24551

Quantile Regression – A Statisticians Approach to the Local Ice Pressure-Area Relationship

Tom Morrison and Paul Spencer, Ausenco, Calgary, Alberta, CANADA

This paper was prepared for presentation at the Arctic Technology Conference held in Houston, Texas, USA, 3-5 December 2014.

Abstract

The ISO 19906 local pressure-area relationship represents a deterministic estimate of the failure pressure of massive ice features. For areas < 10 m^2 the pressure is represented as a power law on area. For areas between 10 and 40 m^2 the pressure has a constant value. In the past, in order to fit the power law, the data were binned and means and standard deviations for each bin were determined. The deterministic value of the local ice failure pressure is the relationship obtained by fitting a line to the values of mean plus three standard deviations so obtained. This implies the probability a local ice pressure exceeds the deterministic relationship is 0.13%, treating the deterministic value as a one-sided upper bound. It is shown herein that a quantitative, repeatable and non-subjective statistical methodology called quantile regression can be applied to this data. Using a nonlinear quantile regression methodology the deterministic relationship can be estimated in one step without the requirement for binning or transformation. Furthermore, the method produces quantile lines for a wide range of quantiles. The nonlinear quantile regression analysis of the local pressure area relationship, fitted to a power law, indicate the deterministic ISO relationship is in fact equivalent to an exceedance of 5% for areas greater than 2 and less than 10 m^2. For areas of less than 2 m^2 the ISO relationship exceedance decreases from 5% at 2 m^2 to less than 1% for areas less than 0.3 m^2.

The purpose of this paper is to introduce this powerful technique to the ice science and ice engineering communities using the local ice pressure data as an example.

Introduction

It is desired to have at hand a recommended design level for local ice failure pressure against a vertically sided structure, including some amount of conservatism. In past ice/structure code developments, such as CSA 471, it has been suggested that a reasonable deterministic estimate of local and global ice failure pressure would be obtained by adding to the mean of measured data at a given area, some multiple of the measured standard deviation of the data points for the same given area. In this fashion the assessment of ice failure pressures would be conservative, but not overly conservative.

The objective is to have a relationship between pressure and area that is exceeded on an infrequent basis. The method that has been used to determine the deterministic line in ISO involves binning, calculating the mean and standard deviation of the binned data, and performing a least-squares fit to the estimate of mean + 3σ (mean + three standard deviations σ) and using this line as the deterministic value as a function of area. In statistical terminology, this line would be referred to as the 99.87% regression quantile. The quantile regression methodology uses the data without binning and directly generates quantiles at a specified exceedance level. The method removes the requirement for subjectively processing the data, thereby resulting in a method that will produce a unique answer. While used in other areas of science and engineering, quantile regression appears to be unknown or at least unused in the ice community.

Whereas a statistician would comment that the desired exceedance at mean + 3σ is the 0.13% exceedance, it is more appropriate to state that the intent of those developing the deterministic line was to estimate a 1% exceedance using the tools and methodologies (binning, etc.) available.

The outline of this paper is as follows: The raw data are first presented. Previous methodologies for estimating the upper bound of the local ice pressure relationship are then discussed. The meaning of estimating a one-sided upper bound three standard deviations above the mean is presented. Quantile regression is introduced for the linear model. The non-linear quantile regression results are then presented. A comparison of the nonlinear and linear results (on the logarithms of area and pressure) is then presented. The paper finishes with a discussion of available software, some issues with quantile regression and further work.

The methodology presented in this paper is available as an easily available software tool (created by others) that can be added to the repertoire of the ice scientist and ice engineering communities.

Spencer's previous reviews and analyses of the variability introduced by different methods of estimating the pressure-area relationship (see Spencer (2013), for example) was responsible for initiating the search by the present authors for a more appropriate and statistically valid approach to analyzing this data set. A companion paper at this conference (Spencer and Morrison, 2014) illustrates practical use of the methodology for the purpose of probabilistic modelling of ice loads, and also expands the use of the methodology to global ice loads calculations.

We have deliberately avoided flooding this paper with the mathematics of the methodology because we believe it more appropriate, as an introduction to quantile regression, to explain our procedural goals in words. We provide references to the quantile regression methodology and code.

In this paper we are concentrating on determining estimates of the parameters of the power law formula for the local pressure-area relationship, thus the use of the nonlinear methodology. Other quantile regression methodologies exist, such as nonparametric quantile regression. While use of these methodologies may be appropriate from a statistical viewpoint, they may not be appropriate from an ice community viewpoint because they will yield a curve for the relationship that cannot be described as simply as the power law relationship or other closed form expression. We are somewhat constrained in that our desire for an easily communicable formula removes candidate methodologies from our list of alternative methodologies.

The Raw Data

The data being analyzed are shown in Figures 1 and 2, first on a linear plot followed by a plot wherein both pressure and area have been transformed using logarithms. The data were kindly provided to the authors by Bob Frederking of NRC Canada.

The data are from a mixture of small and medium scale tests, and estimated failure pressures for areas of 3, 6 and 9 m^2 using load data recorded on the Molikpaq. All of the data were obtained at locations in the Canadian Arctic.

A single test can be responsible for more than a single data point. In many of the larger test data samples a shape was pushed into the ice and several peak loads measured. Cases where large scale failure of the sample occurred before a peak load was obtained were usually removed from data sets.

The range of areas is greater than two orders of magnitude. The range of pressures is one and one half orders of magnitude. Speaking in a statistical manner, the researcher is attempting to estimate the upper 99.87% confidence bound for a data set that: 1) is heteroskedastic (does not have the same variance throughout the range); 2) most likely does not have errors that are distributed as a Gaussian at a given area; 3) has sample selection issues in that the researcher "chose" the areas at which to make the measurements because the size of the test instruments used to sample was not random, and the sizes were convenient for the researcher; and 4) contains a relationship that has a local minimum at an area of 1.5 m^2 and a local maximum at an area of 2.5 m^2, thus the curvature of the data is inconsistent within the range.

To highlight the four points above, consider the schematic illustrations in Figures 3 and 4 below. The first illustration is what a statistician desires – the determination of a smooth line where the data points (not shown) surround the line with local variances equal throughout the data set. The second illustration illustrates the local pressure-area data set under examination: 1) the variance of the data changes throughout the data set; 2) the errors are not necessarily Gaussian; 3) the data points are not spread out evenly throughout the data set; and 4) the data are not monotonic in that local minima and maxima occur within the measured bounds.

This is a difficult data set to analyze. The data require a non-trivial methodology for analysis.

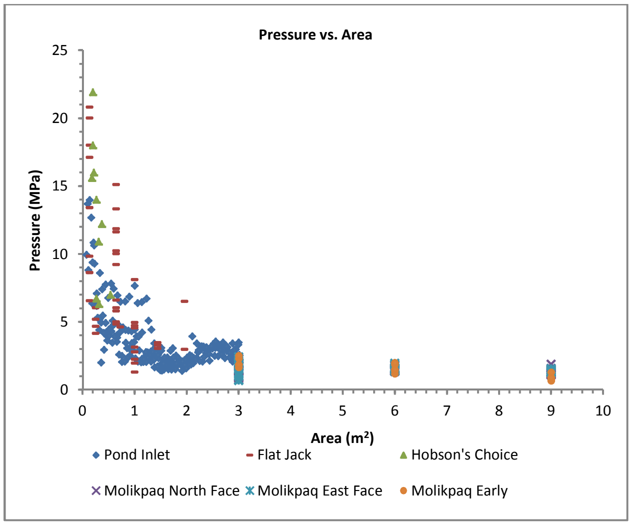

Figure 1 Raw Data on Linear Scales

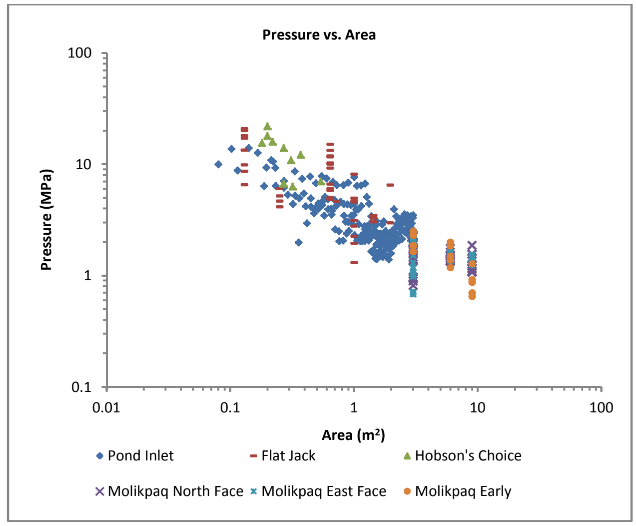

Figure 2 Raw Data After Transforming to Logarithms

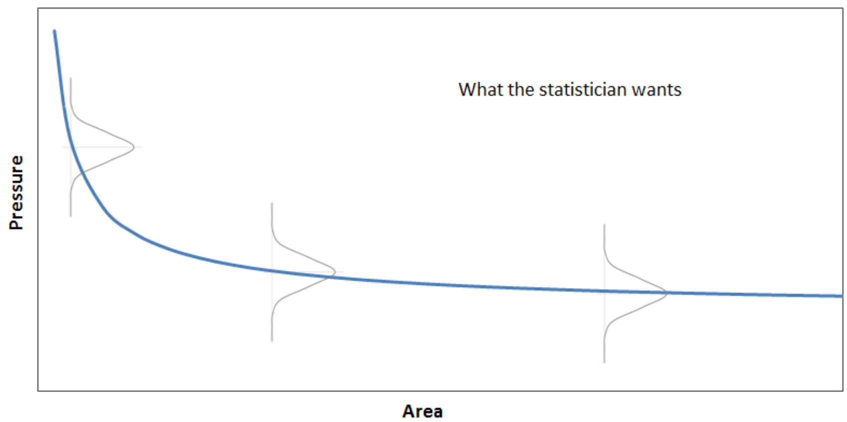

Figure 3 What a Statistician Wants – the Assumptions that must be Satisfied for a Simple Regression Model

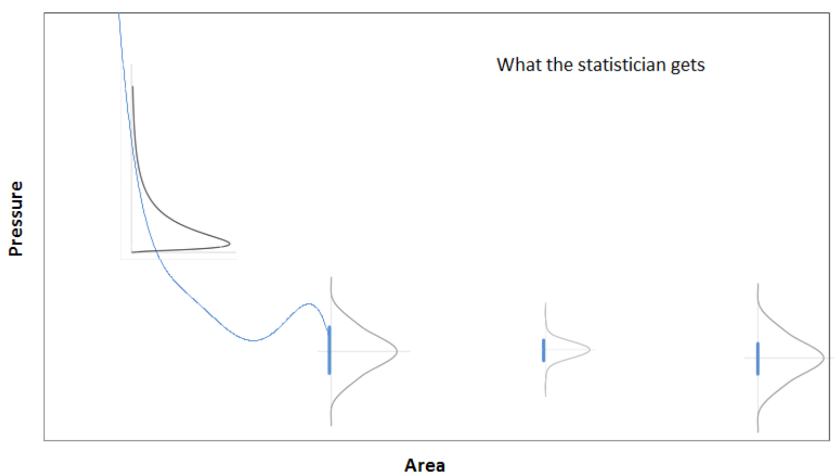

Figure 4 What a Statistician Gets – The Broken Assumptions Presented by the Pressure-Area Relationship

Previous Methodology: Implementing the Mean + 3σ Estimate

In statistical terms the stated desired estimate is the one-sided upper bound of the mean + 3σ, as a function of area, such that a small percentage of local ice loadings would have failure pressures greater than the deterministic relationship so determined. Examples of two- and one-sided upper bounds, for data distributed as a normal or Gaussian distribution, are given in Table 1.

Table 1 Confidence Bounds Represented as Standard Deviations *z*, as a Function of Coverage, for the Normal Distribution Assumption of Distribution of Errors

α	Coverage 100 (1 – α)%	Confidence Bound Two-Sided Normal *z*	Confidence Bound One-Sided Normal *z*
0.32	68	1.00	0.47
0.10	90	1.645	1.28
0.05	95	1.96	1.645
0.025	97.5	2.24	1.96
0.01	99	2.58	2.33
0.005	99.5	2.81	2.58
0.0025	99.75	3.00	2.81
0.0013	99.87	3.24	3.00

A one-sided mean + 3σ upper bound estimates the upper bound is exceeded in about 0.0013 of the total number of interactions. For example, say for a given area the mean was 2 MPa and the standard deviation equal to 1 MPa. The researcher would then report the deterministic value at 5 MPa, and say that only 0.13% of pressures at this area would be higher than 5 MPa over the long term.

Recall Figure 1 above. What is a researcher to do? Spencer (2013) considered three methods of analyzing the data, the first one consisting of the following "standard" steps (see also Masterson et al., (2007)):

1. In linear or log space, bin or group the data according to individual area ranges such that a reasonable number of data points would be in each bin, from which to estimate the mean and 3σ for each bin.
2. Obtain logarithms of the estimates of mean + 3σ pressure values, and the logarithm of the area at the midpoint of each bin.
3. In log-log space, perform a linear least-squares regression on the values so obtained, with each point having equal weight. Use the resulting relationship as representing the deterministic upper bound of local ice pressure.

The methodology is fraught with uncertainty and is not easily duplicable by other researchers unless the original bins are known. One issue is whether the area range within each bin should be the same, or whether the area ranges should be such that the same number of data points reside within each bin. More data points enable, perhaps, a better estimate of the standard deviation and mean for that particular bin, but do not enable, perhaps, a reasonable number of estimates of the means and standard deviations from which to somehow estimate the relationship using linear regression.

Consider a subsequent researcher beginning anew with the raw data having to make his or her own decisions with respect to the binning. They may easily produce a different estimate of the upper bound, adding to the confusion. Spencer (2013) contains some details of variability in the estimate of the upper bound based on methods similar to those above – the three estimates of his local pressure-area relationship are compared with quantile regression estimates in Table 2.

Transforming to log-log Data and Using a Least-Squares Methodology is not Recommended

It may be expected that transforming the data by obtaining logarithms of each parameter, pressure and area in this case, analyzing the relationship using a least-squares methodology on the data so transformed, and back transforming into linear-linear space would yield the same relationship as performing a nonlinear methodology on the data. This is not so. One reason for the breakdown in the estimated relationship is that a systematic, and often neglected, multiplier is required to be included in the back transformed relationship, the multiplier being the exponential of the average of the errors in the linear data set. Secondly, notwithstanding that the errors in the original data may or may not be Gaussian with mean zero and equal standard deviations at each area, the log-log data will have a different error distribution, this error distribution actually depending subtly on the relationship to which one attempting to fit the data.

To be general, the pressure-area data should not be transformed. Fortunately, there exists a methodology that can contribute to the understanding of the relationship, not involving estimation by taking logs, binning the data and using a least squares fit of the estimated mean + 3σ relationship. This methodology is called quantile regression.

Linear Quantile Regression

The meaning behind the statement "the student's mark is at the 90th percentile" is that the student scored a mark equal to or greater than 90% of his or her fellow students. The percentile can also be called a quantile. The 50th percentile, or 50th percent quantile, or 0.50 quantile, is the median of the data set.

Imagine now, a researcher wanting to describe a two-dimensional data set such as the pressure-area relationship. A least-squares regression can be applied to the data to obtain the mean response of the dependent variable (pressure) as a function of the independent variable (area). In fact, as stated by Koenker (2005, page 75): "The classical theory of linear regression *assumes* that the conditional quantile functions of the response variable, *y*, given covariates *x*, are all parallel to one another, implying that the slope coefficients of distinct quantile regressions will be identical." How to obtain the full relationship is the issue. To continue, in classical least-squares regression, confidence bands can be estimated for the line as a whole, with any given coverage property. The bands are curved with the minimum approach to the fit line being at the mean of each of the two variables. The probability distribution of a future measurement *y,* given *x,* can also be estimated. However, this way of representing the two-dimensional relationship may not be exactly what the researcher wants. The situation under study is a very obvious case wherein the researcher desires an upper bound to the relationship such that they can state that all future data will be below the upper bound with a given probability.

Quantile regression comes to the aid of the researcher. In the two-dimensional case, for each given quantile, the methodology calculates an upper bound line (for linear data) or curve (for nonlinear data) such that the chosen percentage of data (for the given quantile) will lie below the upper bound line. The estimated "mean + 3σ" methodology can be replaced by quantile regression at a quantile equivalent to "mean + 3σ." Quantile regression was introduced to the econometrics and statistics community by Roger Koenker and Gilbert Bassett Jr. in 1978 (Koenker and Bassett Jr., 1978).

The quantile regression methodology is a statistically valid, quantifiable, reproducible and non-subjective methodology for the estimation of the deterministic local failure pressure relationship. The quantile regression methodology removes bias due to arbitrarily binning the data as a function of area and performing a least-squares methodology. Though not a panacea -- some issues are described near the end of this paper -- the methodology is a firm step in the proper direction for estimating the deterministic upper bound of local ice pressure.

The quantile regression methodology has several advantages over classical least-squares. These advantages are general, and are not specific to the present context. The quantile regression methodology:

1. is robust, meaning quantile regression estimates are less susceptible to outliers as compared to least-squares;
2. has less variability than least-squares when assumptions of heteroscedasticity and Gaussian error distribution are not valid; and
3. can better describe the total conditional distribution of the dependent variable, yielding a far more complete picture of the total relationship between area and pressure.

In quantile regression the objective is to determine a suite of regression relationships describing the whole of the relationship between two variables. Figure 5, from Koenker (2005), aids the discussion of the objective and methodology.

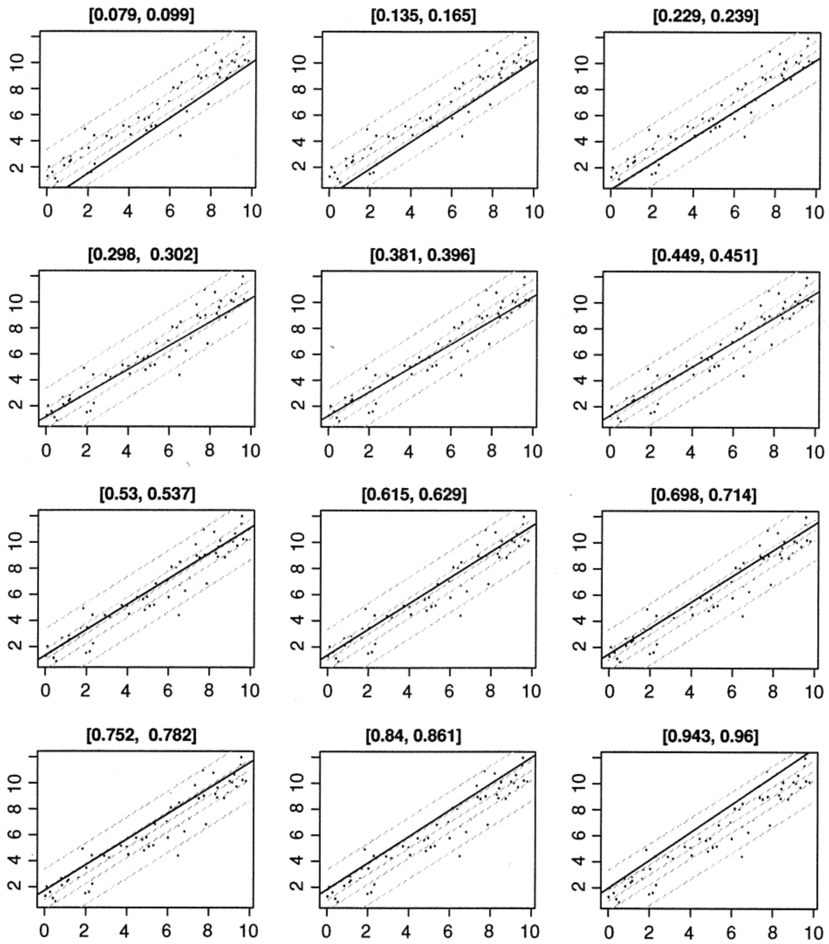

Figure 5 Bivariate Linear iid Quantile Example (Koenker, 2005)

Koenker generated the random data using the following structure in a monte carlo selection,

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

where the u_i are Gaussian with mean zero and some given standard deviation. The conditional quantile relationships of *y* are given by

$$Q_y(\tau|x) = \beta_0 + \beta_1 x + F_u^{-1}(\tau)$$

where $F_u(\tau)$ is the cumulative probability distribution function of the errors as a function of the quantile τ, assumed equal as a function of *x* in this example. In Koenker's example the errors are the same for each *x*, and are described by a Gaussian distribution with mean zero.

In Figure 5, the set of quantiles {0.05, 0.25, 0.50, 0.75, 0.95}, shown as the dashed parallel lines, are the true quantiles of the input relationship. These quantiles are vertically displaced, by definition having different intersections $\beta_0 + F^{-1}(\tau)$ and equal slopes β_1.

Now perform a quantile regression analysis on the sample of 60 data points obtained by monte carlo. The black line is the quantile regression line that marches through the dataset from small to large quantiles. The quantile ranges at the top of each individual plot are related to the line in the plot in that the line in the plot is the optimal quantile regression line for that small range of quantiles. The ranges are shown because for this example data set 66 different quantile regression estimates can be generated. Koenker (2005, page 13) cannot resist making a statistical joke using this data set as an example (recall what the statistician wanted and what the statistician got above):

If real data analysis were always as well behaved as the iid linear model depicted…there would be little need for quantile regression.

In data sets with a small number of data points, say less than 20 or so, it is likely that the quantile regression line for quantiles $\tau = \{0.95, 0.98, 0.99\}$ may all be represented by the same line.

If the local ice pressure-area relationship could be modelled as a bivariate linear quantile regression line with independent and identically distributed (iid) errors, we would say that the conditional relationship would be a set of parallel lines as described in the example above.

The estimates of the regression quantiles are derived using linear programming and interior point methods, familiar tools in the Operations Research and Management Science toolkit. An illustration of the functional form of the linear quantile regression method is provided in Figure 6. The essence of the linear method is a weighting dependent on the quantile of interest, τ, that is different for a point above or below the quantile line being estimated, multiplied by the absolute difference between the point and the line. The minimum of the sums of the product of quantile τ and absolute distance above the line, and the product of $(1 - \tau)$ and absolute distance below the line determines the linear relationship for that specific quantile τ. What the weighting achieves is that the set of the points above the quantile line has equal weight to the set of points below the quantile line.

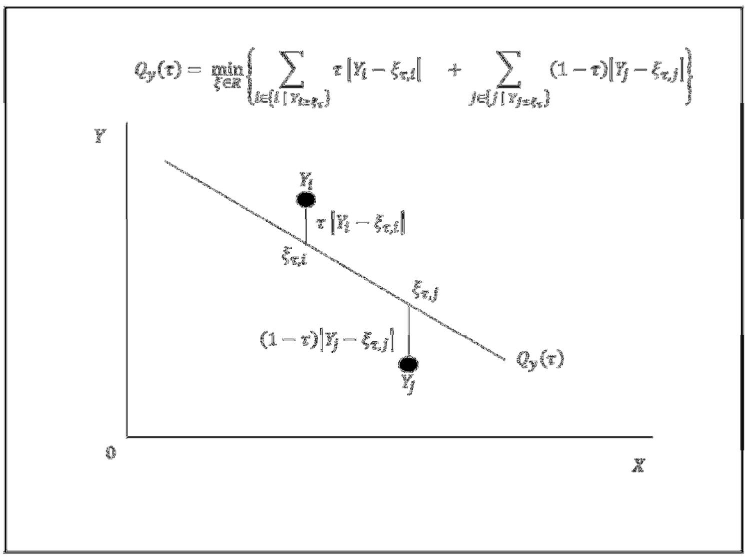

Figure 6 Minimization Function in Linear Quantile Regression

Non-Linear iid Quantile Regression

The non-linear iid quantile regression problem is solved using an interior point algorithm. The nonlinear quantile regression interior point algorithm methodology is described in Koenker and Park (1996).

We are able to obtain estimates of *a* and *b* for the relationship, $y = a\,x^b + \varepsilon,\ \varepsilon \sim N(0, \sigma^2)$, and estimates of the errors of *a* and *b*, using the nonlinear iid quantile regression interior point algorithm. It is the estimates of *a* and *b* that we wish to compare with the deterministic relationship given in ISO 19906, Section A.8.2.5.3.

The interior point algorithm is an iterative procedure that begins with initial estimates of the parameters. Different starting values for *a* and *b* can yield slightly different final estimates of *a* and *b*.

The results, for quantiles from 0.05 to 0.95, using starting values of 10 for *a* and -0.5 for *b*, are presented in Table 2 below, where they are compared with previous work.

The quantile relationship estimates obtained using the nonlinear iid interior point algorithm, are shown in Figure 7. The same estimates, transformed to log-log space, are shown in Figure 8 for comparison with the usual log-log graphs of the local pressure-area relationship as shown in ISO 19906.

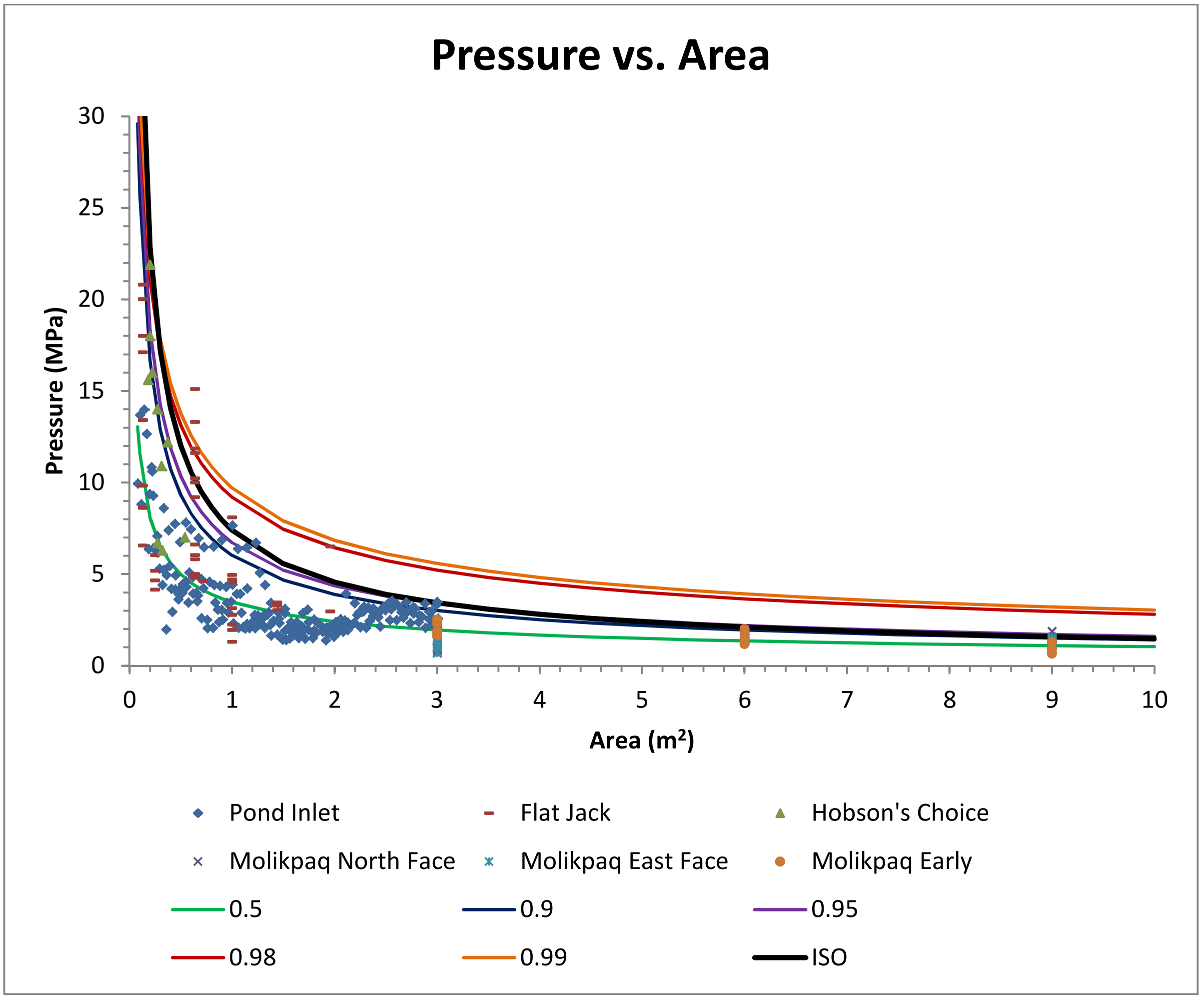

Figure 7 Comparison of ISO 19906 and Results of Nonlinear iid Quantile Regression on Linear Scales

ISO 19906, in Section A.8.2.5.3 (Annex A - Local pressures for thick, massive ice features) provides guidance for the local pressure area deterministic relationship. It is written in the legend of Figure A.8-19 that the deterministic relationship is equal to that of the mean + 3 σ.

The relationship given in ISO 19906 is: Pressure = 7.40 Area$^{-0.70}$.

Recall that the α value for a one-sided upper bound three standard deviations z above the mean is 0.0013, or 0.13%. The comparison of the relationships in Figure 7 indicates the ISO deterministic local ice pressure line corresponds to the 0.95 quantile for areas greater than 2.0 m^2, and is greater than or equal to the 0.98 or 0.99 quantile only for areas less than 0.3 m^2. In other words, for areas greater than about 2 m^2 the ISO local ice pressure deterministic line, at 5% exceedance, is more appropriately described as mean + 1.65 σ rather than mean + 3σ.

The area exponents of the nonlinear iid quantile regression fits are all less than or equal to -0.63 compared to the ISO area exponent of -0.70. The quantile regression estimate indicates that as the area increases, the local failure pressure does not decrease by the same ratio as the ISO deterministic line would have us believe.

The slopes of the nonlinear iid quantile regression lines for τ = {0.05, 0.10, … , 0.95, 0.98, 0.99} are all in the range (-0.495, -0.63) with the shallowest slope at a quantile of about 0.4 and the steepest slopes at the lowest (5%) and highest quantiles (95, 98 and 99%).

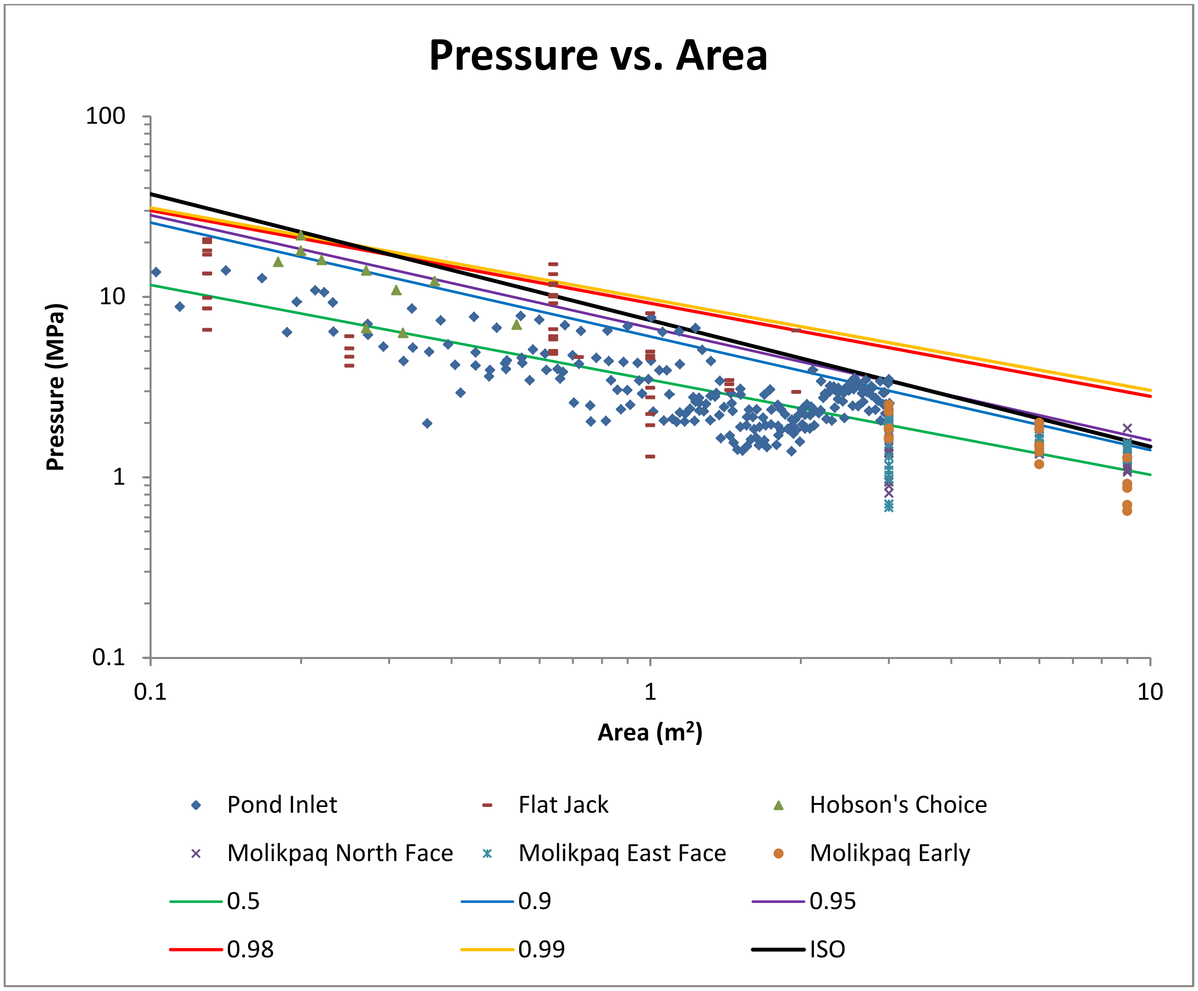

Figure 8 Comparison of ISO 19906 and Results of Nonlinear iid Quantile Regression on log Scales

Comparison of Nonlinear iid Quantile Regression Estimate with ISO 19906 and Other Previous Work, and Interpretation of the Comparison

Spencer (2013) illustrated the results of three methods of estimating the deterministic pressure-area relationship. The confidence intervals are two-sided one standard deviation, or 68% coverage, confidence intervals.

The comparison of previous relationships with those generated herein is shown in Table 2 below. The ISO 19906 deterministic relationship from Section A.8 is first shown, followed by the three relationships from Spencer (2013). The nonlinear iid quantile regression results are next shown, followed by the linear iid quantile regression results obtained by transforming the data using logarithms (recall that we do not recommend performing the analysis using the logarithmic transformation).

The powers estimated in the nonlinear relationship are greatest at about the 0.90 quantile and smallest at the 0.5 and 0.99 quantile. Variable slope estimates for different quantiles is an indication of some sort of heteroscedasticity in the data.

We would suggest that, for the power law relationship similar to that in ISO 19906, the nonlinear iid 99% quantile relationship, highlighted in red in Table 2, be considered. Our companion paper should be read since it contains further discussion of what can be recommended for the pressure-area relationship.

Software for Linear and Non-Linear Quantile Regression

The ice science and ice engineering communities are fortunate in that software is readily available for solving some of the linear and non-linear problems, and that some of the software is free under the GNU license.

The quantile regression package, quantreg, is part of the CRAN (The Comprehensive R Archive Network) repository on the internet. CRAN is a repository of statistical and graphical software that is freely available for use by researchers. See

Koenker (2012) for details of the package.

The benefit of not having to code the basics of the linear and nonlinear quantile regression methodologies is that basic implementation is much simpler than may have been implied in this paper.

Table 2 Comparison of ISO, Other Previous Estimates, and Quantile Regression Estimates

ISO 19906		
Regression Method	**Power Law Fit**	**Power Law Fit Including Standard Errors (68% Coverage)**
Take logs, equal-weight, linear least-squares regression of binned data	7.40 $A^{-0.70}$	Not Available
Spencer (POAC, 2013)		
Regression Method	**Power Law Fit**	**Power Law Fit Including Standard Errors (68% Coverage)**
1: Take logs, equal-weight, linear least-squares regression of binned data	7.40 $A^{-0.704}$	7.40 (± 0.46) $A^{-0.704\ (\pm\ 0.065)}$
2: Equal-weight fit to power function, nonlinear least-squares regression of binned data	7.89 $A^{-0.652}$	7.89 (± 0.63) $A^{-0.652\ (\pm\ 0.059)}$
3: Weighted fit to power function, inverse weighting of binned data to variance of mean + 3σ	6.31 $A^{-0.624}$	6.31 (± 0.46) $A^{-0.624\ (\pm\ 0.065)}$
Nonlinear (iid) Quantile Regression (Starting Estimates: constant = 10 and power = -0.5) Including Standard Errors (68% Coverage)		
Quantile	Power Law Fit	Including Standard Errors
0.50 (median)	3.462 $A^{-0.525}$	3.462 (± 0.159) $A^{-0.525\ (\pm 0.047)}$
0.80	5.090 $A^{-0.619}$	5.090 (± 0.210) $A^{-0.619\ (\pm 0.034)}$
0.90	6.029 $A^{-0.630}$	6.029 (± 0.255) $A^{-0.630\ (\pm 0.025)}$
0.95	6.724 $A^{-0.623}$	6.724 (± 0.436) $A^{-0.623\ (\pm 0.056)}$
0.98	9.193 $A^{-0.515}$	9.192 (± 0.689) $A^{-0.515\ (\pm 0.060)}$
0.99	**9.712 $A^{-0.505}$**	**9.712 (± 0.864) $A^{-0.505\ (\pm 0.081)}$**
Linear Quantile Regression of log-log Data		
Quantile	**Power Law Fit**	Not Available
0.50 (median)	3.376 $A^{-0.470}$	Not Available
0.80	5.066 $A^{-0.593}$	Not Available
0.90	6.050 $A^{-0.641}$	Not Available
0.95	7.073 $A^{-0.694}$	Not Available
0.98	8.100 $A^{-0.750}$	Not Available
0.99	9.550 $A^{-0.742}$	Not Available

Some Limitations of Quantile Regression in the Present Context

Ideally, for small data sets where estimation of the 0.9987 quantile is not possible due to too few data points, we would like to be able to perform an extremal quantile regression. Our specific case is that of nonlinear, non-iid data versus a nonlinear iid data analysis methodology. It has not proven possible to find a nonlinear non-iid methodology nor has it proven possible to find methods that can estimate extremal aspects of a nonlinear (iid or non-iid) data set. Extremal quantile regression methods only appear to be available for the linear iid model.

- What is the greatest quantile that can be calculated for a two parameter (intercept and slope) model containing n data points? If a data set does not have enough data points to enable estimation of quantiles above 99%, a nonlinear extremal quantile regression estimate should be added to the capabilities of the quantile regression methodology.
- Do the quantile curves cross --- i.e. if the 0.90 quantile curve crosses the 0.98 curve we can only be embarrassed due to the non-monotonicity of the estimates.

- In data sets with a small number of data points, say less than 20 or so, it is likely that the quantile regression line for quantiles $\tau = \{0.95, 0.98, 0.99\}$ will all be represented by the same line.
- We have been forced to analyse the data using the nonlinear methodology assuming the errors are iid.

Summary, Conclusions and Further Work

Herein we suggest a statistically valid, quantifiable, reproducible and non-subjective methodology for the estimation of the deterministic local failure pressure relationship for a given quantile or, equivalently, a given probability of exceedance. This is what Masterson et al. (2007) appear to have been attempting to perform, without access to quantile regression – determine a design pressure relationship that has a low probability of exceedance. In other words, they were doing quantile analysis by an indirect method. As demonstrated in this paper, there exists a method that performs the quantile analysis directly and produces a pressure design relationship at a known probability of exceedance.

The greatest quantile calculable depends on the number of data points in the data set. The quantile regression methodology completely removes any bias due to arbitrarily binning the data as a function of area and performing a least-squares methodology. This is a major step forward for analysing this very complicated data set.

The parameter estimates shown herein are for comparison with the deterministic ISO relationship only, and we have provided constructive criticism of the deterministic ISO relationship for small areas. The companion paper uses the quantile regression methodology and examines in more detail the results that can be obtained.

What can be stated is the following: Assuming it is permissible to model the relationship using the iid error assumption, the nonlinear quantile regression analysis of the local pressure area relationship, fitted to a power law, indicates the deterministic ISO relationship is in fact equivalent to an exceedance of 5% for areas greater than 2 and less than 10 m^2. For areas of less than 2 m^2 the ISO relationship exceedance decreases from 5% at 2 m^2 to less than 1% for areas less than 0.3 m^2.

Future work is specifically required for the nonlinear model. In particular, efforts to overcome heteroscedasticity and non iid errors for the nonlinear non iid model are required. An extremal analysis using the nonlinear, non-iid quantile regression model should be developed.

The companion paper by Spencer and Morrison (2014), also presented at this conference, includes applications of the quantile regression methodology to local and large scale estimation of the ice failure pressure relationship.

Acknowledgement

The authors are indebted to Robert Frederking of NRC Canada for providing the data set consisting of the 351 ice failure pressure measurements analyzed in this paper.

References

Canadian Standards Association, (2004). CSA Standard S471-04: General requirements, design criteria, the environment, and loads, February 2004.

International Standards Organization, (2010). International Standard 19906. Petroleum and natural gas industries – Arctic offshore structures. First edition, 2010-12-15.

Koenker, Roger (2005). *Quantile Regression*, Cambridge University Press.

Koenker, Roger (2012). quantreg: Quantile Regression. R package version 4.94. http://CRAN.R-project.org/package=quantreg

Koenker, R., and G. Bassett Jr., (1978). Regression Quantiles, *Econometrica*, Vol. 46, No. 1, pp. 33-50.

Koenker, R., and B. Park, (1996). An Interior Point Algorithm for Nonlinear Quantile Regression, *Journal of Econometrics*, Vol. 71, pp. 265-283.

Masterson, D.M., R.M.W. Frederking, B. Wright, T. Karna and W.P. Maddock, (2007). A Revised Ice Pressure-Area Curve, POAC'07, Dalian, China, Volume 1, pp. 305 – 314.

Spencer, Paul (2013). Unifying local and global ice crushing pressures. POAC'13, Espo, Finland.

Spencer, Paul, and Tom Morrison (2014). Quantile regression as a tool for investigating local and global ice pressures. Arctic Technology

Conference, Paper 24550, Houston, TX.

The Comprehensive R Archive Network. http://cran.r-project.org/