APPENDIX D

Statistical Review

Natural Gas Distribution Pipeline Systems by Analysis of OPS Incident Data 1990 – 2002

Statistical Review

An important use of the performance data analysis is to determine whether there is any trend in total incident data or incidents attributable to system parts, specific causes, and materials of construction. Examination of the various data plots generated in this study shows that there is a high variability from year-to-year in the number of incidents or incidents per mile. In some cases it appears that these could be a trend up or down or none at all. However, it would be beneficial to be able to say with some sound justification whether a trend actually exists and if so what is its direction. In order to determine whether an upward or downward trend exists for the incident rates during the time period of 1990 to 2002, two methods were used: time series plots and the Mann-Kendall test for trends. In addition, the trend component of a times series was analyzed in order to describe the behavior of the time series.

The first step in determining trends is a graphical representation. Graphical representations are important in revealing long-term trends and may show other major types of trends, such as cycles or impules. The second step is to use the Mann-Kendall test to identify whether a statistically significant decreasing or increasing trend may exist for a given data set. The Mann-Kendall test can be used on data sets when at least four observations per time period of data are available. The more data observations per time period, the "better" the picture of the trend. The Mann-Kendall test is a non-parametric test for trends based on the ranks of the data. The fact that the test is non-parametric means that no assumption about the distribution (e.g. normality) was required. The use of the ranked observations rather than the observations themselves helps to minimize the potential impact of outliers on the results and allows the test to detect trends that are monotonically increasing or decreasing but not necessarily linear.

For this study, the test was performed at the 0.05 and 0.10 significance levels (i.e., p-values smaller that 0.05 or 0.10, respectively were required before the test was concluded to be significant). This means that strong evidence (95% or 90% confidence) was required before a trend was labeled as "significant".

Finally, the trend component of the time series was analyzed. Time series have several components that are used to describe their behavior. One of these components is trend. Trend is the long-term movement of the time series. The long-term movement of the time series can take a variety of shapes, such as linear, quadratic, exponential, or logarithmic. In order to find the equation that best describes the trend, the least squares method was used. The trend line with the R-squared value that is closest to 1 is considered to be the "best fit" trend. However, judgment and experience should be taken into consideration when determining the "best fit".

Results of the tests for the various datasets in this report are presented in Table 2 of Appendix B, Part 1 for the fatality and injury incidents data and Table 2 of Appendix C, Part 1 for the fatality and injury counts.