

Appendix D

Theil Method to Determine Statistically Significant Trends

A simple least-squares technique can be used to analyze any time series of data to determine if the overall trend is positive, negative, or zero. A positive slope indicates a trend toward higher values and a negative slope indicates a decreasing trend. However, air quality data examined over multiple years may show considerable fluctuations from one year to the next (Figure D-1). The negative slope calculated for the seven data points in Figure D-1 might indicate overall improved visibility due to air quality improvements. Therefore, the Theil method was employed in this report to indicate whether or not the observed trends were statistically significant.

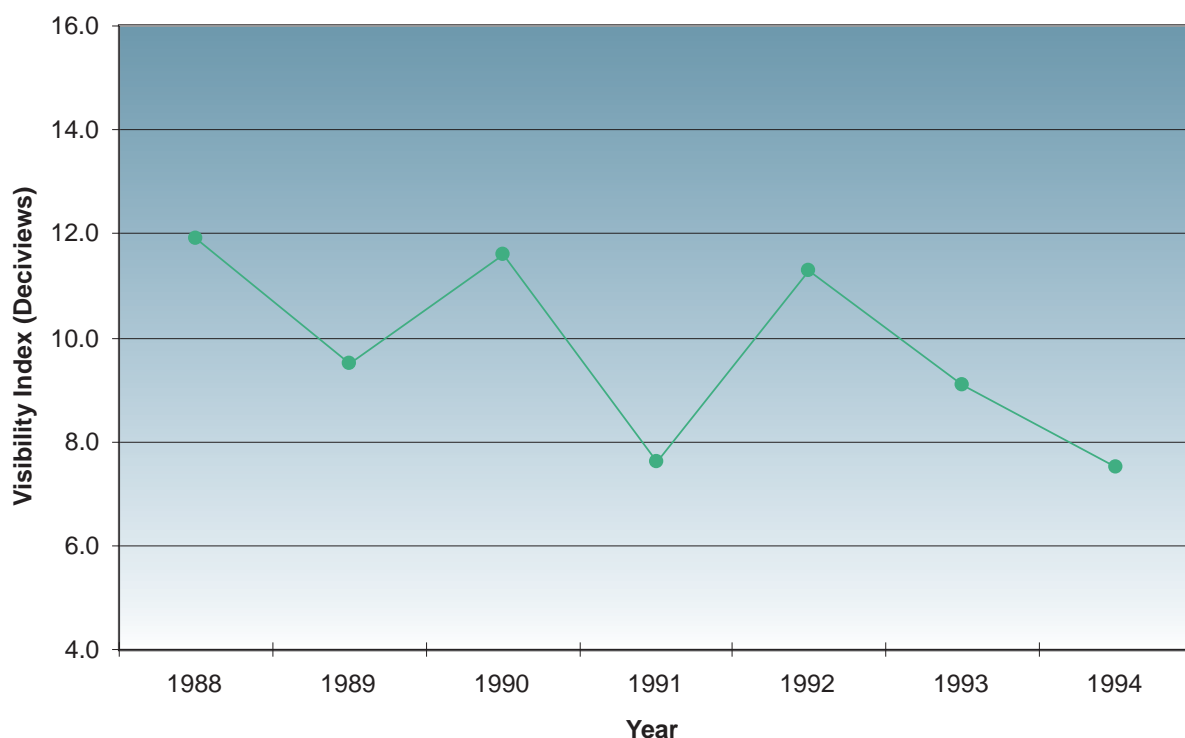


Figure D-1. Sample Data to Illustrate Theil Method

The Theil method utilizes a nonparametric regression technique to determine statistically significant trends. The method is further referenced in the National Air Quality Trends Report (EPA, 1997). All possible pairs of data points (i.e., years) are evaluated to determine whether the visibility index increases, decreases, or remains constant across each pair. Table D-1 displays these combinations for the sample data contained in Figure D-1. For each data pair, values of -1, 0, and +1 are assigned for data that decreases, remains constant, and increases. The assigned values for all pairs are then summed to calculate the S value. The S value is simply a calculated statistic used to determine whether the data exhibit a statistically significant trend. The sample data have an S value of -13.

Table D-1. Calculating S Value for Sample Data.

Starting Data Point	Ending Data Point	Direction (Increase/Decrease/Remain Constant)	Assigned Value
1988	1989	Decrease	-1
1988	1990	Decrease	-1
1988	1991	Decrease	-1
1988	1992	Decrease	-1
1988	1993	Decrease	-1
1988	1994	Decrease	-1
1989	1990	Increase	+1
1989	1991	Decrease	-1
1989	1992	Increase	+1
1989	1993	Decrease	-1
1989	1994	Decrease	-1
1990	1991	Decrease	-1
1990	1992	Decrease	-1
1990	1993	Decrease	-1
1990	1994	Decrease	-1
1991	1992	Increase	+1
1991	1993	Increase	+1
1991	1994	Decrease	-1
1992	1993	Decrease	-1
1992	1994	Decrease	-1
1993	1994	Decrease	-1
TOTAL (S value for 7 data points)			-13

With seven data points, the possible S values range from -21 to +21. For decreasing data sets (i.e., S values less than zero), Figure D-2 shows the probabilities that an S value will be below a certain number for randomly fluctuating data sets. The cumulative probabilities are calculated from the numerical arrays of Kendall's Tau statistics (Kendall, 1990). Reading from Figure D-2, an S value of -13 for seven data points corresponds to a 3 percent cumulative probability. In other words, a randomly fluctuating data set would yield an S value between -21 and -13 in only 3 percent of the cases. If the negative signs of the S values along the ordinate axis in Figure D-2 are made positive, the same graph can be used to determine the probability of increasing trend.

In previous reports (e.g., EPA, 1997), EPA has chosen a 5 percent cumulative probability as the significance criterion to be applied for trend evaluations. If the cumulative probability of a data event is less than 5 percent, then the trend is considered statistically significant. The same criterion was used in

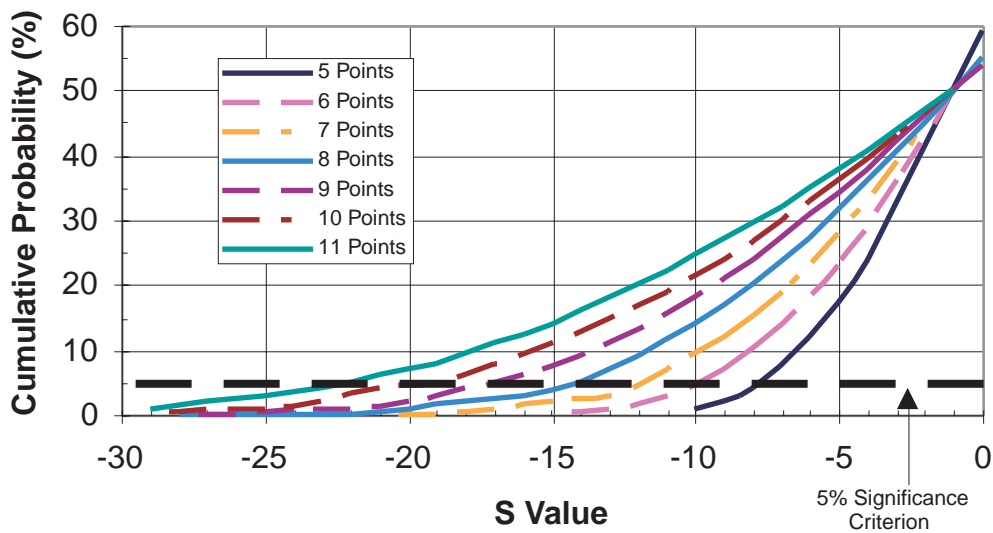


Figure D-2. Probability of Decreasing Trend in a Random Data Series (Kendall's Tau statistic)

this report to determine statistically significant trends of the data sets. The S values were calculated, the cumulative probabilities were determined from Kendall's Tau statistics, and any probabilities less than or equal to 5 percent were considered to indicate statistically significant trends.