SECTION 3: MODEL DEVELOPMENT

SSOs are caused by a combined effect of a large set of interactive factors. Understanding SSOs and their relationship to sewer systems is a complex undertaking. The objective of this study is to capture and quantify a relationship between reported type *B* SSOs and observed levels of some major factors via statistical modeling. In particular, we hope to capture and quantify statistically the impact of sewer system maintenance activities on SSOs, in terms of maintenance intensity and maintenance type.

It is to be mentioned here that any mathematical model is, at its best, an approximation to the true state of the nature. Study results, as consequences of a modeling process, should not be taken as final conclusions, but as reference, information to be used in future studies on similar topics. Particularly in this study, the data was not collected from a carefully designed setting. Although scientific techniques can help to reduce much of the difficulties caused by non-designed survey data, the study remains exploratory in nature.

With this in mind, let us proceed to develop the study model.

3.01 Dependent Variable and Associated Model

Sewer system performances can be gauged by a variety of indicators. Among them, SSOs are probably the most accepted and widely used indicators and are the chosen indicators for system performance in this study. Since SSOs occur in a discrete manner in time and CMU records only report the occurrences and not the amount of SSO, it is natural to study the frequencies via a Poisson Regression Model.

The first step in the study is to aggregate the data in time. Data aggregation is an approximation process. This can be done in different ways according to study objectives. Since we are primarily interested in the macroscopical patterns of SSOs, and since we have more than 14 years of data collected, it is quite reasonable to aggregate the SSO occurrences by month.

We denote the SSO frequency in a given month as Y. It is reasonable to assume that Y is a Poisson random variable with an intensity parameter λ. The monthly SSO frequencies from 1982 to 1997 provided by CMU are independent observations of this random variable. These observations are not identically distributed. The intensity parameter, λ, as a part of the model, is assumed to be a function of many other factors.

Suppose the factors of interest can be measured by variables, X_1, X_2, ... X_k with regard to the SSO frequencies. We assume that the relationship between SSO frequency, Y, and the independent variables, X_1, X_2, ... X_k can be described by the following linear function.

$$\text{(1)} \qquad \log(\lambda)=\mu+\beta_1 X_1+\cdots+\beta_k X_k$$

where $\beta_1, \cdots, \beta_k$ are regression parameters.

3.02 Independent Variables

Before an attempt is made to identify the independent variables, we will revisit the primary question of interest: What causes SSOs? Unfortunately, there are probably no simple or clear-cut answers to the question. For the purpose of this exploratory study, let us adopt a Load-Capacity perspective of sewer system performances, specifically with respect to SSOs.

The Load-Capacity perspective is a simplified filter through which independent variables are selected and interpreted. With this perspective, it is assumed that all SSO factors can be classified into two basic categories. They are load related (Load) and capacity related (Capacity), respectively. In general, it is reasonable to conceptualize the sewer systems as wastewater conveyance systems operating at a capacity level. If the systems are overloaded and its capacity limit is exceeded, then SSOs will occur. Even for a same system, the system capacity is not a constant. It varies according to weather, seasons and many other conditions. For an example, sewer system maintenance is clearly a factor that will affect the conditions and the capacity of the systems. At least, we hope that this study will help to establish the effect of maintenance activities on the system capacity.

In selecting the independent variables, let us first consider rainfall, universally considered one of the most important impacting SSOs. We argue that rainfall should be a secondary independent variable provided the flow volume to the treatment plants is used as an independent variable. There are two steps in our argument for that view.

1. The impact of rainfall on sewer systems is delivered through inflow and infiltration. The process of rainfall becoming inflow and infiltration is a complex one, and not well understood. In gauging the impact of rainfall on the sewer systems, one may bypass the inflow and infiltration process and measure directly the flow to the treatment plants. After all, the impact level of the rainfall is only determined by the amount of rainfall that actually gets into the systems. Of course, flow to the treatment plants not only contains inflow and infiltration by rainfall, but also all other sources of flow. Does that matter? We answer this question in the next step.

2. The main objective of this study is to capture the relationship between sewer system maintenance activities and the sewer system performances. This objective is achieved by examining how much difference maintenance activities can make in system capacity (or system condition), when the system load is controlled. In view of the Load-Capacity perspective defined above, it is sufficient to describe the comprehensive load on the systems, but not necessary to separate the different sources of the load. In other words, as long as the model describes the system load at every point in time, there is no need to specifically describe the proportion of system load induced by rainfall.

After the flow to the treatment plants is adopted as a primary explanatory variable, the amount of rainfall will be numerically gauged in the model as a secondary independent variable. The result will further support the above argument.

Remark: *Although this study does not specifically require a clear mechanism to describe the proportion of inflow and infiltration induced by rainfall, the problem itself is of great importance. The* *industry-standard simulation models for measuring sewer system capacity depends heavily on the value of rain induced inflow and infiltration ratio (I/I ratio). A small difference in the estimated value could lead to a significant difference in the outcomes of the simulation. The estimation problem of I/I ratio deserves a serious separate study.*

Next, let us consider the flow to the treatment plants. Again from the Load-Capacity perspective, many factors could be contributing to the load of the sewer system, but ultimately the combined effect is manifested in the form of the total volume of the wastewater received at the treatment plants. From this viewpoint, the total flow to the treatment plants is naturally an index that will be used in the model to describe the system load.

With regard to groundwater levels, which is also commonly considered as a source of inflow and infiltration, an identical argument to the rainfall can be applied. That is, the portion of groundwater that finds its way into the sewer systems is also included in the total flow to the treatment plants. In fact, a visual inspection of Figure 2 (average monthly flow index versus an adjusted average monthly groundwater level) reveals that the groundwater level is somewhat indicated by the flow to the treatment plants. (Higher value of **WELL** means lower groundwater level.) The correlation coefficient is -0.65. Groundwater level is also considered as a secondary independent variable to be gauged at a later stage of the modeling process.

3.03 Analysis - Stage 1: FLOW

Let us make an attempt to establish a model relationship between SSO frequencies and the flow to the treatment plants, the primary independent variable with regard to system load. First we aggregate both SSO and flow data monthly.

To capture the relationship between SSO and the flow, it is necessary to determine a stable frame of reference in time. The sewer system under CMU's jurisdiction has been expanding continuously in time over the last several decades. We first identify a particular region of the systems which was in place before 1984 and call it "the stable region". This region includes areas with **AGE** values A, B, and C. The dependent variable is defined as

Y = the total monthly SSO frequency in the stable region.

Likewise, adjustments must be made to the total monthly flow to the treatment plants to account for the continuous expansion of the sewer systems in time. An index for the system load is defined to be the total monthly flow to the treatment plants divided by the average length of the system, i.e.,

FLOW = (Total Monthly Flow in MG) / (Sewer Length in Miles),

where MG is millions of Gallons.

Let us consider first the model

(2) $$\log(\lambda)=\mu+\beta\,\mathbf{FLOW}.$$

Using the GENMOD Procedure of SAS version 6.12, (see Appendix C for SAS output,) we have 170 observations,

1. μ is estimated to be 2.2584, the standard error is estimated to be 0.1127 and the p-value of the test statistic for the hypothesis of $\mu=0$ is less or equal to 0.0001.
2. β is estimated to be 0.8789, with an estimated standard error of 0.1285, and the p-value of the test statistic for the hypothesis of $\beta=0$ is less or equal to 0.0001.

This indicates that there is strong evidence suggesting that the SSO frequency, as defined above, is positively related to the flow index. A higher level of the flow index, **FLOW**, leads to a higher probability of an SSO, or a higher average of a monthly SSO frequency.

At this point, we introduce an intuitive way of interpreting a statistic associated with the Poisson regression methodology. The statistic is Deviance. Deviance is a special statistical distance measuring how much of the fluctuation of SSO frequency in time is explained by the model employed here.

It is often useful to keep in mind that if a model could explain completely why SSO frequency fluctuates in time, then we would have had the complete knowledge of what were the factors of SSO, qualitatively and quantitatively. In reality, we do not have that kind of knowledge and, we rely on statistical distances such as Deviance to tell us how much of the total variation (or deviance) is explained by a specific factor (or independent variable).

In the current study, the total deviance is 803.92. This is the deviance after the constant μ is fitted. The deviance, after **FLOW** is fitted, is 730.13. The difference, 73.79 or 9.2% of the total deviance, is explained by the linear term **FLOW** in the model. The most important statistic here is the percent 9.2%. This value projects how much fluctuation in SSO that can be attributed to the change in the flow index.

3.04 Analysis - Stage 2: Seasons

Next, we consider the seasonal effect on the SSO frequency after the flow index is included. It is to be pointed out that the seasonal effect considered here is a Capacity effect, not a Load effect. Seasonal trend is very clear in the SSO frequency plot in Figure 3. This trend is caused by a combination of two separate trends: one is the seasonal trend of flow into the system by varying natural and human behaviors, and the other is caused by the change in the condition of the system in conveying wastewater. Again, we consider the seasonal effect in the framework of the Load and Capacity perspective discussed above.

The seasonal trend of flow into the system has already been captured by the flow index, as clearly shown in Figure 3. By adding a seasonal factor after **FLOW** is fitted, the new factor is expected to capture only the seasonal fluctuation in the condition (or the capacity) of the sewer systems.

The seasonal factor is introduced into the model by categorical variables $\mathbf{M}_k$, k=1,2, ..., 12. For example, M_1 is for the month of January. $\{M_1=1\}$ means that the month is January, and $\{M_1=0\}$

means that the month is not January. The index k is for the k^{th} month of a year, i.e., 1 (January), 2 (February), 3 (March), 4 (April), 5 (May), 6 (June), 7 (July), 8 (August), 9 (September), 10 (October), 11 (November) and 12 (December).

The model at this stage is

(3) $$\log(\lambda)=\mu+\beta\,\mathbf{FLOW}+\Sigma_{k=1\text{ to }11}\,\beta_k\,\mathbf{M}_k.$$

There are only 11 terms for the season in the above model. This is so because the month of December is indicated by $\mathbf{M}_k$=0, k=1, 2, ..., 11. These 11 terms form one group of variables to gauge seasonal change in sewer system condition.

The SAS GENMOD Procedure showed the following major results, (see Appendix C for SAS output,) with 170 observations.

1. μ is estimated to be 2.7535, the standard error is estimated to be 0.1949 and the p-value of the test statistic for the hypothesis of μ=0 is less or equal to 0.0001.
2. β is estimated to be 0.5554, with an estimated standard error of 0.1915, and the p-value of the test statistic for the hypothesis of β=0 is equal to 0.0037.
3. The estimated values for the parameters β_1 through β_{12} are as tabulated in the following table.

Parameter	Estimate	p-value < or =
β_1	0.0919	0.4984
β_2	0.0114	0.9344
β_3	-0.0039	0.9776
β_4	-0.1964	0.1801
β_5	-0.2805	0.0605
β_6	-0.4457	0.0052
β_7	-0.5861	0.0003
β_8	-0.6563	0.0001
β_9	-0.5614	0.0008
β_{10}	-0.2856	0.0599
β_{11}	-0.0819	0.5671

β_{12}	0.0000	NA

One may interpret these estimates as follows. Using β_{12} as a level of reference, in January, February, March, April and November, the conditions of the sewer systems are not very different from that of December. From May to October, the system condition (or capacity) is significantly better, and the likelihood of SSO decreases as manifested by the negative values of the estimates.

The monthly average of SSO frequencies from 1983 to 1997 is graphed in Figure 4. The above table is also graphed in Figure 4. As seen in the comparison, the observed seasonal trend is largely captured by the estimates of the parameters.

In terms of the deviance explained by the current model, **FLOW** still explains 73.8 in the total deviance (803.9), or 9.2%; the group of season variables, M_1, ..., M_{12}, explains an additional 193.58 in the total deviance, or 24.1%.

3.05 Analysis - Stage 3: General Maintenance

At this point, a cumulative 33.3% of the total deviance has been described by the model employed. Conversely, the remaining 66.7% of the total deviance is not explained. In Stage 1, we started with 100% of the total deviance, and we then used one factor, flow index, to describe the system load, and a second factor, season index, to describe the seasonal change of system capacity. Now we face 66.7% of the total deviance. What other factors are important in explaining the remainder?

With regard to the performance of the sewer systems, one may view maintenance activities as means of improving the system capacity. It is natural to gauge the model relationship with maintenance activities. To do so, we must develop reasonable index measures for comprehensive maintenance intensity.

In the data provided by CMU, we have identified 9 different specific operation codes (see Data Description) that are considered pro-active with regard to controlling SSO. These data are aggregated yearly.

Since the sewer systems under CMU jurisdiction have expanded continuously over the years, the comprehensive activity data kept by CMU must be adjusted for the fixed region of reference, the stable region since 1984. After some considerable consultation, we decided that all the yearly maintenance data should be converted to measures per unit (per linear mile of sewer). Furthermore, the adjusted maintenance measures are all normalized to account for their vastly different scales and variations over the years.

Before the adjustments, it is to be pointed out that there are three types of maintenance operations, CHAMPS Codes 9, 10, and 14, that are not completely pro-active. Code 9 (X09) is the footage of sewer cleaned by Jets & Combination Machines. Code 10 (X10) is the footage of sewer cleaned by Rodder. Code 14 (X14) is footage of sewer inspected with TV cameras. These activities, in addition to the regularly scheduled maintenance, are ordered to respond to each reported SSO. According to CMU, on the average, each reported SSO requires a section of sewer of length 250 feet to be cleaned and inspected. To take this passive portion of the maintenance, caused directly by SSOs, out of the maintenance intensity measures, we define, for each year,

1. X09s = X09 - 250 • (Total yearly number of SSOs),
2. X10s = X10 - 250 • (Total yearly number of SSOs),
3. X14s = X14 - 250 • (Total yearly number of SSOs).

Next, for each year, let

1. X08* = X08 / (Total system length in miles),
2. X09* = X09s / (Total system length in miles),
3. X10* = X10s / (Total system length in miles),
4. X11* = X11 / (Total system length in miles),
5. X12* = X12 / (Total system length in miles),
6. X14* = X14s / (Total system length in miles),
7. X15* = X15 / (Total system length in miles),
8. X16* = X16 / (Total system length in miles),
9. X17* = X17 / (Total system length in miles).

Finally, we standardize these variables.

1. Z08 = [X08* - (the mean of X08*)] / (the standard deviation of X08*),
2. Z09s = [X09* - (the mean of X09*)] / (the standard deviation of X09*),
3. Z10s = [X10* - (the mean of X10*)] / (the standard deviation of X10*),
4. Z11 = [X11* - (the mean of X11*)] / (the standard deviation of X11*),
5. Z12 = [X12* - (the mean of X12*)] / (the standard deviation of X12*),
6. Z14s = [X14* - (the mean of X14*)] / (the standard deviation of X14*),
7. Z15 = [X15* - (the mean of X15*)] / (the standard deviation of X15*),
8. Z16 = [X16* - (the mean of X16*)] / (the standard deviation of X16*),
9. Z17 = [X17* - (the mean of X17*)] / (the standard deviation of X17*).

To describe the general intensity level of CMU's pro-active sewer maintenance, the most natural statistic to use is probably the average of the above 9 individual indices, which will be denoted by **Z**.

Z = (Z08 + Z09s + Z10s + Z11 + Z12 + Z14 + Z15 + Z16 + Z17)/9.

Can this general pro-active maintenance index explain some of the remaining deviance from Stage 2?

The model at this stage is

(4) $$\log(\lambda)=\mu+\beta\,\mathbf{FLOW}+\Sigma_{k=1\text{ to }11}\,\beta_k\,\mathbf{M}_k+\alpha\,\mathbf{Z}.$$

The SAS GENMOD Procedure showed the following major results, (see Appendix C for SAS output,) with 170 observations. **Z** is represented in the output as **ZMEAN**.

The estimated values for the parameters are as tabulated in the following table.

Parameter	Estimate	p-value < or =
μ	2.8952	0.0001
β	0.3765	0.0227
β_1	0.0807	0.4942
β_2	-0.0092	0.9394
β_3	-0.0074	0.9513
β_4	-0.2185	0.0870
β_5	-0.2954	0.0232
β_6	-0.4793	0.0006
β_7	-0.6060	0.0001
β_8	-0.6739	0.0001
β_9	-0.5819	0.0001
β_{10}	-0.2995	0.0235
β_{11}	-0.0866	0.4872
β_{12}	0.0000	NA
α	-0.4268	0.0001

First we notice that the estimated α is -0.4268 with strong statistical evidence (p-value is less or equal to 0.0001) that the true value of α is negative. **That means that higher level of Z, the pro-active maintenance index, leads to lower level of λ, and in turn a down shift of the probability distribution of SSO frequency.**

The deviance, explained by this variable, is 131.99 or 16.42% of the total deviance (803.9205).

Cumulatively the model can explain 49.68% the total deviance at this point.

3.06 Analysis - Stage 3*: Individual Maintenance Types

At this stage, we will examine the impact of each individual type of maintenance activity, as reflected by the model.

It is very important to keep in mind, as we run through the individual types of activities, that each type is embedded in a comprehensive maintenance program. Any relationship established in this section of the analysis should be interpreted in conjunction with the comprehensive nature of the overall maintenance. It is hoped that the interpretations provided here will be taken as preliminary suggestions.

The base model used here is

(5) $$\log(\lambda)=\mu+\beta\,\mathbf{FLOW}+\Sigma_{k=1\text{ to }11}\,\beta_k\,\mathbf{M}_k.$$

The justification for that is that this model covers the system load change and the system capacity change because of the seasons. With these two major factors under consideration, it is reasonable to ask whether or how much each individual type of maintenance activity can help to explain the fluctuation of SSO frequency.

To do that, we use the following model.

(6) $$\log(\lambda)=\mu+\beta\,\mathbf{FLOW}+\Sigma_{k=1\text{ to }11}\,\beta_k\,\mathbf{M}_k+\alpha_i\,\mathbf{Z}_i$$

With **Zi**, i =1, 2, ..., 9, being any one of the nine individual maintenance intensity indices, we will estimate α_i and gauge its statistical significance.

Rapid Response. A rapid response crew carries out a work order immediately after a sewer related problem is reported. This type of activity is coded as Z08. Let Z08 be the Z_i in (6). The SAS GENMOD Procedure showed the following major results, (see Appendix C for SAS output,) with 170 observations.

The estimated α_i is -0.3188 with a p-value less or equal to 0.0001 in testing the hypothesis that $\alpha=0$. This implies that, with high statistical confidence, rapid responses tend to reduce SSO frequency. In consultation with CMU operators, they suggested that this relationship may be attributed to the ability of averting a potential SSO before it actually occurs.

This variable, Z08, explains an additional 242.5 (30.16%) in the remainder deviance from the base model (5). This brings the total deviance explained by the model up to 509.78, or 63.42%.

Jets & Combination Machines. Jets & Combination Machines stands for machines used in cleaning procedures with high-pressure water and debris vacuuming capability. This type of activity is coded as Z09s. Let Z09s be the Z_i in (6). The SAS GENMOD Procedure showed the following major results, (see Appendix C for SAS output,) with 170 observations.

The estimated α_i is 0.0851 with a p-value 0.0095 in testing the hypothesis that α_i=0. At a first glance, this may seem to imply that, with a positive estimate for α_i, such cleaning procedure may lead to a higher SSO frequency, although slightly. We are reminded of the existence of the comprehensive maintenance program. A positive estimate here only suggests that such procedure is not as effective as some other cleaning procedures. More usage of Jets & Combination Machines may be taking away resources from other more effective maintenance activities. It still does not mean that this procedure can be replaced by a more effective one. It simply suggests that the spectrum of situations, when such procedure was called for, as in CMU's current practice, might have been wider than it should be. In fact, according to CMU operators, CMU has already started to shift to a more effective procedure.

This variable, Z09s, explains only an additional 21.36 (2.66%) in the remainder deviance from the base model (5). This brings the total deviance explained by the model up to 288.73, or 35.92%.

Rodder. Rodder (Root Removal) is a machine with a root-removing device used in cleaning procedures. This type of activity is coded as Z10s. Let Z10s be the Z_i in (6). The SAS GENMOD Procedure showed the following major results, (see Appendix C for SAS output,) with 170 observations.

The estimated α_i is -0.2211 with a p-value less or equal to 0.0001 in testing the hypothesis that α_i=0. The negative value of the estimate implies such cleaning procedure tends to lead to lower SSO frequency. By a comparison with Jets & Combination Machines, this procedure seemed much more effective, at least from a modeler's point of view.

This variable, Z10s, explains an additional 132.01 (16.42%) in the remainder deviance from the base model (5). This brings the total deviance explained by the model up to 399.38, or 49.68%.

Off-Street. Off-Street Cleaning stands for a labor-intensive procedure in which maintenance workers manually clean and remove roots or debris in hard-to-reach areas where the use of other machinery is not practical. This type of activity is coded as Z11. Let Z11 be the Z_i in (6). The SAS GENMOD Procedure showed the following major results, (see Appendix C for SAS output,) with 170 observations.

The estimated α_i is -0.1710 with a p-value less or equal to 0.0001 in testing the hypothesis that $\alpha_i=0$. The negative value of the estimate implies such cleaning procedure tends to lead to lower SSO frequency.

This variable, Z11, explains an additional 84.60 (10.52%) in the remainder deviance from the base model (5). This brings the total deviance explained by the model up to 351.97, or 43.78%.

Right-of-Way Mowing. Right-of-Way Mowing is an activity to clear or maintain access paths to sewer lines by the creeks. This type of activity is coded as Z13. Let Z13 be the Z_i in (6). The SAS GENMOD Procedure showed the following major results, (see Appendix C for SAS output,) with 170 observations.

The estimated α_i is -0.1633 with a p-value less or equal to 0.0001 in testing the hypothesis that $\alpha_i=0$. The negative value of the estimate implies such procedure tends to lower SSO frequency.

This variable, Z13, explains an additional 97.95 (12.18%) in the deviance remainder from the base model (5). This brings the total deviance explained by the model up to 365.32, or 45.44%.

T.V. T.V stands for the use of television camera in inspecting the sewer pipes. This type of activity is coded as Z14s. Let Z14s be the Z_i in (6). The SAS GENMOD Procedure showed the following major results, (see Appendix C for SAS output,) with 170 observations.

The estimated α_i is -0.0354 with a p-value 0.3106 in testing the hypothesis that $\alpha_i=0$. The large p-value indicates that there is no sufficient evidence to suggest that the true value of α_i is non-zero. This procedure does not seem to have a very significant impact on SSO frequency.

This variable, Z14s, explains only an additional 3.54 (0.44%) in the remainder deviance from the base model (5). This brings the total deviance explained by the model up to 270.91, or 33.70%.

Herbicide. Herbicide stands for the application of herbicide to control root growth in sewer pipes. This type of activity is coded as Z15. Let Z15 be the Z_i in (6). The SAS GENMOD Procedure showed the following major results, (see Appendix C for SAS output,) with 160 observations (due to some missing values of the Herbicide data).

The estimated α_i is -0.2296 with a p-value less or equal to 0.0001 in testing the hypothesis that $\alpha_i=0$. The negative value of the estimate implies such procedure tends to lead to lower SSO frequency.

With only 160 observations available, the total deviance in the sample is also changed to 782.4079. This variable, Z15, explains an additional 92.89 (11.78%) in the remainder deviance from the base model (5), which is 527.36. (See Appendix C.) This brings the total deviance explained by the model up to 347.93, or 44.47%.

Manhole Inspection and Cleaning. Manhole Inspection and Cleaning is largely an alternative when a weather condition prevents other regular maintenance activities to be carried out in any meaningful way. This type of activity is coded as Z16. Let Z16 be the Z_i in (6). The SAS GENMOD Procedure showed the following major results, (see Appendix C for SAS output,) with 170 observations.

The estimated α_i is 0.0875 with a p-value 0.0057 in testing the hypothesis that $\alpha_i=0$. The positive value of the estimate implies that such activities tend to lead to higher SSO frequency. Why should inspection and cleaning of manhole do any harm to the sewer maintenance? They do not. This, in fact, is an excellent example to illustrate that the relationships established in this section must be interpreted in conjunction with the other maintenance activities. This particular activity is known to be inefficient in the sense that it wastes resources, which may otherwise be used to achieve greater good for system maintenance.

This variable, Z16, explains on an additional 24.69 (3.07%) in the remainder deviance from the base model (5). This brings the total deviance explained by the model up to 292.06, or 36.33%.

Inspection. This inspection stands for regular scheduled sewer system inspection. This type of activity is coded as Z17. Let Z17 be the Z_i in (6). The SAS GENMOD Procedure showed the following major results, (see Appendix C for SAS output,) with 160 observations, again due to some missing values in inspection data.

The estimated α_i is -0.2872 with a p-value less or equal to 0.0001 in testing the hypothesis that α_i=0. The negative value of the estimate implies such cleaning procedure tends to lead to lower SSO frequency.

This variable, Z17, explains an additional 164.41 (20.41%) in the remainder deviance from the base model (5). This brings the total deviance explained by the model up to 431.45, or 53.67%.

Relative Ranking. The above analysis suggests that, these individual maintenance activities may be relatively ranked according to their ability in explaining the deviance remainder from the base model (5). Consider a empirical score for relative strength, RS, defined as follows.

(7) RS = – (Sign of Estimated α_i) • (Proportion of Deviance by Zi).

In a decreasing order, we have the following ranking.

TYPE	RS
Rapid Response (Z08)	30.16%
Inspection (Z17)	20.41%
Rodder (Z10s)	16.42%

Right-of-Way Mowing (Z13)	12.18%
Herbicide (Z15)	11.78%
Off-Street (Z11)	10.58%
T.V. (Z14s)	0.44%
Jets & Combination Machines (Z09s)	-2.66%
Manhole Inspection and Cleaning (Z16)	-3.07%

Remarks.

The ordering of the maintenance activity types here should not be taken as a rank of importance of these activities out of context. Each Relative Strength is calculated for a particular activity without others being considered in the model. Since all types of maintenance activities are used concomitantly, it is probably best to interpret such ordering in the following fashion.

Sewer system maintenance is a complex task. The resources for maintenance are often limited. There is usually a large variety of maintenance situations that may call for different procedures. It would not be reasonable to claim one particular procedure is better than another is in general. Rather the mixture of different maintenance procedures in an existing program can be viewed, by either design or tradition or convenience, as a means to utilize the combined effect of these component procedures.

If one accepts the above viewpoint, then one cannot help to ask what the optimal (or most efficient) mixture of the component activities may be to control SSO frequency. The answer to such question is unknown, and not easy to obtain. The ordering provided in the above table may be interpreted nicely in this context. Given the current (or the last 14 years' average) mixture of the component maintenance procedures at CMU, the table suggests that the overall effect may be improved if the activities with higher RS scores are increased and those with lower scores, particularly the negative ones, are decreased.

In summary, the RS scores are meaningful only with respect to the current state of the mixture of maintenance activities in a particular program.

3.07 Analysis - Stage 4: Maintenance Management

Having provided a ranking for the various types of the maintenance activities, let us return to Stage 3 where the model considered is

$$\log(\lambda)=\mu+\beta\,\mathbf{FLOW}+\Sigma_{k=1\text{ to }11}\,\beta_k\,\mathbf{M}_k+\alpha\,\mathbf{Z}. \quad (8)$$

According to CMU, the management style of the maintenance program for its sewer system has three clearly different phases during the last 14 years. Before 1990, there was a regular maintenance schedule, developed according to Schaaf's suggestion and CHAMPS data. Although the schedule algorithm was not computerized then, the actual execution of the schedule was believed to be reasonably close to what was intended. In 1990, there was a management change, and therefore, the philosophy of regular sewer maintenance was changed. Between 1990 and 1994, the regular maintenance was largely carried out by means of cleaning a whole neighborhood or subdivision, when there was a reported sewer problem near by. Most maintenance orders were issued based on subjective judgment and convenience. The argument for such a method is twofold. One is that a reported problem is usually an indication that this particular area needs maintenance. The other is that it is cost-efficient to clean the area with a reported problem while a crew is already in the area. This philosophy provides a contrast to the Schaaf's methodology, in the sense that Schaaf's methodology relies on a balance between reported current problems and cleaning history. The philosophy adopted by CMU between 1990 and 1994 weighted much on the reported current problems. Did they weight it too high? We will attempt to answer that question at this stage. It is to be mentioned first that, from 1995 on, CMU has again moved to carry out their regular maintenance based on Schaaf's methodology. In fact, this time around, the scheduling algorithm is computerized.

With the above information, it is reasonable to define an independent variable, say, **SCHAAF**, to distinguish the period from 1990 to 1994, from the other two periods. Let {**SCHAAF**=1} mean the time when Schaaf-based maintenance schedule was implemented, and let {Schaaf=0} stand for the time when Schaaf's method was not followed.

Based on (8); let us consider

(9) $\log(\lambda)=\mu+\beta$ **FLOW**$+\Sigma_{k=1 \text{ to } 11}\, \beta_k$ $\mathbf{M_k}$$+\alpha$ **Z**$+\delta$ **SCHAAF**,

where δ is the parameter corresponding to variable **SCHAAF**.

The SAS GENMOD Procedure showed the following key results, (see Appendix C for SAS output,) with 170 observations.

Parameter	Estimate	p-value
β	0.0369	0.8395*
α	-0.5447	0.0001
δ	-0.3188	0.0001

The negative estimate of δ suggests that, when Schaaf's methodology is used, the average frequency of SSO is decreased, if levels of all other factors are held constant.

(*) The p-value here, for the hypothesis that $\beta=0$, is 0.8395. This indicates that the newly included variable, **SCHAAF**, is somewhat correlated with the variable **FLOW**. It is beyond doubt that **FLOW** is a useful variable in influencing the frequency of SSOs. The Type 1 Analysis (see Appendix C) shows that if one adds **SCHAAF** in the model after the **FLOW**, the season and the general maintenance are already fitted, an additional deviance of 36.05 or approximately 4.48 % of the total deviance is explained. At this stage, our main goal is to examine whether **SCHAAF** increase the power of the model. We will go back to re-gauge variable **FLOW** at a later stage.

At this stage, the variables representing the flow index, the seasons, the general maintenance intensity, and the Schaaf's methodology, together explains 54.16% of the total deviance.

3.08 Analysis - Stage 5: Hugo

In 1989, Hurricane Hugo hit Charlotte-Mecklenburg area in late September. The system was overwhelmed by the storm, and its after-effect lingered for several months. Many of the SSOs are presumably Hugo related. To figure the Hugo effect in the model, we let {**HUGO**=1} stand for the year when the hurricane hit, and {**HUGO**=0} for other years. We consider model

(10) $\log(\lambda)=\mu+\beta\,\mathbf{FLOW}+\Sigma_{k=1\text{ to }11}\,\beta_k\,\mathbf{M_k}+\alpha\,\mathbf{Z}+\delta\,\mathbf{SCHAAF}+\theta\,\mathbf{HUGO}$,

where θ is the parameter corresponding to variable **HUGO**.

The SAS GENMOD Procedure showed the following key results, (see Appendix C for SAS output,) with 170 observations.

Parameter	Estimate	p-value
β	0.1477	0.4127**
α	-0.5015	0.0001
δ	-0.2468	0.0030
θ	0.3124	0.0003

The positive estimate of θ suggests that Hugo may have been the reason why SSO frequency surges in 1989.

(**) We notice that, with **HUGO** in the model, this p-value decreases considerably (as compared to the previous model in (9)). This suggests that **HUGO** may have explained some of the correlation between **FLOW** and **SCHAAF**.

At this stage, the variables representing the flow index, the seasons, the general maintenance intensity, the Schaaf's methodology, and Hugo together explains 57.55% of the total deviance.

3.09 Analysis - Stage 6: FLOW Revisited

Now let us go back and investigate variable **FLOW** a little further.

Thus far, **FLOW** has been classified as a Load variable. Since the flow to the treatment plants also has demonstrated seasonal trend over the years, the 9.2% of the total deviance attributed to **FLOW** (See Stage 1) may reflect some contribution from Capacity factors, for example, the seasons. Is **FLOW** really a Load variable? To answer this question, we expand the model to include a non-linear term in **FLOW**. The need of a non-linear term in **FLOW** is also suggested by the fact that at Stage 4, when we include variable, **SCHAAF**, the p-value for β not equal to zero is greatly inflated. This fact suggests that **SCHAAF** may be linearly related to **FLOW**, and an added non-linear term may help the calculation, the power and the validity of the model.

Let us consider

(8) $\log(\lambda) = \mu + f(\mathbf{FLOW}) + \Sigma_{k=1 \text{ to } 11} \beta_k \mathbf{M_k} + \alpha\ \mathbf{Z} + \delta\ \mathbf{SCHAAF} + \theta\ \mathbf{HUGO}$.

where f(**FLOW**)= a **FLOW** + b $\mathbf{FLOW}^2$, and a, b are parameters.

The SAS GENMOD Procedure showed the following key results, (see Appendix C for SAS output,) with 170 observations.

Parameter	Estimate	p-value
a	1.9856	0.0855
b	-0.9634	0.1069
α	-0.4984	0.0001
δ	-0.2229	0.0077
θ	0.2971	0.0006

The Type 1 and 3 Analysis (see Appendix C) shows that the non-linear expansion of the model with respect to **FLOW** is supported by the data. All the variables in the model at this stage, together, explain 430.74 in deviance or 58.29% of the total deviance.

3.10 Analysis - Stage 7: The Final Model

A 1998 study by American Society of Civil Engineers and Black & Veatch, LLP, for US EPA suggests in Section 1.4 that the sewer system aging process is indexed by the remaining value of the system, and such value decreases in time at a constant yearly rate. The basic point of reference adopted in that study is that, without any maintenance the system will deteriorate at a constant rate for about one hundred years. The role of sewer maintenance activities is then to slow or reverse the aging of the system.

To complete this study, it is necessary to, after all reasonable factors are considered, include Time as a final term to see if the amount of maintenance by CMU in the last 14 years has prevented the aging of the system. Let **T** be the yearly time from 1983 to 1997, we consider

(11) $\log(\lambda) = \mu + f(\mathbf{FLOW}) + \Sigma_{k=1 \text{ to } 11}\, \beta_k \mathbf{M_k} + \alpha\, \mathbf{Z} + \delta\, \mathbf{SCHAAF}$
$+ \theta\, \mathbf{HUGO} + \kappa\, \mathbf{T}$,

where $f(\mathbf{FLOW}) = a\, \mathbf{FLOW} + b\, \mathbf{FLOW}^2$, and a and b are parameters associated with **FLOW**, and κ is the unknown parameter describing the unit rate of decay corresponding to yearly time in SSO frequency.

The SAS GENMOD Procedure showed the following major results, (see Appendix C for SAS output,) with 170 observations.

Parameter	Estimate	p-value
μ	-119.3656	0.0001
a	2.5839	0.0175
b	-1.4066	0.0133
$\beta 1$	0.1135	0.2648
$\beta 2$	-0.0200	0.8471
$\beta 3$	0.0312	0.7660
$\beta 4$	-0.2321	0.0343
$\beta 5$	-0.3067	0.0063
$\beta 6$	-0.5159	0.0001
$\beta 7$	-0.6359	0.0001
$\beta 8$	-0.6908	0.0001
$\beta 9$	-0.6213	0.0001
$\beta 10$	-0.3215	0.0045
$\beta 11$	-0.0883	0.4090
$\beta 12$	0.0000	.
α	-0.2324	0.0044
δ	-0.5280	0.0001
θ	0.3658	0.0001
κ	0.0611	0.0001

The positive value of the estimated κ suggests that the system is aging despite of the maintenance effort.

This model explains 64.05% of the total deviance.