, since the covariate of distance from point source (i.e., ore dock) for the reference site would be very large, thereby creating a high-leverage point and obscuring the trends of interest near the point source. |

# 1.0 INTRODUCTION

## 1.1 Background

Baffinland Iron Mines Corporation (Baffinland) is the owner and operator of the Mary River Project (the Project), an operational open-pit iron ore mine located on North Baffin Island in the Qikiqtani Region of Nunavut. Project Certificate No. 005, amended by the Nunavut Impact Review Board (NIRB) on 27 May 2014, authorized Baffinland to mine up to 22.2 million tonnes per annum (Mtpa) of iron ore from Deposit No. 1. Of this 22.2 Mtpa, Baffinland is currently authorized to transport 18 Mtpa of ore by rail to Steensby Port for year-round shipping through the Southern Shipping Route (via Foxe Basin and Hudson Strait), and 4.2 Mtpa of ore by truck to Milne Port for open water shipping through the Northern Shipping Route using chartered ore carrier vessels (the Approved Project). A production increase to ship 6.0 Mtpa from Milne Port was approved for 2018 and 2019.

In accordance with existing Terms and Conditions of the Project Certificate, Baffinland is responsible for the establishment and implementation of environmental monitoring studies conducted over a sufficient time to meet the following objectives:

- measure the relevant effects of the Project on the marine environment
- improve understanding of local environmental processes
- confirm that the Project is being carried out within the pre-determined terms and conditions relating to the protection of the marine environment
- assess the accuracy of the predictions contained in the Final Environmental Impact Statement (FEIS) for the Project

Since the commencement of Project activities under the ERP in 2014, Baffinland has conducted an ongoing annual Marine Environmental Effects Monitoring Program (MEEMP) in Milne Inlet. The MEEMP was developed and implemented in consideration of potential Project-related adverse impacts to the marine environment as identified in the Final Environmental Impact Statement (FEIS) (Baffinland 2012) and Early Revenue Phase (ERP) Addendum (Baffinland 2013) and to meet monitoring requirements outlined in Terms and Conditions of Project Certificate No. 005.

The MEEMP includes monitoring of marine water and sediment quality, marine invertebrates, marine vegetation, and fish and fish habitat. The sampling design is based on the Metal Mining Environmental Effects Monitoring (EEM) guidelines (Environment Canada 2012) and includes statistical approaches to detecting potential Project-induced impacts on the marine environment.

## 1.2 Objectives

The objective of this memorandum is to provide results of a power analysis conducted in accordance with Baffinland’s commitment following DFO Technical Comment 3.9.2 and TM2-DFO-NEW7. The power analysis was conducted using marine biophysical data collected as part of the MEEMP including marine sediment (e.g., percent fines, sediment iron concentrations), benthic infauna and marine fish (e.g., weight-to-length relationships).

## 2.0 METHODS

The statistical power to detect annual differences in sediment and biological data was calculated. Standard power analysis software and methods are limited in the complexity of models to which they can be applied, and were not appropriate for most of the analyses conducted. The power to detect statistically significant effects was estimated using simulation methods as described in Seavy et al. (2005). The general approach was to simulate data based on the fitted model, the effect sizes (EFs) of interest, and the variability in the observed data, and re-run the models that were used for analysis using the simulated data. The data simulation and analysis were repeated at least 1,000 times, and the proportion of repetitions where yearly differences were statistically significant was interpreted as the power. Significance was assessed at the 0.05 level for all analyses. Results are presented in terms of the effect size required to achieve a power of 0.8, which is a common goal for power analyses. Details about how the power analyses were implemented for each collected dataset are described within each analysis section. All power analyses were performed in the statistical package R v.3.5.3 (R 2019) using the package emmeans (Length 2019).

### 2.1 Sediment – Percent Fines

Power to detect statistically significant differences in percent fines between 2018 and previous years of data collection was estimated using simulations. The power analysis used data that were simulated as follows:

1) Data for each year from 2014 to 2017 were simulated based on the fitted model (log10-transformed percent fines as a multiplicative function of transect, distance from transect origin, and year) and the variability of 2014 to 2017 data (calculated as standard deviation [SD] of residuals within each year and transect).

2) The 2018 data were simulated by applying various effect sizes to the 2018 percent fines values and using the variability (as SD) from the 2018 data at each transect. The effect sizes were calculated as multiples of the SD values within each transect, ranging from -4 (i.e., a reduction in percent fines equal to three times the SD of residuals) to +4 (i.e., an increase in percent fines equal to three times the SD of residuals). Since the effect of interest is the influence of the ore dock, the effect size was applied linearly, so that the full effect size was applied at the origin of each transect, and no effect size was applied at the farthest station of the transect.

The simulated dataset was then analyzed using the same model as in the original analysis (Golder 2019), which was an ANCOVA that included main effects of distance from transect origin, year (as a categorical variable), transect, and all possible interactions between the three variables. The effect of distance was modeled as a second-degree orthogonal polynomial to account for the non-linearity in percent fines relative to distance from transect origin. Following the ANCOVA, multiple comparisons between years were performed at the following covariate values: distances of 0 m (except for the Coastal Transect), 500 m, 1,000 m, and 1,500 m for all transects, in addition to 4,000 m for the Coastal Transect. At each distance / transect combination, the simulated 2018 percent fines were compared to each of 2014 to 2017 percent fines to identify significant differences between years. An approximation to Dunnett's adjustment of P-values was used to correct for error rate. An adjusted P-value less than 0.05 was interpreted as a significant difference between years. This simulation was repeated 5,000 times for each effect size and the proportion of repetitions with P-values less than 0.05 was interpreted as the power to detect annual difference in percent fines.

The comparison of simulated 2018 data to simulated data for each year between 2013 and 2017 can be considered the estimated statistical power of the original ANCOVA in Golder (2019). This test corresponds to the 0% effect size in plots of power versus effect size (Figure 1), because there was no additional effect size added to the observed differences between years. Effect sizes from 1 SD to 4 SD on these plots refer to effect sizes in addition to the observed differences in the response variable between years.

## 2.2 Sediment – Iron Content

Power to detect statistically significant differences in the content of iron between 2018 and previous years of data collection was estimated using simulations, following the same process as described above for percent fines. However, the ANCOVA was expanded to account for the relationship between iron content and percent fines, as described in Golder (2019). The model therefore described iron content using main effects of distance from transect origin, year (as a categorical variable), transect, and all possible interactions between the three variables, as well as a main effect of percent fines.

The power analysis used data that were simulated as described for percent fines but using the iron content as the response variable. The value of the percent fines covariate in the simulated data was set to the minimum value of fines content at each transect across years and distances. The simulated dataset was analyzed using the same model as in the original analysis (Golder 2019), which was an ANCOVA that included main effects of distance from transect origin, year (as a categorical variable), transect, and all possible interactions between the three variables, as well as a main effect of fines. The estimation of power was performed as described above for percent fines, however due to the increased computation time relative to the percent fines analysis, the iron content simulations were only performed 1,000 times.

## 2.3 Benthic Infauna

The design of the MEEMP benthic infauna sampling is based on a radial gradient (RG) where the same replicates (stations) located along a distance gradient are re-sampled at specific time intervals (years). This design is recommended by Environment Canada (2012) and advocated by Ellis and Schneider (1997) as an alternative to the BACI design. Environment Canada recommends the RG design, in which a broader geographic area is sampled, as particularly useful in non-homogenous environments, such as in a marine environment which often has complex current and circulation patterns or a variety of equally important habitat classes or gradients.

The regression analysis used in the RG design examines biological responses to a potential project-related effect over a spatial gradient (distance) from the source over time and helps to determine a spatial extent of the effect. The RG helps to determine not only the extent of the effect, but also the direction of the effect. The regression approach used in gradient designs is an alternative to the BACI analyses of variance (ANOVA) design and is considered more sensitive to change and therefore more powerful than the simple comparison between control and impact locations before and after impact.

The gradient design enables physical, chemical, and biological changes to be assessed as a function of distance from a point source. This design is very effective at elucidating the spatial scale of impacts and therefore can provide considerable insights into potential mitigations and/or alterations to Project activities to address any observed negative environmental effects. Radial gradient designs are effective at quantification of effects (e.g., contaminant level) as a function of distance. Influence of other confounding factors, such as depth, can also be addressed using the sampling design.

In general, the RG design doesn't necessitate use of a reference site because transects cover a large spatial extent and sampling stations located at far distances from the source point ultimately serve as reference sites. For instance, one of the transects — the Coastal Transect (CT) — extends north along the eastern shore of Milne Inlet, outside of the predicted ZOI of project activities. The Coastal Transect overlaps with Reference Site 1, established in 2013, thereby maximizing the use of existing baseline data. From a modeling perspective, the inclusion of a standalone reference area in a gradient analysis is not meaningful, since the covariate of distance from point source (i.e., ore dock) for the reference site would be very large, thereby creating a high-leverage point and obscuring the trends of interest near the point source.

A power analysis of the benthic invertebrate monitoring program was undertaken using benthic infauna data collected on the Eastern, Western, and Northern transects surveyed in 2018 (5 stations per transect, 3 samples per station). The power analysis was performed as a simulation. In this framework, a certain scenario is assumed (e.g., effect size, number of samples, etc.). The objective was to estimate power under scenarios of different effect sizes and number of stations along the transect. A defined number of samples was drawn from a normal distribution that was based on the model and scenario, and these formed the simulated dataset. The model was rerun, and the *P* value of the effect of interest was retained. The process was repeated 1000 times per scenario. The proportion of simulations with a significant *P* value was then recorded as the statistical power. For benthic invertebrates (species richness and density), an ANCOVA model was used, with a continuous covariate of distance. Several effect sizes were applied, from a reduction of 3 standard deviations (SD) to an increase of 3 SDs, in steps of 1 SD. The SDs are those of the residuals of the original ANCOVA. It was assumed that the effect was linear with distance, with full effect at the ore dock and zero effect at the end of the transect. Sample sizes of 3 to 20 were simulated for each station.

## 2.4 Fish Length-weight

Power to detect statistically significant year effects in the length-weight relationship was estimated using simulation methods for various effect sizes and sample sizes. The power analysis used simulated data representing two years: 1) simulated 2017 data based on model estimates and variability from 2017; and 2) a hypothetical year calculated by applying various effect sizes to the 2017 length-weight slope and the same variability as 2017. For both simulated years, predicted values of weight were calculated using the linear regression equation with the 2017 intercept estimate, a slope estimate (2017 estimate or slope based on a particular effect size), and body length values randomly sampled from the 2017 data. For each observation in the simulated data set, the weight value was drawn from a normal distribution where the mean was the predicted value calculated using the regression equation for that observation, and the standard deviation was standard deviation of the 2017 model residuals. The body length was the same randomly sampled value associated with the predicted weight value for that observation. Linear regressions were re-run on the simulated data. An F-test was used to compare the multiplicative model (i.e., including year×ln(length) interaction) to an additive model with year and ln(length) as predictors, where a P-value less than 0.05 was interpreted as significantly different slopes between years. This simulation was repeated 1,000 times for each species and each effect size and the proportion of repetitions with P-values less than 0.05 was interpreted as the power to detect annual difference in the length-weight relationship.

# 3.0 RESULTS

## 3.1 Sediment – Percent Fines

Most of the observed effect sizes (as based on multiple comparisons between 2018 to each preceding year in each transect and distance) were small, with EFs larger than 2 SDs only observed in seven cases: two comparisons along the West Transect (distance 0 m, comparison with 2016 and distance of 1500 m, comparison with 2017), three comparisons along the East Transect (distance 0 m, comparison with 2015, 2016, and 2017), one comparison along the Coastal Transect (distance of 1500 m, comparison with 2014), and one comparison along the North Transect (distance 0 m, comparison with 2014; Table 2). Of these seven cases, the estimated statistical power was greater than 0.8 in three cases. Estimated statistical power was less than 0.8 for all other yearly comparisons. The estimated statistical power was zero or close to zero for all yearly comparisons at the Coastal and Northern transects and greater at the Eastern and Western transects (0.06 to 0.91).

Various effect sizes were added to the observed effect size to investigate how large of an effect size would be required to detect a significant difference between simulated 2018 data and the simulated 2014 to 2017 data. The power estimates depended on transect, year of comparison, distance from transect origin, and the simulated effect size (Figure 1). At the observed effect size (denoted as added EF of zero in the figure), the test had sufficient power (i.e., ≥0.8) to detect differences between 2018 and 2016 (distance of 0 m, West Transect), 2018 and 2015, and 2018 and 2017 (distance of 0 m, East Transect). The interannual differences between 2018 and 2014 to 2017 varied by distance, transect, and year. For example, percent fines data collected along the East Transect were lower in 2018 than in 2014, 2015, and 2018 at the transect origin, but higher than 2014 to 2017 data at a distance of 1,000 m (Golder 2019). Therefore, the change in effect size required to identify a significant difference between years differed by transect, distance, and year. In the same example of East Transect, the 2018 percent fines would have needed to be lower by approximately 1 SD to identify a significant difference between 2018 and 2014 at the transect origin.

The estimated power for the Coastal Transect was 0 at the observed effect size and remained very low (0 to 0.4) at nearly all simulated effect sizes up to 4 SD, with the exception of a few of the distances in 2014 and 2018 (Figure 1). The estimated power for the Northern Transect was 0 at the observed effect size, but power of 0.8 was achieved at effect sizes of approximately 3 to 4 SD at distances of 0 and 500 m. At the East and West transects, a power of 0.8 was generally achieved at effect sizes of 1 to 4 SD for the 0, 500, and 1000 m distances, but power was very low (<0.4) at all effect sizes at 1500 and 4000 m.

**Table 2: Estimated Power under Observed Effect Size for Sediment Per Cent Fines**

| Transect | Distance (m) | Year (Compared to 2018) | | | |
|---|---|---|---|---|---|
| | | 2014 | 2015 | 2016 | 2017 |
| West Transect | 0 | 0.51 (-1.03) | 0.54 (-1.36) | **0.91 (-2.25)** | 0.13 (-0.06) |
| West Transect | 500 | 0.22 (-0.30) | 0.10 (-0.04) | 0.14 (0.17) | 0.26 (-0.39) |
| West Transect | 1000 | 0.2 (0.32) | 0.22 (0.35) | 0.59 (0.79) | 0.21 (0.41) |
| West Transect | 1500 | 0.61 (0.82) | 0.12 (-0.17) | 0.23 (-0.39) | 0.28 (2.35) |
| East Transect | 0 | 0.4 (-1.42) | **0.99 (-3.92)** | 0.77 (2.69) | **0.96 (-3.50)** |
| East Transect | 500 | 0.39 (0.78) | 0.07 (-0.33) | 0.25 (0.54) | 0.1 (0.19) |
| East Transect | 1000 | 0.66 (1.39) | 0.26 (0.86) | 0.23 (-0.6) | 0.61 (1.44) |
| East Transect | 1500 | 0.09 (0.38) | 0.06 (-0.36) | 0.25 (-0.71) | 0.1 (0.23) |
| Coastal Transect | 0 | 0 (-0.52) | 0 (-0.93) | 0 (-1.37) | 0 (0.70) |
| Coastal Transect | 500 | 0 (0.62) | 0 (-0.37) | 0 (-0.86) | 0 (0.20) |
| Coastal Transect | 1000 | 0 (1.48) | 0 (0.06) | 0 (-0.46) | 0 (-0.13) |
| Coastal Transect | 1500 | 0 (2.06) | 0 (0.37) | 0 (-0.17) | 0 (-0.29) |
| Coastal Transect | 4000 | 0 (0.71) | 0 (0.14) | 0 (-0.20) | 0 (1.50) |
| North Transect | 0 | 0.02 (2.27) | 0 (-0.83) | 0 (0.29) | 0 (1.32) |
| North Transect | 500 | 0 (-0.05) | 0 (-0.06) | 0 (0.44) | 0 (0.42) |
| North Transect | 1000 | 0 (-0.96) | 0 (0.38) | 0 (0.58) | 0 (0.06) |
| North Transect | 1500 | 0 (-0.49) | 0 (0.48) | 0 (0.70) | 0 (0.23) |

Note: Values are the power of the multiple comparisons between 2018 and each of 2014 to 2017 percent fines at each transect/distance combination. Values in parentheses are effect sizes, as number of standard deviations. Power of 0 indicates that power was less than 0.01. Values of estimated power greater than 0.8 and their corresponding effect sizes are shown in bold text.

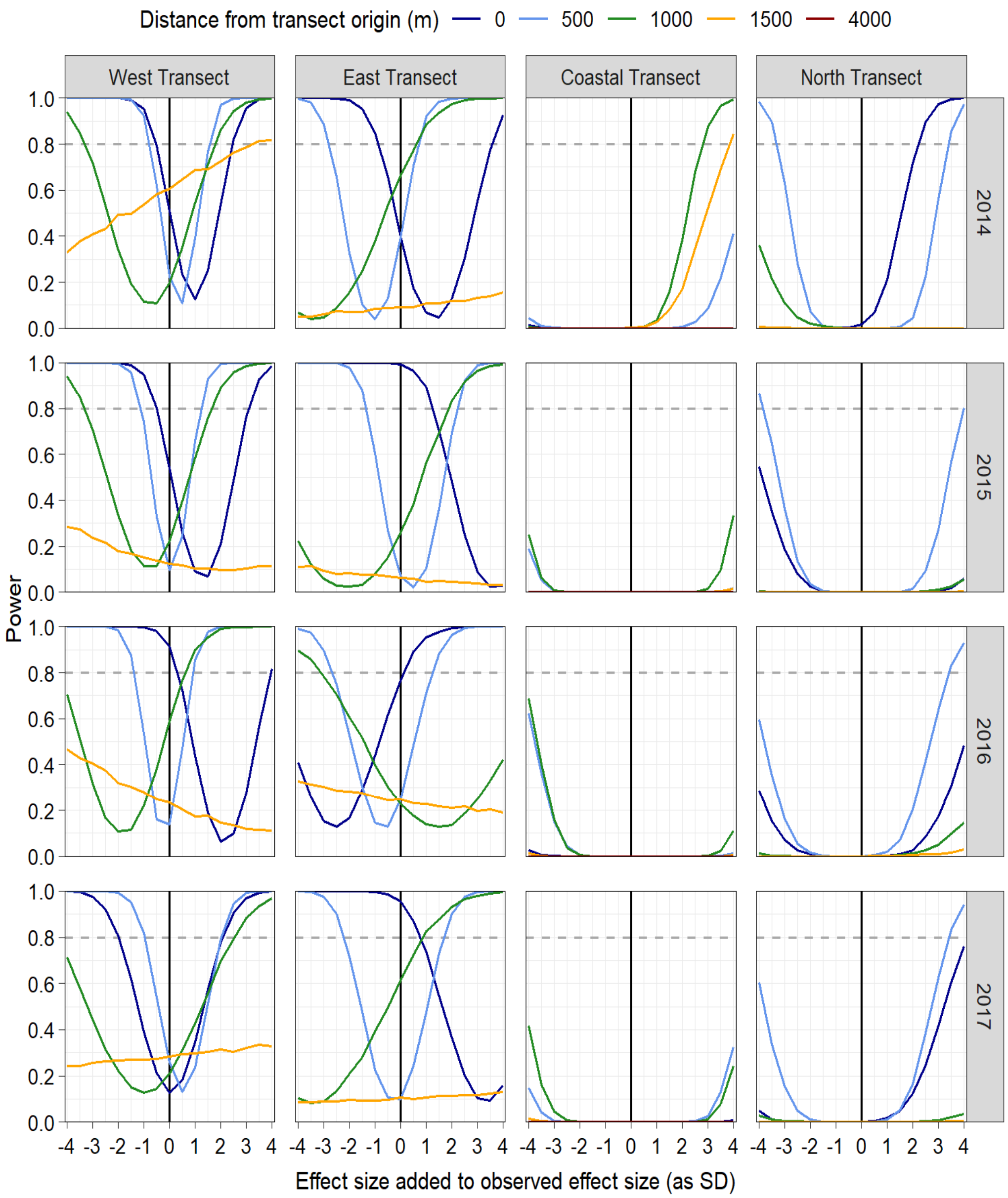



**Figure 1: Estimated Power to Detect Significant Differences in Percent Fines between 2018 and Previous Sampling Years (2013-2017) for Varying Effect Sizes and Distances from Transect Origin.**

**Note: Power of 0.8 (Standard Goal) is Denoted as Dashed Line. Vertical Solid Line Denotes Effect Size of Zero (i.e., Observed Effect Size Only).**

## 3.2 Sediment – Iron Content

In comparisons of iron content between 2018 and other years at particular transects and distances at the observed effect sizes, the estimated statistical power was greater than 0.8 at five of the year-distance combinations at the West transect and seven of the year-distance combinations at the East transect (Table 2).

The observed effect sizes for these comparisons with >0.8 power at the East and West transects ranged from 1.13 to 14.92 SD. At the Coastal and Northern transects, comparisons of iron content between 2018 and other years had statistical power less than 0.8 for all years and distances. The observed effect sizes at the Coastal and Northern transects ranged from -9.72 to 10.25 SD, although the majority of effect sizes were less than 3 SD different than the 2018 values.

Overall, the results suggest reasonable ability to detect significant differences in iron content at the West and East transects, at effect sizes as low as approximately 1 SD. On the other hand, statistical power was very low at the Coastal and Northern transects, even when effect sizes were greater than 2 SD.

**Table 3: Estimated Power under Observed Effect Size for Sediment Iron Content**

| Transect | Distance (m) | Year (Compared to 2018) | | | |
|---|---|---|---|---|---|
| | | 2014 | 2015 | 2016 | 2017 |
| West Transect | 50 | 0.15 (1.63) | 0.23 (5.17) | 0.06 (0.82) | 0.31 (3.11) |
| West Transect | 500 | 0.74 (1.37) | **1.00 (2.95)** | **0.95 (2.14)** | 0.29 (0.83) |
| West Transect | 1000 | **0.88 (1.32)** | **1.00 (2.63)** | **0.99 (1.94)** | 0.03 (0.21) |
| East Transect | 50 | **0.98 (5.24)** | 0.13 (1.85) | **1.00 (14.92)** | 0.32 (2.10) |
| East Transect | 500 | **0.82 (1.30)** | **1.00 (2.24)** | **0.95 (1.75)** | 0.26 (-0.55) |
| East Transect | 1000 | **0.87 (1.13)** | **1.00 (1.86)** | 0.74 (1.07) | 0.12 (0.11) |
| Coastal Transect | 50 | 0.02 (10.25) | 0.01 (9.25) | 0.00 (9.72) | 0.04 (-9.17) |
| Coastal Transect | 500 | 0.64 (2.80) | 0.52 (2.76) | 0.2 (2.00) | 0.05 (0.99) |
| Coastal Transect | 1000 | 0.28 (2.00) | 0.04 (1.50) | 0.03 (1.33) | 0.17 (1.75) |
| Coastal Transect | 4000 | 0.13 (2.41) | 0.00 (-0.04) | 0.11 (2.35) | 0.01 (0.03) |
| North Transect | 50 | 0.29 (1.49) | 0.06 (2.35) | 0.02 (0.52) | 0.23 (1.94) |
| North Transect | 500 | 0.17 (0.51) | 0.23 (0.90) | 0.40 (1.10) | 0.05 (-0.36) |
| North Transect | 1000 | 0.09 (0.30) | 0.14 (0.62) | 0.12 (0.51) | 0.02 (-0.16) |

Note: Values are the power of the multiple comparisons between 2018 and each of 2014 to 2017 percent fines at each transect/distance combination. Values in parentheses are effect sizes, as number of standard deviations. Power of 0 indicates that power was less than 0.01. Values of estimated power greater than 0.8 and their corresponding effect sizes are shown in bold text.

Various effect sizes were added to the observed effect size, to investigate how large of a difference in iron content would be required to detect a significant difference between simulated 2018 data and the simulated 2014 to 2017 data. The power estimates depended on transect, year of comparison, distance from transect origin, and the simulated effect size (Figure 2). At the Coastal and Northern Transects, power was greatest at 500 and 1000 m and very low at 50 and 4000 m. Although statistical power was low at the Coastal Transect (range of 0.04 to 0.64 at 500 and 1000 m distances), power greater than 0.8 was achieved at effect sizes of 1 to 2 SD for all years between 2014 and 2017, compared to 2018. The results were similar for the Northern transect, where at distances of 500 to 1000 m, power greater than 0.8 was achieved at positive effect sizes between 1 and 3 SD. At the East and West transects, estimated power was variable depending on the distance and year being compared, with 2017 having the lower power relative to 2018. But overall, estimated power was greater than 0.8 for many (approximately half) of the annual comparisons, and for those that were not, power of 0.8 was often achieved at effect sizes between 1 and 3 SD (Figure 2).

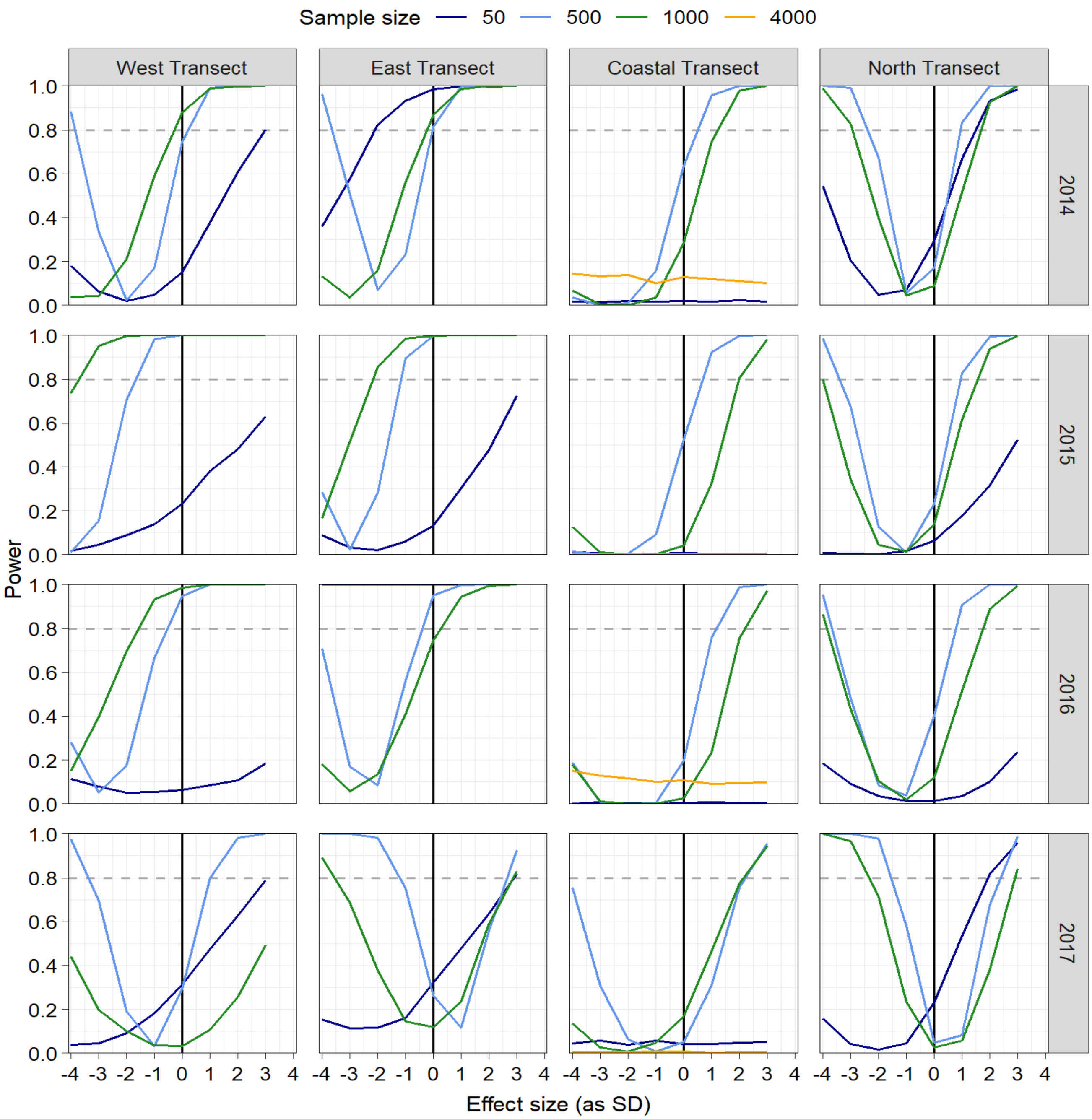



**Figure 2: Estimated Power to Detect Significant Differences in Sediment Iron Content between 2018 and Previous Sampling Years (2014-2017) for Varying Effect Sizes and Distances from Transect Origin.**

**Note: Power of 0.8 (Standard Goal) is Denoted as Dashed Line. Vertical Solid Line Denotes Effect Size of Zero (i.e., Observed Effect Size Only).**

## 3.3 Epifauna and Macroflora

Due to the non-linear relationship between distance from transect origin and both macroflora cover and epifauna abundance, it was not possible to use ANCOVA to analyze yearly differences in these two variables (Golder 2018). Instead, the analysis of percent macroflora cover and epifauna abundance in 2014 to 2017 was performed using ANOVA, with distance binned into 250 m intervals. The discretization of continuous data usually renders data less informative. In addition, not all distance bins were sampled in all years, and the 2017 North Transect data were omitted from analysis, because no macroflora was observed in the entirety of the transect, and because epifauna abundance was very high in two distinct parts of the transect. These two high-abundance segments of the transect had a very high density of smaller brittle stars, likely following a large settlement event. Due to these limitations of the dataset, power analyses for epifauna and macroflora were not conducted.

## 3.4 Benthic Infauna

Benthic infauna samples were collected as part of the MEEMP program for the first time in 2018. At least two years of sampling are required for comparison between years using regression analysis and to determine whether the statistical power is sufficient to detect differences. SEM (2015) reported that “the power analyses determined the sample size requirements to detect a change in benthic community were prohibitive (D. Schneider, Pers. Comm.), both in terms of sample collection effort and analytical costs” and, therefore, benthic infauna was originally excluded as a monitoring target for the MEEMP.

The lack of statistical power in benthic infauna data collected is attributed to the sample collection methods used. Benthic infauna samples from 2010 to 2018 were collected using a Petit Ponar grab sampler, which has a small opening area (15 x 15 cm) and is normally used in a freshwater environment. In the marine environment, sampling with the Petit Ponar grab results in high variability of area and volumes sampled and, consequently, high variability in density and diversity of detected organisms. Use of sampling equipment that is more adequate for the marine environment, such as Van Veen or Ponar grabs, may increase the statistical power of the analysis.

With only five sampling stations, there was not enough power to detect a ±2 SD change on any of the three transects for most samples sizes >5, although the North Transect had the highest (and sometimes sufficient) power (Figure 3). The simulation was then adjusted to mimic having more stations along the transects (15 stations at 100 m intervals). For this scenario, power was much higher, with the majority of transects being able to detect 2SD change with 3 samples per station (power approximately 0.75 for 3 samples per station; Figure 4). Based on these results, the 2019 MEEMP sampling design has been revised to include an increased number of benthic infaunal sampling stations along the North, West, and East transects (increase from 5 to 15 sampling stations on each transect with three subsamples collected per station). Furthermore, Baffinland has introduced an additional transect to the MEEMP study design oriented in a northeast direction starting at the new ore dock. This transect will also include 15 samplings stations with three sub-samples collected per station (Figure 5).

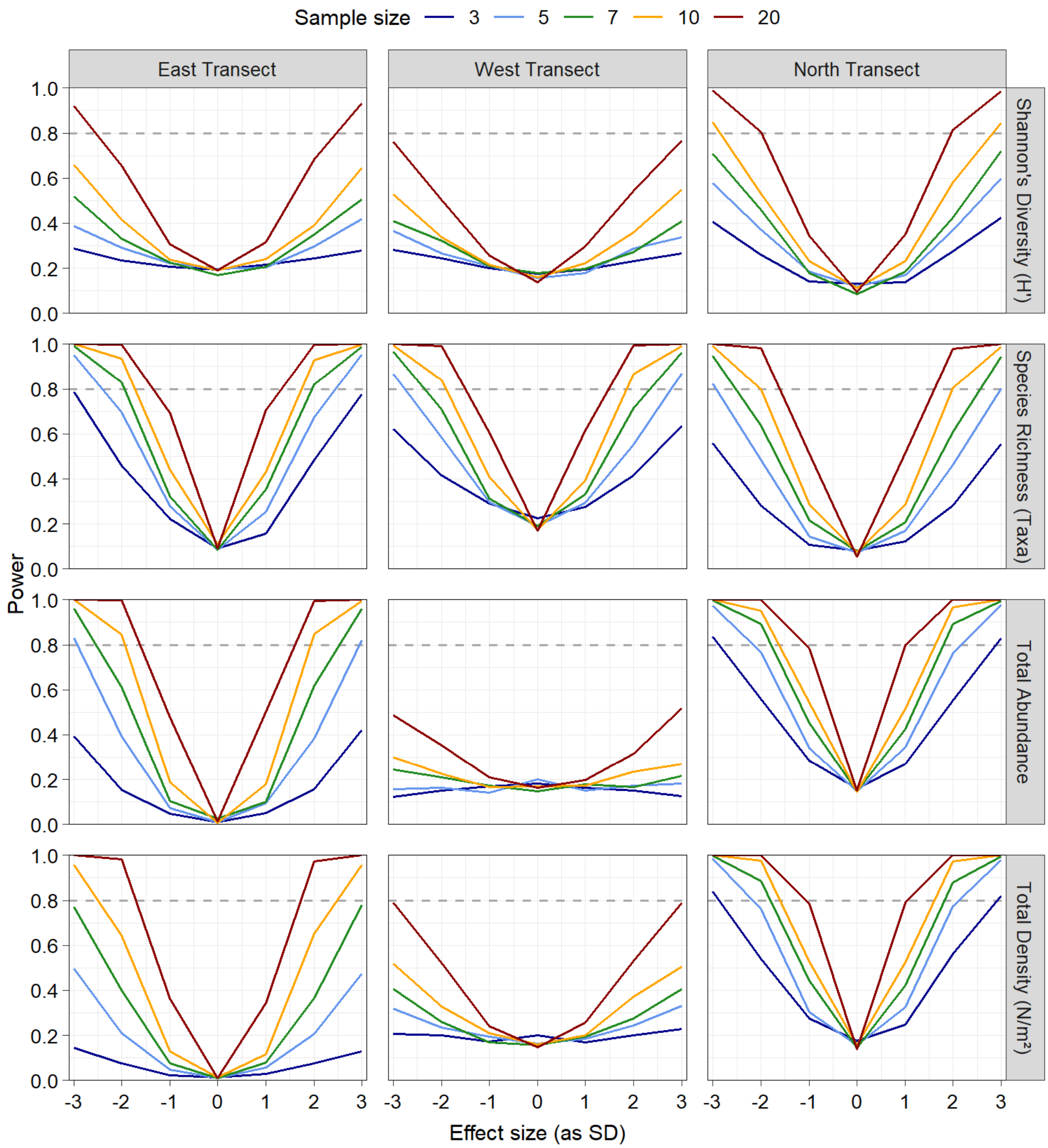



**Figure 3: Estimated Power to Detect Significant Differences in Benthic Infauna Based on 5 Sample Stations.**

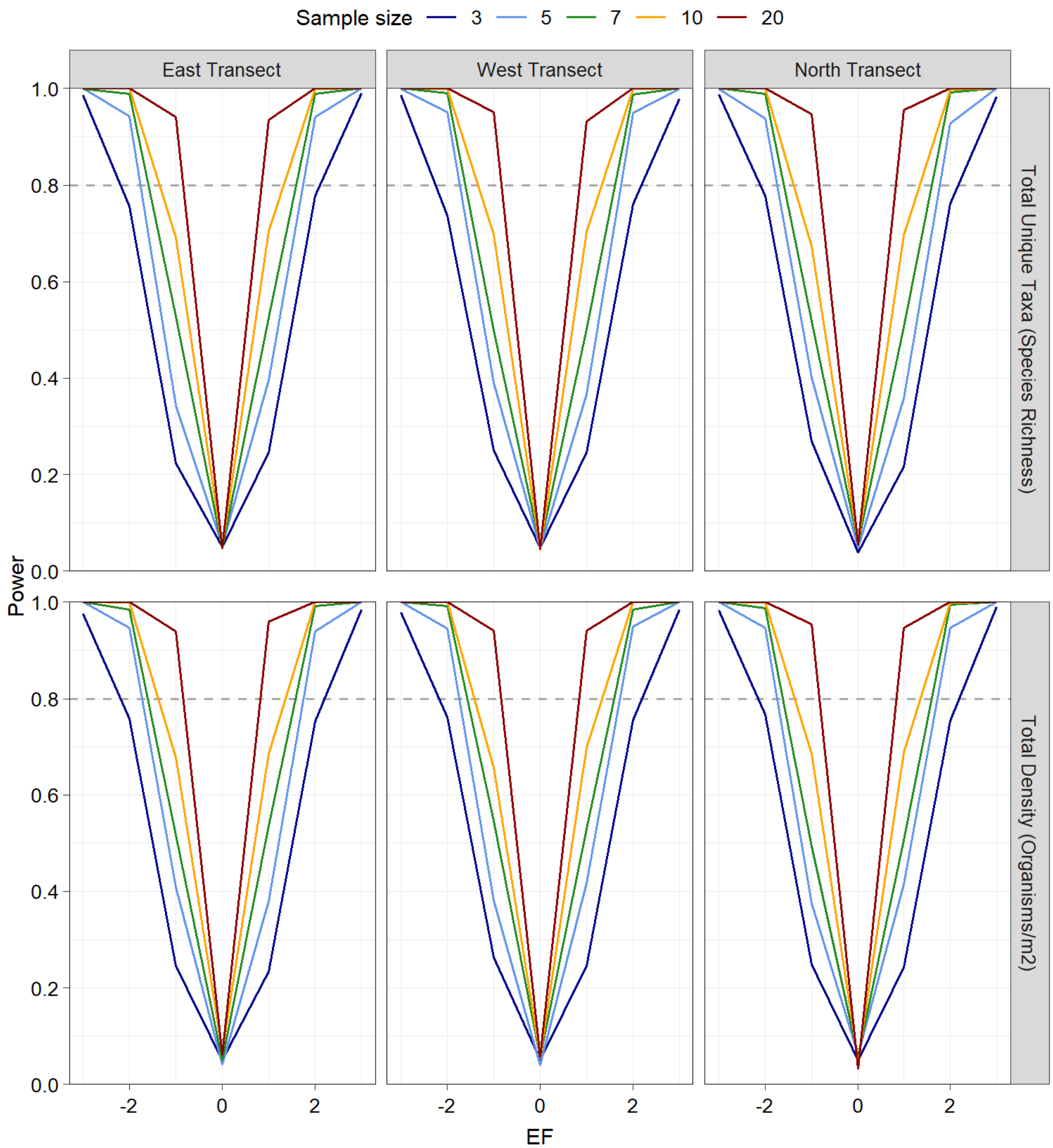
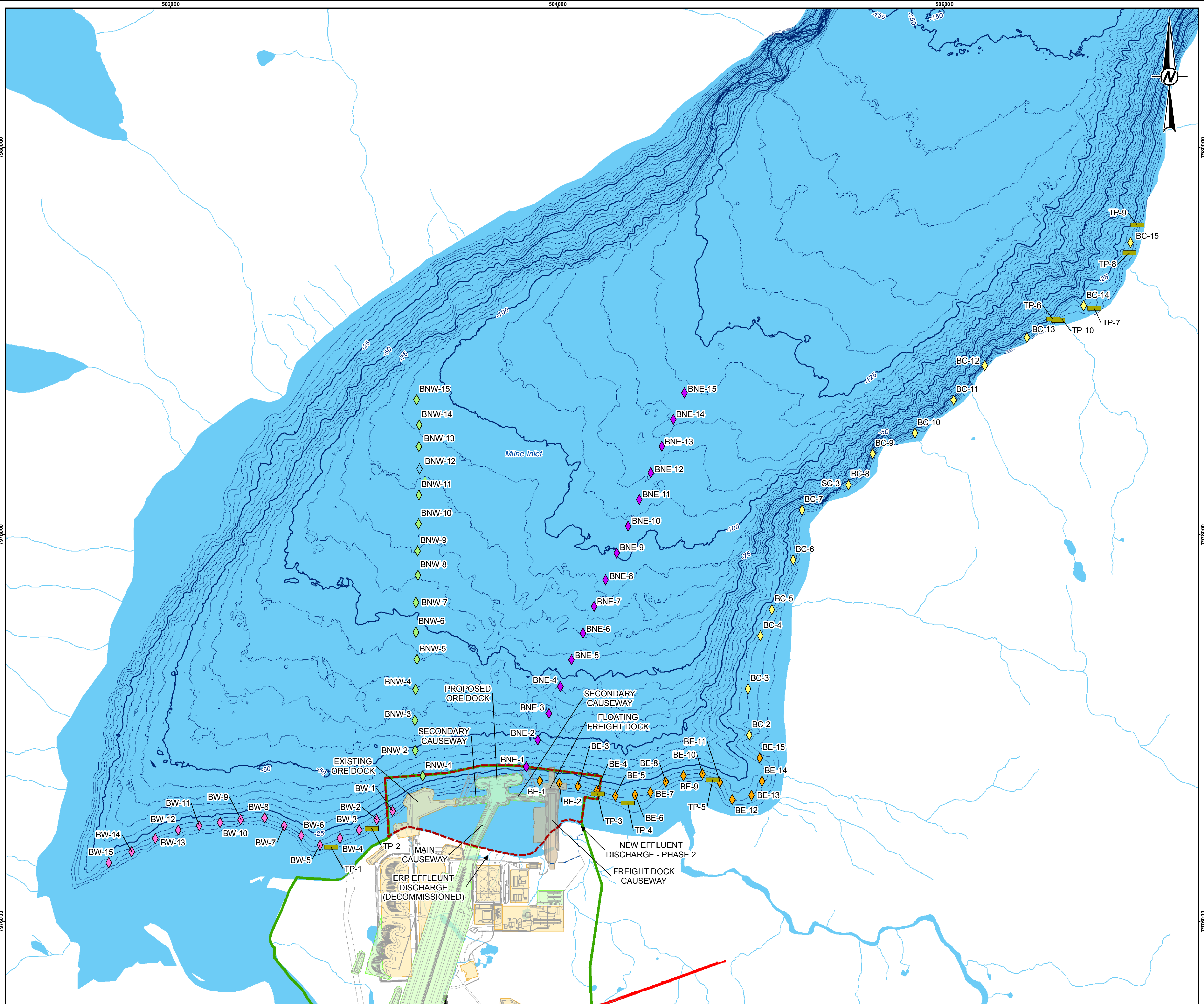



**Figure 4: Estimated Power to Detect Significant Differences in Benthic Infauna Based on 15 Sample Stations.**




**LEGEND**

- BELT TRANSECT

**BENTHIC INFAUNA AND SEDIMENT SAMPLING STATIONS**

- COASTAL TRANSECT
- EAST TRANSECT
- NORTH TRANSECT
- NORTH-EAST TRANSECT
- WEST TRANSECT
- BATHYMETRIC CONTOUR (5 m INTERVAL)
- BATHYMETRIC CONTOUR (25 m INTERVAL)
- NEW EFFLUENT DISCHARGE - PHASE 2
- PDA / QIA COMMERCIAL LEASE
- MILNE INLET TOTE ROAD
- PROPOSED NORTH RAILWAY
- WATERCOURSE
- INAC FORESHORE LEASE
- EXISTING FREIGHT DOCK AND CAUSEWAY
- EXISTING ORE DOCK
- PROPOSED SECOND ORE DOCK AND CAUSEWAYS
- WATERBODY

0 500 1,000

1:20,000 METRES

**REFERENCE(S)**

MILNE PORT INFRASTRUCTURE DATA OBTAINED FROM CLIENT, MAY 28, 2018, AND BY HATCH, JANUARY 25, 2017, RETRIEVED FROM KNIGHT PIESOLD LTD. FULCRUM DATA MANAGEMENT SITE MAY 19, 2017. BATHYMETRY CREATED BY GOLDER FROM MULTIPLE DATA SOURCES. HYDROGRAPHY AND TOPOGRAPHY DATA BY EAGLE MAPPING (2005), RETRIEVED FROM KNIGHT PIESOLD LTD. FULCRUM DATA MANAGEMENT SITE, MAY 2017. HYDROGRAPHY, POPULATED PLACE, AND PROVINCIAL BOUNDARY DATA OBTAINED FROM GEOGRATIS, © DEPARTMENT OF NATURAL RESOURCES CANADA. ALL RIGHTS RESERVED.
PROJECTION: UTM ZONE 17 DATUM: NAD 83

CLIENT
BAFFINLAND IRON MINES CORPORATION

PROJECT
MARINE MONITORING PLAN (MMP) - MARY RIVER PROJECT – PHASE 2

TITLE
**MARINE SEDIMENT QUALITY AND BENTHIC COMMUNITY MONITORING STATIONS**

CONSULTANT
GOLDER

| | |
|---|---|
| YYYY-MM-DD | 2019-09-13 |
| DESIGNED | AO |
| PREPARED | AA |
| REVIEWED | PR |
| APPROVED | PR |

| PROJECT NO. | CONTROL | REV. | FIGURE |
|---|---|---|---|
| 1663724 | 30000 | 0 | 5 |

## 3.5 Fish Length-weight

In the analysis of power of fish weight-length regression, the weight-length analysis had sufficient power to detect a 7% increase in slope for Arctic Char, and a 15% increase in slope for both Fourhorn Sculpin and Shorthorn Sculpin relative to 2017 data (Figure 6). These results are in agreement with the estimated significance of interannual differences in ANCOVA slopes (Golder 2019), where the slopes for Arctic Char, Fourhorn Sculpin, and Shorthorn Sculpin were 0.3%, 31%, and 11% higher in 2018 than in 2017, and only the interannual difference for Fourhorn Sculpin was statistically significant (P=0.04).

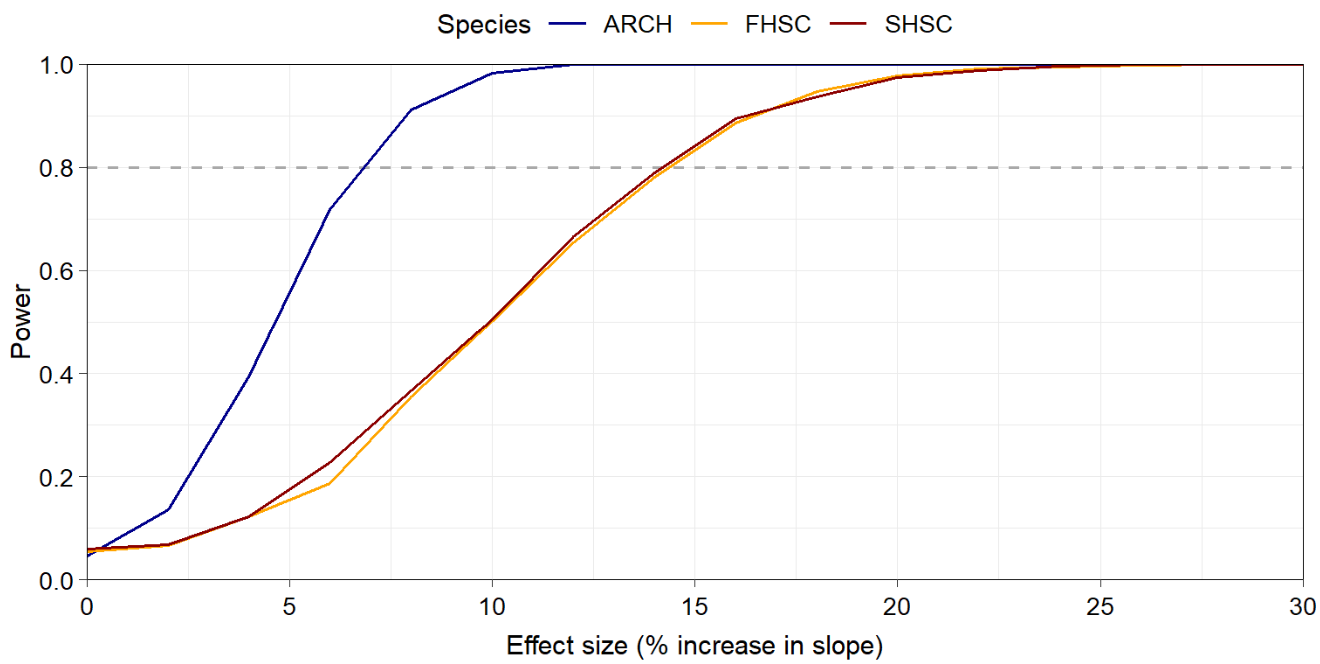



**Figure 6: Estimated Power to Detect Significant Differences in the Slope of Length-weight Analysis (Relative to 2017 Sampling Program) for Arctic Char, Fourhorn Sculpin, and Shorthorn Sculpin for varying effect sizes. Power of 0.8 (standard goal) is Denoted by Dashed Line.**

## 4.0 DISCUSSION

Simulation-based statistical power analyses offer a flexible tool to estimate statistical power for models regardless of model type or complexity. These tools therefore provide a single framework of power analyses.

The estimated power to detect annual differences in percent fine sediment at the Eastern and Western transects ranged from 0.06 to 0.91, over all years and distances. Most observed effect sizes were small, with only 14% (7 cases) of effect sizes for percent fines larger than 2 SDs of the collected 2018 data. Of these seven cases, the estimated statistical power was greater than 0.8 in three cases. At the Coastal and Northern transects, the estimated statistical power was zero or close to zero for all yearly comparisons, which was partly attributed to the small observed EFs (mostly <1 SD). Increasing the effect size to 3 to 4 SD larger than the observed difference results in sufficient power (>0.8) for the Northern transect but not for the Coastal Transect.

In yearly comparisons of iron content, estimated statistical power was reasonable for the Eastern and Northern transects at the observed effect sizes, with power greater than 0.8 for many of the comparisons with effects sizes as low as 1 SD compared to 2018. Statistical power was low at observed effect sizes at the Coastal and Northern transects, even when effect sizes were greater than 2 SD. Increasing effect sizes an additional 1 to 3 SD relative to 2018 generally resulted in sufficient power (>0.8) at the Coastal and Northern transects at distances of 500 and 1000 m, but not for distances of 50 or 4000 m.

Simulations of radial-gradient analysis of benthos data indicated that an increase in number of stations per transect from 5 to 15 considerably increased statistical power, so that power to detect a 2SD change in species richness or organism density was approximately 0.75 for a simulation of 3 samples per station. In comparison, when the analysis was performed using 5 stations per transect, power to detect a 2SD change in species richness or organism density ranged from less than 0.1 to 0.55, depending on variable and transect. These results have prompted a change in the benthic infauna sampling design from 5 stations per transect to 15 stations per transect, as per Baffinland's outlined commitment in Table 1.

Fish length-weight regressions had sufficient statistical power (>0.8) to detect a 7% increase in weight-length slope for Arctic Char, and a 15% increase in slope for both Fourhorn Sculpin and Shorthorn Sculpin relative to 2017 data. These effects sizes are well within the recommendations provided by Environment Canada (2012).

## 5.0 CLOSURE

We trust that this technical memorandum provides sufficient information for your present needs. Please direct any questions to the undersigned at Sima_Usvyatsov@golder.com or 506-343-4083.

**Golder Associates Ltd.**

Sima Usvyatsov, PhD
*Biological Scientist*

Shawn Redden, RPBio
*Associate, Senior Fisheries Biologist*

SU/SR/asd

o:\final\2016\3 proj\1663724 baff_marinemammalsurvey_ont\1663724-117-tm-rev1\1663724-117-tm-rev1-power analysis 07oct_19.docx

## 6.0 REFERENCES

Baffinland Iron Mines Corporation (Baffinland), 2012. Mary River Project – Final Environmental Impact Statement.

Baffinland, 2013. Mary River Project – Addendum to the Final Environmental Impact Statement.

Ellis, J.L. and D.C. Schneider. 1997. Evaluation of a Gradient Design for Environmental Impact Assessment. Environ. Monit. Assess 48:157–172.

Environment Canada. 2012. Metal mining technical guidance for environmental effects monitoring. EN14-61/2012E-PDF.

Golder (Golder Associates Ltd.). 2018. Mary River Project 2017 Marine Environmental Effects Monitoring Program (MEEMP) and Aquatic Invasive Species (AIS) Monitoring Program. Prepared for Baffinland Iron Mines Corporation, Oakville, Ontario. Golder Doc. No. 1663724-048R-Rev1; 9 April 2018. 504 p.

Golder (Golder Associates Ltd.). 2019. Mary River Project 2018 Marine Environmental Effects Monitoring Program (MEEMP) and Aquatic Invasive Species (AIS) Monitoring Program. Prepared for Baffinland Iron Mines Corporation, Oakville, Ontario. Golder Doc. No. 1663724-092-R-Rev0-14000; 22 February 2019. 714 p.

Length R. 2019. emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.3.4. https://CRAN.R-project.org/package=emmeans

R Core Team (R). 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL www.R-project.org

Seavy, N.E., S. Quader, J.D. Alexander, and C.J. Ralph. 2005. Generalized linear models and point count data: statistical considerations for the design and analysis of monitoring studies. USDA Forest Service General Technical Report PSW-GTR-191, pp. 744-753.

Sikumiut Environmental Management Ltd. (SEM), 2015. Marine Biological and Environmental Baseline Surveys Milne Inlet 2014. Prepared for Baffinland Iron Mines Corporation, Oakville, Ontario.