Journal of Environmental Statistics

February 2010, Volume 1, Issue 1. http://www.jenvstat.org

Estimating Oil Spill Occurrence Rates: A Case Study for Outer Continental Shelf Areas of Gulf of Mexico

Ted G. Eschenbach
University of Alaska Anchorage
TGE Consulting

William V. Harper
Mathematical Sciences
Otterbein College

Cheryl M. Anderson
Operations Research Analyst
Offshore Oil & Gas Program of MMS

Richard Prentki
Oceanographer, Environmental Studies
Alaska OCS Region of MMS

Abstract

About 25% of U.S. oil production comes from about 3400 active oil platforms in the Outer Continental Shelf (OCS) areas of the Gulf of Mexico, so estimating the rate of oil spills is important. This article reports on a portion of a study sponsored by the U.S. Minerals Management Service (MMS) that focused on extrapolating from the data on past operations in the Gulf of Mexico to potential future operations in the Alaska OCS areas of the Beaufort and Chukchi Seas. That extrapolation required the development of statistical models for the GOM that advanced current practice. Important differences from published MMS work include exact Poisson confidence intervals, exact binomial confidence intervals, detailed analyses for the exposure variables of pipeline mile-years and platform-years, the use of the larger spill data set of spills exceeding 50 barrels to estimate spill rates at higher thresholds, and the inclusion of more recent data (through 2005). A declining rate of platform spills is statistically verified, so that platform results are generally based on spills 1990 to 2005, while pipeline results are based on data from 1972 to 2005. It is suggested that some of the techniques may be applicable to other problems with a limited number of occurrences spread over decades.

Keywords: Poisson confidence intervals, exposure variables.

1. Introduction

A major source of U.S. oil production is the Outer Continental Shelf (OCS) region of the Gulf of Mexico (GOM). This is done under the oversight of the Minerals Management Service (MMS) within the U.S. Department of the Interior. The database on past oil spills has been used in an extensive history of statistical analysis of oil spill occurrence rates in the area and combined with production in OCS areas of the Pacific, for example see Smith et al. (1982); Lanfear & Amstutz (1983); LaBelle & Anderson (1985); Anderson & LaBelle (1990, 1994, and 2000). Note that there is now production from Northstar from the OCS area north of Alaska.

This article presents key GOM results and methodology from a larger study (Eschenbach and Harper, 2006) focused on estimating spill rates for the Arctic based on that data. As a measure of the breadth of that report it included 52 figures and 76 tables with another 15 figures and 9 tables in an appendix. Because of the need for additional oil and gas resources and the environmental impact of large oil spills when they occur, statistics must use the available data to develop the best models possible. A continued search for improvements in statistical approach is particularly important in light of increased US dependence on OCS oil, corollary decisions by the federal government to open previously closed areas off of Virginia and Alaska to leasing, and recent advances in deepwater production within the Gulf of Mexico.

The focus of this article is the methodology that was used to analyze data that extended over decades, yet only had a relatively small number of data points. In particular when estimating spill rates for spills over 1000 barrels (bbls), there were only 16 pipeline and 8 platform spills in 35 years. The authors conclude that models derived from spills over 50 bbls and spill size distribution models are more effective than models based on smaller numbers of data points. This also allows extrapolation to spill sizes larger than have occurred, where there is no direct evidence. An answer of no model possible is just not good enough.

Current MMS spill occurrence rates are expressed in terms of spills per billion barrels (Bbbl) produced. For example, the rates for US OCS spills of 1000 or more barrels (bbl) per billion barrels (Bbbl) handled based on data from 1964 to 1999 were reported as 1.33 pipeline and 0.32 platform spills in Anderson & LaBelle (2000, p. 311). Like earlier estimates these are reported as point estimates with no confidence limits. For other studies MMS has relied on subsets of this data (pipelines from 1985 to 1999 and platform from 1980 to 1999) for estimated spill rates of 1.38 for pipelines and between 0 and 0.13 for platforms (Anderson & LaBelle, 2000, Table 6).

The following list summarizes the issues and approaches that are addressed in this article.

1. A spill rate is expressed as a number of spills divided by an exposure variable, and it follows a probability distribution. The current MMS exposure variable of production volume (Bbbl) is not as directly linked to the probability of oil spills as the potential exposure variables of pipeline-mile years or platform-years. All can be used to analyze spill rates for the current infrastructure in the GOM, but as will be discussed in Section 3, the infrastructure variables of pipeline-mile years and platform-years are favored for new developments. Which statistical distributions for oil spill rates should be used for modeling and testing of hypotheses? Oil spill occurrence rates are modeled as Poisson distributions, but goodness of fit testing is done with the associated exponential distributions.

2. All spill rate point estimates must include valid confidence intervals. Exact Poisson and binomial intervals were used instead of approximations based on the Normal distribution.

3. Can the larger data base of spills $\geq$50 bbl be used to estimate spill rates at larger thresholds? Yes as shown in Section 5, but this requires integrating results from spill size distributions and binomial proportions with Poisson spill rates. This larger data base supports modeling pipeline spill sizes with a Weibull distribution and platform spill sizes with a lognormal distribution. Note that differences between platforms and pipelines in spill causes and the speed of detection are believed to drive the statistical fits to the different distributions.

4. Regulatory, design, and technology standards have changed over the 1972 - 2005 data period. Are the estimated spill rates stable over time? As described in Section 6, pipeline spill rates are stable over time, but platform spill rates are not.

5. A key issue of the larger study that is not addressed in this article is how to account for spill causes that are specific to the Gulf of Mexico, such as hurricanes and ships dragging anchors or trawling gear, and causes that are specific to the Arctic, such as ice keel gouging and strudel scour. Note: ice keels extend beneath ice floes and sheet ice, just as sails extend above. As the ice moves trenches are gouged in the bottom. Strudel scour is the removal of bottom sediments from spring river overflows draining through holes in the ice. This can uncover a buried pipeline that is subject to thermal expansion stress from the hot oil.

A key difficulty in addressing these issues is that the available data are relatively limited. For example since 1972 there have been 16 pipeline and 8 platform spills over 1000 barrels. It is worth noting that since the key data are based on oil spills, it is good that the number of data points is not larger - even though this makes the statistical analysis more difficult. Platform and pipeline spills were analyzed separately in the original study. For brevity and expository clarity this article sometimes presents the results for only one type of spill.

As this study was conducted over several years, it addressed the robustness of the results as the underlying data was refined and extended by time. As the data records are examined there are significant periods with no spills - especially at larger spill thresholds, and brief periods with several spills. This is intrinsic in the stochastic nature of this problem, but for decision-makers to rely on the study's results; those results should be reasonably stable if analyzed over different periods of time. The results were stable, and interested readers are referred to the original study (Eschenbach and Harper, 2006).

2. Data

Data for this study came from MMS records, and it is detailed in the file posted on the JES site. Data from earlier years was included in a few analyses, but all results reported here are based on spill data and exposure data from 1972 to 2005. As noted by one reviewer of this article, the number of spills ($<$ 1000 bbls) shown in Table 1 declined from 16 (1972 - 1989) to 4 (1990 - 2005). Even with this decline, we could not demonstrate a statistically significant drop over time in the rate of pipeline spills. The rate for platform spills could be shown to have changed, so most study results for platform spills are based on the later period.

MMS maintains a database of "all" spills, but data on spills $\geq$ 1000 bbls have received the most scrutiny and checking. Nevertheless, it was found that using the larger database of spills $\geq$ 50 bbls was statistically desirable. For pipeline spills this expanded the data base from 16 to 36, and for platform spills this expanded the data base from 8 to 78. This data base includes information on location, date reported, fluid spilled, spill volume, primary spill cause, secondary spill cause, etc. Table 1 summarizes some of the key elements of the pipeline spill data.

There is one point to be noted and two caveats with this data. This data set contains multiple values estimated as round numbers such as 50 or 100 barrels, which were adjusted to 50, 51, ... so that each spill volume was unique for statistical testing. This adjustment has no real impact on the magnitude of the data, but it supports distributions and tests which are based on the assumption of continuous data. For this study this includes the Anderson-Darling goodness of fit test, the probability plot, and the empirical cumulative distribution fit. This common practice aids both the visual assessment of fit as well as the associated statistical properties of the methods.

There are two caveats with respect to the spills due to hurricanes. First, in several of these cases, the spill volume is the total of multiple oil spills, where they have been aggregated in keeping with historical MMS practice due to their common cause. The second caveat is that the data for hurricanes Ivan (2004), Katrina (August 2005), and Rita (September 2005) differ from the data reported in earlier years for other hurricanes. Although losses of fluids stored on destroyed and missing structures were not always accounted for by MMS in earlier years, for these three recent spills MMS has asked for the reporting of fluids presumed lost with destroyed platforms and the linked damage to pipelines. Thus, there are more pipeline spills counted as part of each hurricane. Specifically, the Ivan total comes from 8 spills from 95 to 1720 bbl, the Katrina total comes from 6 spills, and the Rita total from 8 spills. Recent updates to these hurricane spill volumes on the MMS web site have the same total volume for Ivan, 1247 bbl for Katrina, and 4474 bbl for Rita. Hurricanes Ivan, Katrina and Rita represent exceptional losses of oil and gas structures (and the products stored on them). For example, Katrina destroyed or extensively damaged about 65 platforms and 22 drilling rigs, and for Rita there were 103 platforms and 33 drilling rigs destroyed or extensively damaged. Hurricanes Rita, Katrina, and Ivan were the 4th, 6th, and 9th most intense Atlantic hurricanes in recorded history respectively (based on lowest recorded barometric pressures). Having three major storms passing through significant portions of the OCS oil and gas infrastructure at peak intensity all within a 12-month period is uncommon. Roughly 75% of all OCS oil and gas facilities in the Gulf of Mexico were exposed to at least one of the three storms.

The primary cause of each pipeline spill is included in Table 1 to provide an overview of the wide variety of causes. While not part of this article, GOM-specific causes (such as hurricanes, fishing trawlers, and freighter anchors for pipelines and such as hurricanes for platforms) had to be specifically addressed in the original study‘s extrapolation to the Arctic. The 78 platform spills are summarized in the data posted on the JES website. These spills do not have a primary spill cause for two reasons. First, in several cases it is simply not possible to identify a primary cause. As a simple example, when a spill occurs during a diesel transfer from a coupling that breaks or separates is it a mechanical failure or human error because the transfer was unattended? Second, unlike pipelines the only GOM-specific cause is hurricanes,

Spill Date	Spill Size (bbl)	Cause
1972-06-13	100	Corrosion
1973-05-12	5,000	Corrosion
1974-04-17	19,833	Anchor drag
1974-05-21	65	Operational: anchor from derrick barge
1974-09-11	3,500	Hurricane Carmen
1976-02-29	414	Corrosion: after pipeline kinked by anchor
1976-12-18	4,000	Shrimp trawl damaged valve
1977-03-29	250	Natural: mud slide
1977-06-05	50	Operational: lay barge anchor
1977-10-18	300	Anchor drag
1978-04-08	135	Mechanical/operational
1978-07-17	900	Operational: anchor drag 600' from platform
1979-07-15	50	Operational: workboat searching for rig anchor
1980-01-29	100	Trawler drag broke valve
1981-08-05	80	Corrosion: external or metal fatigue
1981-12-11	5,100	Operational: service vessel anchor
1983-01-20	80	Natural: mud slide
1985-02-16	323	Operational: pipeline dented during construction
1985-11-09	50	Operational: spud barge anchor
1986-02-03	119	Mechanical/operational: pinhole leak during shut-down
1986-12-30	210	Mechanical/operational: anchor or original construction
1988-02-07	15,576	Ship illegally dropped and dragged anchor
1990-01-24	14,423	Fishing net or anchor
1990-05-06	4,569	Trawler net drag
1992-01-03	190	Unknown
1992-08-31	2,000	Rig broke loose during Hurricane Andrew
1993-06-17	50	Operational: workboat anchor
1994-11-16	4,533	Trawl net damaged valve
1998-01-22	800	Mechanical damage: probably anchor
1998-01-26	1,211	Operational: anchor during overboard rescue
1998-09-29	8,212	Hurricane Georges
1999-07-23	3,200	Operational: jackup rig set down on pipeline
2000-01-21	2,240	Operational: anchor drag from drill rig
2004-09-15	3,445	Hurricane Ivan
2005-08-29	553	Hurricane Katrina
2005-09-24	8,162	Hurricane Rita

Table 1: Gulf of Mexico OCS Pipeline Oil Spills (N = 36)

thus there was no reason in the original study to further distinguish between the causes of platform spills.

The exposure variables shown in Table 2 were reported on an annual basis. These included the volume of crude oil produced, the number of platforms, the number of pipeline segments, and pipeline-miles for various product codes. These exposure variables were chosen because they can logically be linked to the probability of an oil spill and because the data were available. Oil production was the exposure variable used in earlier MMS studies.

The data on exposure variables within the GOM are highly correlated. For example the correlation between pipeline mileage and each of the other variables is 99% for time, 91% for number of pipeline segments, 93% for number of platforms, and 78% for GOM oil production. One implication of this collinearity is that multiple exposure variables cannot reliably be used together for estimating. As pointed out by two reviewers, using differencing, logarithms of ratios, or loglinear models are other ways to tackle the estimation. This represents an area for further research and we appreciate the comments. The approach taken in the publication provides exact confidence intervals and does not rely on asymptotic properties of maximum likelihood estimates. The sample sizes of the spill data are not large.

MMS has made substantial efforts to ensure that this data is as “clean” as possible, and the scope of the statistical study that underlies this article focused on use of the data not further validation of it. For purposes of estimating oil spill rates, it does not really matter if declines in mileage, segments, or platforms are due to a shift in status to inactive, intentional dismantling, or destruction. The exposure variables represent what was in use for a particular year.

3. Choosing the Best Exposure Variables and Spill Rate Distributions

Historical MMS practice used the volume of oil production as the exposure variable, so that the rate for US OCS pipeline spills was reported as 1.33 spills of 1000 or more barrels (bbl) per billion barrels (Bbbl) handled in Anderson & LaBelle (2000, p. 311). This exposure variable has the advantage that it does not require a detailed development scenario to apply it to a hypothetical OCS development of an undiscovered resource. Only the volume of oil to be produced is needed to estimate the number of oil spills. As noted by a reviewer, it also directly links how much “good/oil” we get in return for how much “bad/spills” that we have to endure. Obviously, such an extrapolation assumes that the potential new development is enough like the existing GOM infrastructure and operating environment that the extrapolation is valid. The weakness of this exposure variable is that the extrapolation may not be as accurate for newer OCS developments in the GOM that are in deeper waters and further from shore.

While volume of oil produced may not be the best exposure variable for extension to the Arctic, it was analyzed extensively for the Gulf of Mexico analysis because it is the measure that has historically been used by MMS. Many of the results in this article are for spill rates with the oil production exposure variable.

The probability of an oil spill from the number of platforms and the network of pipelines clearly depends in some fashion on the number of platforms and the length of pipeline. The more that is there, the more likely it is that an anchor from a freighter, corrosion, mud-slides, etc. will cause a spill. Similarly, the probability of an oil spill is linked in some fashion to the number of segments of pipeline between joins and terminations. It is this logic and the

				Oil Production (Million bbl)	
Year	Pipe Segments	Miles	Total of Platforms	Gulf of Mexico	Total OCS
1972	1249	1740	1623	373	396
1973	1312	1932	1691	366	385
1974	1344	2049	1749	338	355
1975	1389	2200	1798	310	325
1976	1442	2451	1869	301	315
1977	1497	2563	1956	284	296
1978	1568	2818	2075	276	288
1979	1643	2956	2179	318	334
1980	1695	3120	2296	265	275
1981	1766	3343	2411	263	283
1982	1837	3511	2562	286	315
1983	1907	3703	2735	320	351
1984	2012	3925	2854	355	385
1985	2095	4063	3009	350	380
1986	2134	4238	3052	356	384
1987	2173	4345	3159	328	359
1988	2213	4451	3307	301	333
1989	2250	4562	3391	291	324
1990	2294	4737	3444	275	304
1991	2345	4836	3456	295	326
1992	2362	4979	3434	305	348
1993	2395	5030	3478	309	359
1994	2436	5287	3466	314	372
1995	2462	5536	3513	345	417
1996	2501	6148	3557	369	433
1997	2531	6433	3601	412	466
1998	2576	6753	3551	444	491
1999	2578	6996	3585	495	534
2000	2584	7247	3595	523	559
2001	2594	7466	3575	558	592
2002	2578	7651	3573	567	602
2003	2463	7842	3578	561	595
2004	2466	8516	3530	535	567
2005	2388	8369	3435	459	488
sum 72-05		161,796	100,087	12,445	13,535
sum 90-05		103,826	56,371	6,766	7,454

Table 2: Exposure Variables for Gulf of Mexico OCS Oil Spills

availability of data on these infrastructure variables, that drove the choice of which exposure variables to consider. In addition, those professionals with the most knowledge of the data believed that the mileage data were much cleaner than the data on segments which represented a mix of physical, geographical, and labeling definitions. Thus the segment data was only used in checks of correlation and data consistency.

There are also spill causes such as hurricanes which might be best modeled using time as an exposure variable. Because of the high correlations between the exposure variables in the GOM, consistent results for the GOM on an aggregate basis are obtained with any of the exposure variables. However, for estimating the spill rate for a new development that is different from the average GOM production facility, it is clearly better to use an exposure variable that is linked to the amount of infrastructure needed for the potential development.

In this case all exposure variables were satisfactory independent variables for modeling of the number of oil spills as Poisson distributions (Eschenbach and Harper, 2006, Chapter 3) or the equivalent exponential inter-arrival distribution. While the fit of the Poisson distributions could have been directly tested, this would have required some level of binning for the discrete data. And with 36 pipeline spills over 34 years many bins would have only a few data points. Thus, the testing of distributions was done for inter-spill intervals using the exponential distribution. For pipeline spills the infrastructure variable became thousands of pipeline mile-years, which was computed using linear interpolation for the number of days in between spills. Thus, each interval between spill was converted from days to thousands of pipeline mile-years. Production volume was tested similarly. Again, use of all of exposure variables as independent variables passed a goodness of fit test for an exponential distribution of the inter-spill interval.

When translated into estimating the spill rate for the Poisson distribution the exposure variable becomes the total over the modeled interval. These values were shown at the bottom of Table 2. For the logic and availability reasons detailed in this section, the chosen infrastructure variable for pipeline spills is the number of thousands of miles of pipeline in active use. The infrastructure variable for platform spills is the number of thousands of platforms in active use. If there were a much larger number of spills, it might be possible to consider other factors, such as the size of a platform or the number of wells or pipelines linked to a platform, but fortunately the number of spills is too small for this.

4. Estimates of Oil Spill Rates

4.1. Exact Versus Approximate Confidence Intervals for Poisson Spill Rates

Historical spill rate estimates by MMS were point estimates without confidence intervals. After Givens (2002) and Zeh (2002) critiqued the lack of confidence intervals, MMS updated their website (http://www.mms.gov/eppd/sciences/osmp/pdfs/ConfidenceIntervals2.pdf accessed November 9, 2009) with confidence intervals based on counting the number of spills in a bin of a fixed production amount. For example, the first estimate in Table 3 used 12 bins of a billion barrels of production from 1964 to 1999. This approach assumes independent observations, and with small sample statistics typically assumes that the data are from a normal distribution. Thus one task of the study was to provide a better approach to the calculation of confidence intervals for estimated spill rates.

Equation 1 is modified from Johnson and Kotz (1969, pp. 96) by using λ (the more common Poisson mean notation) instead of the θ used in Johnson and Kotz. The confidence interval (λ_L, λ_U) places $\alpha/2$ in each tail.

$$\lambda_L = \frac{\frac{1}{2}x^2_{2x,\frac{\alpha}{2}}}{\sum ExposureVariable};\ \lambda_U = \frac{\frac{1}{2}x^2_{2(x+1),1-\frac{\alpha}{2}}}{\sum ExposureVariable} \qquad (1)$$

Generally in statistics increasing the sample size decreases the width of confidence intervals. In this case, the subscript x is the number of spills, and the basis for the number of degrees of freedom for the chi-square variable. Then the amount of the exposure variable is in the denominators of Equation 1. To better understand this formula, assume that the number of spills and the amount of exposure are both doubled (which keeps the spill rate constant). Doubling the number of degrees of freedom more than doubles the lower chi-square value. Since the exposure also doubled, the lower limit goes up and is closer to the estimated average. In like fashion doubling the number of degrees of freedom less than doubles the upper value, so the upper limit has decreased. Thus, a larger sample at the same spill rate has a tighter or narrower confidence interval. As shown in Table 3, these exact intervals are generally tighter than the inaccurate intervals based on dividing the exposure variable into artificial subunits.

OCS Pipelines	# Spills	Volume (Bbbl)	Rate (Spills/Bbbl)	MMS Confidence Interval	Exact 95% Poisson Confidence Interval
1964-1999	16	12.00	1.333	(0.54, 2.12)	(0.76, 2.17)
1985-1999	8	5.81	1.377	(0.00, 2.77)	(0.59, 2.71)

Table 3: 95% Confidence Intervals for Spills $\geq$ 1000 bbl Based on Anderson & LaBelle (2000) Pipeline Data with Production Volume as the Exposure Variable

4.2. Estimated Spill Rates for Different Exposure Variables

The principal results of the study with respect to spill rates in the Gulf of Mexico are summarized in Table 4. The 50 bbl spill size used for Table 4 is much smaller than the 1000 bbl size used in Table 3, thus the spill rates are about twice as high. The tabulated lower and upper confidence limits are for 95% confidence intervals, calculated using Equation 1.

It is worth noting that if the spill rates are estimated using the exponential distribution and the inter-spill data, then slightly different results are generated because the data for the Poisson includes the time before the first spill and after the last spill. If the Poisson data is restricted to the period between the first and last spills, then the Poisson rate and the exponential rate are for the same data and they match.

The results for platform spill rates are reported for two time periods, because as will be shown in Section 6 of this article, the spill rate for platforms is not homogeneous over time.

Pipelines 1972-2005	# Spills	Exposure Variable	Sum Exposure Vairable	Rate	LCL	UCL
Spills/1000 miles-years	36	KMile-years	161.8	0.22	0.16	0.31
Spills/Bbbl	36	Bbbl Production	13.5	2.66	1.86	3.68
Spills/year	36	Times, years	34	1.06	0.74	1.47
Platforms 1972-1989						
Spills/1000 platform-years	56	KPlatform-years	43.7	1.28	0.97	1.66
Spills/Bbbl	56	Production Bbbl	5.68	9.86	7.45	12.8
Spills/year	56	Times, years	18	3.11	2.35	4.04
Platforms 1990-2005						
Spills/1000 platform-years	22	KPlatform-years	56.4	0.39	0.25	0.59
Spills/Bbbl	22	Production Bbbl	6.77	3.25	2.04	4.92
Spills/year	22	Times, years	16	1.38	0.86	2.08

Table 4: Spill Rates and 95% Confidence Intervals (Spills $\geq$ 50 bbls) for Gulf of Mexico

5. Estimating Larger Oil Spills from Data Base of Smaller Spills

5.1. Overview of Approach

The simplest approach to estimating the spill rate for larger spills, such as 500, 1000, or more barrels is to count the number of that size and directly compute the spill rate and associated confidence interval. For example, from 1972 to 2005 there were 16 pipeline and 8 platform spills over 1000 barrels in the OCS area of the Gulf of Mexico (the 1000 barrel threshold is the most frequently reported spill rate for MMS publications). In Eschenbach and Harper (2006) spill rates are reported for this approach at spill size thresholds of 50, 100, 500, and 1000 barrels for each of the exposure variables discussed earlier. These thresholds were chosen because they represent a range of spills that are not uncommon.

Reporting on spill rates at thresholds of 100, 500, and 1000 barrels allowed the comparison of the direct approach of counting spills above each threshold with alternate approaches that relied on the larger data base of spills over 50 barrels. The alternate approaches began with (1) estimating the spill size distribution using the database of spills $\geq$ 50 barrels. (2) Then this distribution was used to estimate the fraction of the spills that exceed each of the higher thresholds. (3) Those fractions times the spill rate for spills $\geq$ 50 barrels became the point estimates of the spill rates for the higher thresholds. These first three steps are described in Section 5.2. (4) The exact binomial confidence intervals were then calculated for the fraction of the spill distribution exceeding each threshold (see Section 5.3). (5) Finally, appropriate confidence intervals were calculated for the spill rate at each threshold (see Section 5.4).

5.2. Point Estimates for Spill Rates Using Distribution for Spill Sizes

For pipeline spill sizes at all four spill size thresholds a three-parameter Weibull distribution was selected as the “best-fit” using maximum likelihood estimation. Each distribution was not rejected based on an Anderson-Darling goodness of fit test. Each passed visual inspections of the probability plot and fit to the empirical CDF. The parameters for the size distribution for spills $\geq$ 50 barrels were shape = 0.440, scale = 1391, and threshold = 50 (fitted). With this distribution the conditional probabilities of exceeding 100, 500, and 1000 barrels are respectively 0.793, 0.544, and 0.429 (see Eq. 2 for example calculation).

$$1 - F(1000) = 1 - \left[1 - exp\left(-(\frac{x-\nu}{\alpha})^{\beta}\right)\right] = exp\left[-\left(\frac{1000-50}{1391}\right)^{0.4398}\right] = 0.429 \qquad (2)$$

These values are quite close to the proportions in the data which are 0.806, 0.528, and 0.444 respectively. The consistency of the distributions was checked by overlaying the conditional distributions and by calculating odds ratios for different spill size. For example what are the odds of exceeding 1000 barrels versus exceeding 2000 barrels. By this measure the distributions based on thresholds of 500 and 1000 barrels and their smaller number of spills were less reliable. This suggests that the estimates based on counting the number of spills exceeding 500 or 1000 barrels may also be less reliable.

Thus, estimated pipeline spill rates for 100, 500, and 1000 barrel thresholds (2.16 spills/Bbbl, 1.48, and 1.17 respectively) were derived by multiplying the spill rate for $\geq$ 50 barrels by the conditional probabilities above. Because the report focused on extrapolating from GOM statistics to the Arctic, the analysis of GOM data focused on the exposure variable of oil production traditionally used by MMS (spills/Bbbl).

A similar approach was used to analyze platform spills, but as detailed in Section 6, this data is not homogeneous over time - the platform spill rate has declined. So the spill rate calculations were done using data from 1990 to 2005. However, the platform spill size distribution could not be shown to be changing over time, so all 78 spills from 1972 to 2005 were used to fit the platform spill size distribution. In this case a three-parameter lognormal distribution with a location parameter of 4.27, a scale parameter of 2.08, and a threshold of 50.0 was the best fit for spills $\geq$ 50 barrels. With this distribution the conditional probabilities of exceeding 100, 500, and 1000 barrels are respectively 0.57, 0.19, and 0.11. Note that these are much smaller than for pipelines, because many platform spills are linked to the transfer or use of refined petroleum products such as diesel fuel. These spills simply cannot be as large as pipeline spills. Because platform spills are smaller, the methodology was consistent, and for brevity's sake the platform spills are not discussed further in Section 5.

5.3. Exact Clopper-Pearson Binomial Confidence Intervals for Proportion of Pipeline Spills Exceeding Threshold

The exact Clopper-Pearson (1934) binomial confidence limits shown in Equation 3 were implemented as Excel functions available from the second author (Harper, 2005) and used to calculate lower and upper confidence limits for the proportion of spills exceeding higher spill thresholds. (They were also used in developing spill rate estimates that could be extended from the GOM to the Arctic.) Unlike the approximate normal based intervals, exact binomial confidence intervals are generally asymmetric. The difference between exact and Normal based approximations to the binomial are most significant for small sample sizes and proportions that are close to 0 or 1. This technique could be extended to calculate the spill rate for 10,000 or 100,000 barrel oil spills (see Section 5.5).

$$p_L^{\alpha}(n,B) = \frac{B}{B+(n-B+1)f_{\frac{\alpha}{2},2(n-B+1),2B}};\ p_U^{\alpha}(n,B) = 1 - p_L^{\alpha}(n,n-B) \qquad (3)$$

where B is the number of successes in the n Bernoulli trials f_{γ,n_1,n_2} and is the upper γ percentile of the F distribution with n_1 and n_2 degrees of freedom.

More formally, the Weibull based model for pipeline spill volumes for $\geq$ 50 bbl is used to estimate the conditional probability of exceeding larger thresholds. Then assuming this probability represents a binomial proportion for the N = 36 observations in the $\geq$ 50 bbl data set, exact Clopper-Pearson binomial confidence intervals are calculated (see Table 5). Table 5 was constructed using production volume rather than pipeline-mile-years as the exposure variable.

Threshold	P($\geq$ Threshold)	Lower Binomial Confidence Limit	Upper Binomial Confidence Limit	Actual Proportion
100	0.793	0.640	0.918	0.806 = 29/36
500	0.544	0.381	0.721	0.528 = 19/36
1000	0.429	0.255	0.592	0.444 = 16/36

Table 5: Exact Clopper-Pearson 95% Binomial Confidence Intervals for the Proportion of the $\geq$ 50 bbl Model Pipeline Spills Exceeding the Larger Threshold with Production as Exposure Variable

5.4. Confidence Intervals for Spill Rates

In building confidence intervals for spill sizes with larger thresholds from the data on pipeline spills over 50 bbl, the study used two "lower bound" approaches and one "upper bound" approach for pipeline spills. We would also like to thank a reviewer for another suggested approach, that we will explore if possible. One input for these calculations is the spill rate for a threshold of 50 bbl. This value is 2.72 pipeline spills/Bbbl with its associated 95% confidence interval of (1.91, 3.77). The second input is the conditional probabilities of exceeding 100, 500, and 1000 barrels and the 95% confidence limits given in Table 5.

The two lower bound (in the sense of width of the confidence interval) approaches consider one of the two inputs as fixed and the other as defined by its confidence limits. The upper bound or wide confidence interval equals (two lower limits multiplied, two upper limits multiplied). The approaches are illustrated in Figure 1 for a spill threshold of 500 barrels. See Eschenbach and Harper (2006) for the complete tabulated results. Note that Figure 1 and Section 5.5 were developed for this article, and were not part of the cited study (Eschenbach and Harper, 2006). More precisely the values for each approach can be defined as follows:

- LB_1: assumes binomial fraction is known and uses confidence interval for rate.
  For 500 bbl threshold: $0.544 \times (1.91, 3.77)$

- LB_2: assumes 50 bbl spill rate is known and uses confidence interval for binomial.
  For 500 bbl threshold: $2.72 \times (0.381, 0.721)$

- UB: assumes both lower limits or both upper limits define limits.
  For 500 bbl threshold: $(1.91 \times 0.381, 3.77 \times 0.72)$

For comparison purposes, Figure 1 also includes the value and 95% confidence intervals for the exact Poisson intervals defined by the 36 spills over the 50 bbl threshold and the 19 spills over the 500 bbl threshold. Each lower bound assumes one source of uncertainty is known, and the upper bound assumes the extremes of each occur simultaneously. The confidence limits for higher threshold spills based on 50 bbl spill data should lie in between these limits. Other approaches for deriving these limits represent possible future research.

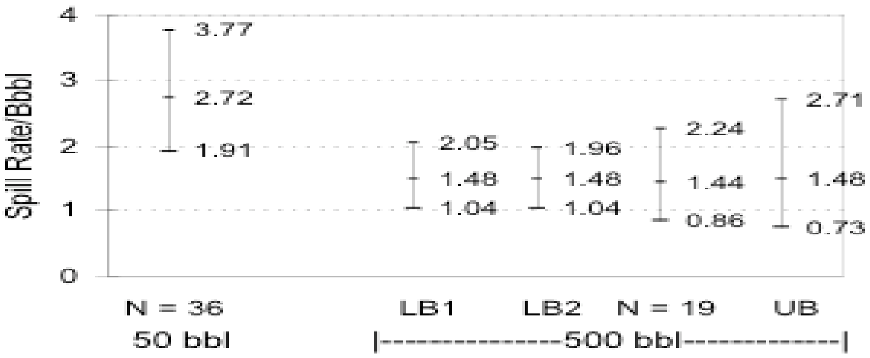

Figure 1: Estimated Confidence Limits for Spills ≥ 500 bbl

5.5. Confidence Intervals for Spill Rates above Largest Observed

In Eschenbach and Harper (2006) it was noted that the approach summarized in Sections 5.2 - 5.5 could be extended to spill thresholds above the largest observed (but it did not do so). This section reports on that extension. Pipeline spills above a 20,000 bbl threshold and platform spills above a 10,000 bbl threshold did not occur in the 1972 - 2005 time interval. If the data is used directly, then the point estimates above the 20,000 bbl threshold for pipelines and above the 10,000 bbl threshold for platforms are unrealistically equal to 0. However, larger spills can clearly occur as demonstrated with the offshore spill in Australia in 2009, since platforms and pipelines are connected to oil wells rather than the fixed quantities in a ship or a tank farm. If the time period is expanded to include the late 1960s, then in 1967 there was a pipeline spill of 160,638 barrels and in 1969 the Santa Barbara platform spill of 80,000 barrels occurred (OCS but not GOM). In 1970 there were two other platform spills of 30,000 and 53,000 barrels in the GOM. While stronger regulation and improved technology has reduced the risk of these large oil spills, that risk has not been eliminated.

Before these alternate approaches can be extended to larger spill sizes of 10,000 or 100,000 barrels, it is necessary to understand and validate their performance at lower thresholds. Table 6 and Figure 2 explore the relationships between the binomial confidence limits based on the empirical data (solid lines in Figure 2) and the binomial confidence limits based on the proportion from the fitted Weibull probability distribution (dashed lines in Figure 2). It should not be surprising that the point estimates and the limits based on the data do not form smooth lines, because of the randomness of any real data set. The distribution-based point estimate and limits shown in Figure 2 are constructed by connecting the points tabulated in Table 6.

The first step in calculating the distribution based values is to calculate the point estimate for the proportion of spills exceeding each threshold value from the fitted 3-parameter Weibull distribution (shape = 0.44, scale = 1391, and threshold = 50). Then the sample size of 36 is multiplied by the proportion and rounded to the nearest integer, which is then used to calculate the lower and upper limits. When this integer matches the empirical data the two sets of limits are the same (shaded values in Table 6). When the empirical data has a different integer number of spills, the results are different. The point estimates do not match because of the rounding to the nearest integer. As pointed out by a reviewer, the distribution based limits would follow the same curve if integer values for the number of spills and the inverse Weibull were used to calculate the exact threshold - but then we could not compare the two approaches at the same thresholds.

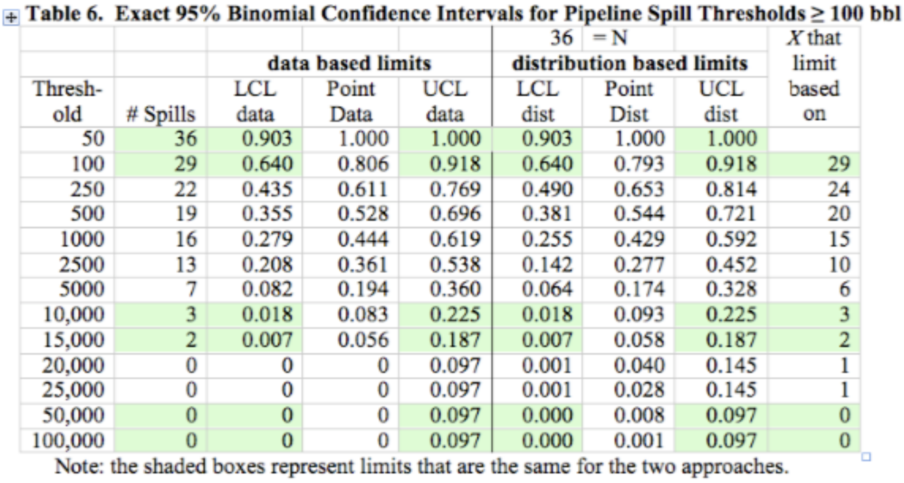

Table 6. Exact 95% Binomial Confidence Intervals for Pipeline Spill Thresholds ≥ 100 bbl

					36	= N		*X* that limit based on
		data based limits			**distribution based limits**			
Thresh-old	# Spills	LCL data	Point Data	UCL data	LCL dist	Point Dist	UCL dist	
50	36	0.903	1.000	1.000	0.903	1.000	1.000	
100	29	0.640	0.806	0.918	0.640	0.793	0.918	29
250	22	0.435	0.611	0.769	0.490	0.653	0.814	24
500	19	0.355	0.528	0.696	0.381	0.544	0.721	20
1000	16	0.279	0.444	0.619	0.255	0.429	0.592	15
2500	13	0.208	0.361	0.538	0.142	0.277	0.452	10
5000	7	0.082	0.194	0.360	0.064	0.174	0.328	6
10,000	3	0.018	0.083	0.225	0.018	0.093	0.225	3
15,000	2	0.007	0.056	0.187	0.007	0.058	0.187	2
20,000	0	0	0	0.097	0.001	0.040	0.145	1
25,000	0	0	0	0.097	0.001	0.028	0.145	1
50,000	0	0	0	0.097	0.000	0.008	0.097	0
100,000	0	0	0	0.097	0.000	0.001	0.097	0

Note: the shaded boxes represent limits that are the same for the two approaches.

Note that the point estimates and confidence limits decline approximately linearly when spill thresholds are graphed on a logarithmic scale in Figure 2. The results are consistent so it is reasonable to extend results to somewhat higher values. However, once the estimated number of spills is less than 0.5 rounded to 0 spills over the time period, then the upper confidence limit bottoms out at about 0.1 spills per year, even as the point estimate continues to decline towards 0.

These results can be combined with the earlier estimates for spill rates of 50 bbl to produce LB1, LB2, and UB bounds on the confidence limits for spills above the threshold. For larger spill thresholds such as above 2500 bbl where the fitted distribution predicts 10 or fewer spills, it is suggested that the fitted distribution is likely to provide more reliable results than the empirical distribution with its small sample data for larger spill sizes.

While the values for each approach can be calculated, the upper confidence limits for the binomial proportion are about 2 to 3 orders of magnitude larger than the point estimates for 25,000 and 100,000 bbl oil spills. Thus, the LB2 and UB values calculated next seem likely to be overly conservative. The complete results for these representative large spill thresholds are shown in Table 6.

- LB_1: assumes binomial fraction is known and uses confidence interval for rate

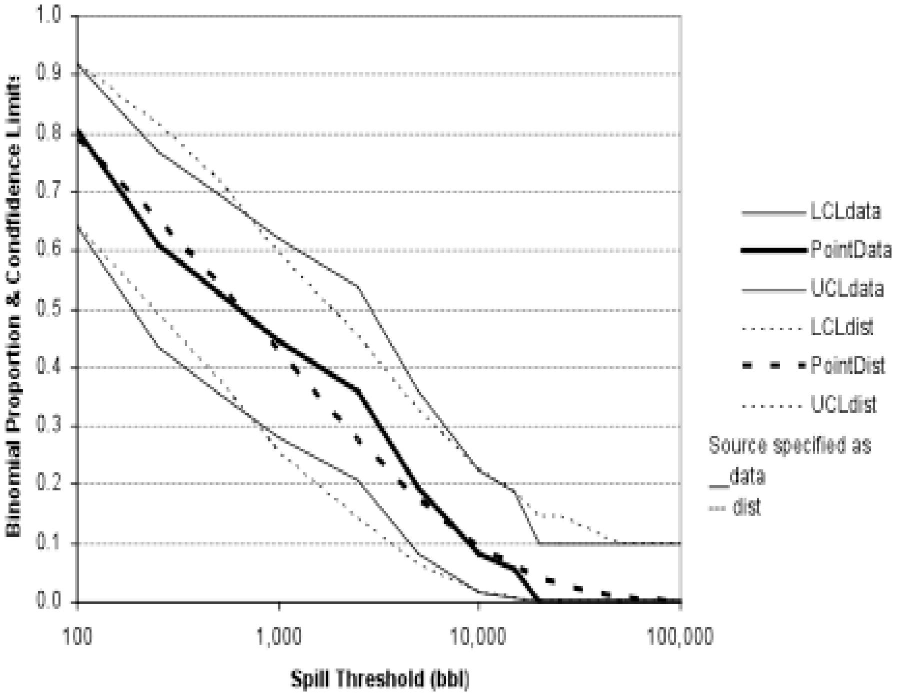

Figure 2: Estimated 95% Binomial Confidence Limits for Pipeline Spill Thresholds ≥ 100 bbl

For 25,000 bbl threshold: $0.0284 \times (1.91, 3.77)$
For 100,000 bbl threshold: $0.00142 \times (1.91, 3.77)$

- LB_2: assumes 50 bbl spill rate is known and uses confidence interval for binomial
  For 25,000 bbl threshold: $2.72 \times (0.000703, 0.145)$
  For 100,000 bbl threshold: $2.72 \times (0, 0.0973)$

- UB: assumes both lower limits or both upper limits define limits
  For 25,000 bbl threshold: $(1.91 \times 0.000703, 3.77 \times 0.145)$
  For 100,000 bbl threshold: $(1.91 \times 0, 3.77 \times 0.0973)$

6. Are Spill Rates Constant Over Time?

Tests of the spill rates for homogeneity over time were extensive as it was hoped that spill rates would be declining for all spill measures, and especially for those more directly linked to exposure variables of pipeline-mile years and platform-years. These tests were not significant for the pipeline spills, but the platform spill rate was shown to be declining.

When the number of days between platform spills was regressed against spill year the R^2 was only 8.7%, but the slope was statistically significant (p-value of 0.009). The average increase in days between spill was 5.77 days/year. Confidence limits for future average inter-spill days were included in the report (for example 185 to 409 days by 2010), but the high variability in

		LCL	Point Estimate	UCL
25,000 bbl	LB$_1$	0.054	0.077	0.11
	LB$_2$	0.0019	0.077	0.40
	UB	0.0013	0.0.77	0.55
100,000 bbl	LB$_1$	0.0027	0.0039	0.0053
	LB$_2$	0	0.0039	0.26
	UB	0	0.0039	0.37

Table 6: Estimated Lower & Upper Bounds on Pipeline Spill Rate Estimates for 25,000 & 100,000 bbl Spill Size Thresholds

the data resulted in prediction limits that were so wide as to be of little value (0 to 650 days by 2010).

For consistency with past MMS practice and the time block approach of 1972-2007, the platform spill data was examined to choose a “good” dividing point. It appeared that dividing the data into 1972-1989 and 1990-2005 would be appropriate. Figure 3 summarizes the analysis of means done for platform spills. The centerline shows an overall average over the full 1972 - 2005 time frame of 155.1 inter-spill days. The “stair step” lines represent upper or lower bounds around this grand average. The interval is tighter for the earlier data (N = 56) vs. (N = 21) for the later data. Since both averages fall outside the bounds, both are significantly different from the overall grand average. Hence the spill rate is non-stationary.

Similar statistical result may be seen in the two sample t-test (p-value of 0.012) comparing the two means and the nonparametric Kruskal-Wallis test (p-value of 0.001) comparing the two corresponding medians. The average and median inter-spill days have about doubled and tripled respectively for the later time interval. The Poisson rates for the two time periods were also shown to be significantly different using the test for ratio between two Poisson rates (Chapman, 1952).

7. Summary and Conclusions

This article has summarized key results and the methodology for estimating oil spill rates in the Gulf of Mexico from a larger study of extrapolation from that base to the Beaufort and Chukchi Seas. This has included exact Poisson confidence intervals, exact binomial confidence intervals, results for the exposure variables of pipeline mile-years and platform-years, and the inclusion of more recent data (through 2005). A declining rate of platform spills is statistically verified, so that platform results are generally based on spills 1990 to 2005, while pipeline results are based on data from 1972 to 2005.

In order to deal with a limited number of spills (especially larger ones) spread over decades, the larger spill data set of spills exceeding 50 barrels was used to estimate spill rates at higher thresholds. It is suggested that this approach may be useful in other problems where estimates must be made for rare large events where smaller events are more common.

This article has advanced the original study by applying this methodology to pipeline spill size thresholds of 25,000 and 100,000 bbls which are above those observed values in the analyzed

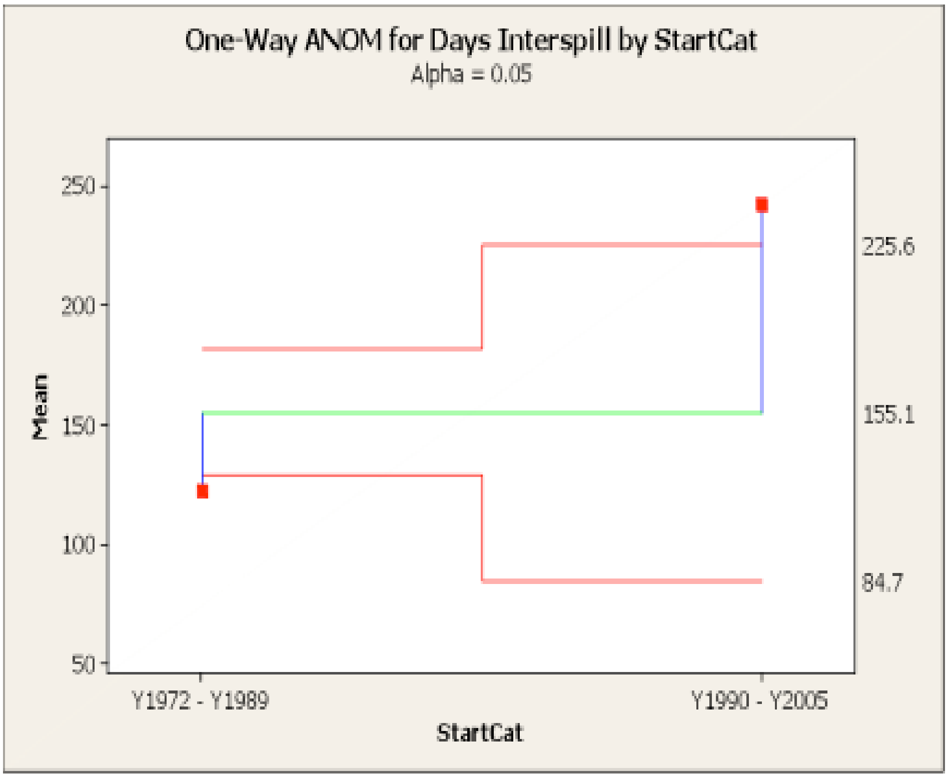

Figure 3: Analysis of Means Comparing Inter-Spill Averages between Data for 1972-1989 versus 1990-2005 for Platform Spills ≥ 50 bbl

data. It is suggested that this approach may be useful in other problems where estimates must be made for rare events that are larger than represented in the time period being analyzed.

8. References

Anderson, C. M., and LaBelle, R. P. (1990), “Estimated Occurrence Rates for Analysis of Accidental Oil Spills on the US Outer Continental Shelf,” *Oil and Chemical Pollution*, Vol. 6, 21-35.

–, (December 1994), “Comparative Occurrence Rates for Offshore Oil Spills,” *Spill Science & Technology Bulletin*, Vol. 1, No. 2, 131-141.

–, (2000), “Update of Comparative Occurrence Rates for Offshore Oil Spills,” *Spill Science & Technology Bulletin*, Vol. 6, No. 5/6, 303-321.

Chapman, D.G. (1952), “On Tests and Estimates for the Ratio of Poisson Means,” *Annual Institute of Statistical Mathematics*, Vol. 4, 45-49.

Clopper, C. J., and Pearson, E.S. (1934), “The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial,” *Biometrika*, 26, 404-413.

Eschenbach, T. G., and Harper, W. V. (2006), Alternative Oil Spill Occurrence Estimators for the Beaufort/Chukchi Sea OCS (Statistical Approach), OCS Study MMS 2006-059, http://www.mms.gov/alaska/reports/2006rpts/akpubs06.HTM (accessed 11/14/09).

Givens, G. H. (December 2002), “Review of Update of Comparative Occurrence Rates for Offshore Oil Spills,” North Slope Borough Science Advisory Committee.

Harper, W. V. (July 2005), “Excel Functions to Compute Exact Binomial Confidence Intervals,” *Proceedings of the 25th European Meeting of Statisticians*, Oslo, Norway, CD.

Johnson, N. L., and Kotz, S. (1969), *Discrete Distributions*, John Wiley & Sons.

LaBelle, R. P., and Anderson, C. M. (1985), The Application of Oceanography to Oil-Spill Modeling for the Outer Continental Shelf Oil and Gas Leasing Program, *Marine Technology Society Journal*, Vol. 19, No. 2, 19-26.

Lanfear, K. J., and Amstutz, D. E. (1983); A Reexamination of Occurrence Rates for Accidental Oil Spills on the U.S. Outer Continental Shelf, *Proceedings of 1983 Oil Spill Conference*, American Petroleum Institute, Washington D.C., 355-359.

Smith, R.A., Slack, J.R., Wyant, T., and Lanfear, K.J. (1982), “The Oil Spill Risk Analysis Model of the US Geological Survey,” USGS Professional Paper 1227, US Geological Survey.

Zeh, J. E. (December 2002), “Review of Oil Spill Risk Analyses for Alaskan Arctic OCS Development,” North Slope Borough Science Advisory Committee.

Affiliation:
Ted Eschenbach
TGE Consulting
E-mail: tede1@ak.net
4376 Rendezvous Circle
Anchorage, AK 99504
Telephone: 907-333-7817
Fax: 337-2928

William V Harper
Mathematical Sciences
Otterbein College
Towers Hall 139, 1 Grove St
Westerville OH 43081-2006 USA
Telephone: 614-823-1417
Fax: 614-823-3201
E-mail: wharper@otterbein.edu

Cheryl M. Anderson
Operations Research Analyst
Offshore Oil Gas
Program of MMS
E-mail: cheryl.anderson@mms.gov

Richard Prentki
Oceanographer, Environmental Studies
Section, Alaska OCS Region of MMS
E-mail: richard.prentki@mms.gov

Journal of Environmental Statistics
Volume 1, Issue 1
February 2010
http://www.jenvstat.org
Submitted: 2009-03-10
Accepted: 2009-11-24