Appendix G

National Estimate of SSO Frequency and Volume

G.1 Summary

National estimates of SSO frequency and volume were generated based on reporting data for more than 33,000 SSO events provided by 25 states during calendar years 2001, 2002, and 2003. The discussion of SSOs in this report does not account for discharges from points after the headworks of the treatment plant, regardless of the level of treatment, or backups into buildings caused by problems in the publicly-owned portion of the SSS. Therefore, these estimates of SSO volume and frequency do not account for discharges occurring after the headworks of the treatment plant or backups into buildings. Estimates were generated by extrapolating these data across the remaining 25 states and Washington, D.C., through a five-step procedure:

1. **Tabulate the total number of SSO events and the SSO volume for each of the reporting states.**
2. **Estimate the total number of SSO events per year for each non-reporting state** based on the number of sewer systems in the state or the population served by sewer systems in the state.
3. **Divide the total number of events in each non-reporting state into different categories describing the cause of the SSO event** based on observed regional differences among the 25 reporting states.
4. **Calculate the SSO volume for each cause category in each non-reporting state** to account for observed regional differences.
5. **Calculate national estimates** by summing the total number of events by state and the total volume across all states.

A range of estimates corresponding to different assumptions regarding the nature of the reported data was generated. Results of the analysis indicate that the annual frequency of SSO events is between 23,000 and 75,000 events per year, with a corresponding volume of 3 to 10 billion gallons per year. This relatively large range is due to uncertainty regarding the extent to which the reported SSO events reflect all SSO events that occurred during calendar years 2001, 2002, and 2003.

This appendix summarizes how the estimates of SSO frequency and volume were generated. The discussion is grouped into the following sections:

- Data available
- Inital data assessment
- Assumptions
- Calculation procedure
- Results

G.2 Data Available

The national estimates of SSO frequency and volume were based on data provided by 25 states over three years (January 1, 2001 to December 31, 2003). Data were provided by:

- California
- Colorado
- Connecticut
- Florida
- Georgia
- Hawaii
- Indiana
- Kansas
- Maine
- Maryland
- Massachusetts
- Michigan
- Minnesota
- New Hampshire
- Nevada
- North Carolina
- North Dakota
- Oklahoma
- Rhode Island
- South Carolina
- South Dakota
- Utah
- Washington
- Wisconin
- Wyoming

These data were obtained directly from the NPDES authority in each of these states. Data were typically compiled in either a database or spreadsheet. Specific data elements tracked by NPDES authorities are summarized in Table G.1.

Table G.1 Data Elements Tracked by NPDES Authorities with Electronic Systems

State	Date & Time Reported	Start Date & Time	End Date & Time/ Duration	Total Overflow Volume (gallons)	SSO Location[a]	SSO Cause	Response Measures Taken[b]	Receiving Water Identified
CA	●	●	●	●	●	●	●	●
CO	●	●	●	●	●	●	●	●
CT		●	●	●	●	●		
FL		●		●	●	●		●
GA	●	●		●	●			●
HI	●	●	●	●	●	●	●	●
IN		●	●	●	●	●		
KS		●	●		●			
MA	●	●	●	●	●	●	●	●
MD		●	●	●	●	●		●
ME	●	●	●	●	●	●	●	●
MI	●	●	●	●	●	●	●	●
MN	●	●	●	●	●	●	●	
NC		●	●	●	●	●		●
ND	●	●	●	●		●	●	●
NH	●	●	●	●	●	●	●	●
NV	●	●		●	●	●	●	●
OK		●	●	●	●	●	●	
RI	●	●	●	●	●	●	●	●
SC	●			●	●			●
SD	●	●	●	●	●	●	●	●
UT						●		
WA		●		●	●	●	●	●
WI		●	●	●	●	●	●	●
WY		●				●	●	●

[a] May not include exact SSO location point
[b] May include cleanup activities, volume recovered, and corrective or preventive measures

A total of 36,325 SSO event records were collected from the NPDES authorities and compiled into a single data management system. No attempt was made to verify the quality or accuracy of the data with the individual jurisdictions reporting the SSOs; however, a quality check of the data was performed to identify discrepancies (e.g., gallons versus million gallons, dates outside of the 2001-2003 range, and records with no dates). Reported events with missing event volumes were used to generate frequency estimates, but were not used for volume

estimates. Reported events with dates outside of the 2001-2003 range and events with no dates were not used to estimate SSO frequency or volume. After the screening, a total of 33,213 records remained describing event frequency, and 28,708 records remained describing event volumes.

Some states did not provide data for the entire three-year period. SSO frequency data for these states were adjusted proportionally to account for the missing months. For example, the number of observed events for a state providing 21 months of reports out of the 36-month period were scaled up by a factor of 36/21.

Basic information describing the sewered universe in each state was obtained from the *2000 CWNS* and included:

- Total number of collection systems by state
- Number of SSSs by state
- Population served by SSSs by state

Lastly, data for each state were grouped by EPA Region to facilitate analysis of regional differences. The number of SSSs and the population served are presented by Region in Table G.2.

Table G.2 Number of SSSs and the Population Served by the SSSs for each Region

EPA Region	Total number of SSSs	Population served by SSSs (in millions)
1	705	6.18
2	1,518	14.51
3	2,149	15.49
4	2,678	29.89
5	4,296	27.05
6	2,983	25.67
7	2,619	7.58
8	1,437	7.78
9	1,003	33.38
10	823	6.36

G.3 Initial Data Assessment

Several analyses were conducted to assess the data prior to estimating SSO frequency and volume. The initial data assessment included:

- Analysis of climatic conditions
- Analysis of the variability of discharge volume, number of systems reporting at least one event, population served, and SSO event frequency
- Regional characterization of frequency, volume, and cause category
- Analysis of volume discharged by cause category
- Statistical regressions to predict frequency of events for non-reporting states

These analyses are discussed in more detail below.

Analysis of Climatic Conditions

Annual precipitation statistics developed by the National Climatic Data Center (NCDC) suggest that no obvious bias towards wet or dry conditions was observed. Climatic conditions in the reporting states over the three-year period ranged from record drought to record rainfall. The results of NCDC's analysis for 2001-2003 are presented in Table G.3 for the 24 of the reporting states. Data for Hawaii were not available.

Table G.3 National Precipitation Summary by State for 2001-2003

State	2001	2002	2003
CA	Near normal	Below normal	Near normal
CO	Near normal	Record driest	Below normal
CT	Below normal	Near normal	Much above normal
FL	Near normal	Above normal	Above normal
GA	Below normal	Near normal	Above normal
IN	Above normal	Above normal	Above normal
KS	Near normal	Below normal	Near normal
MA	Near normal	Near normal	Much above normal
MD	Much below normal	Near normal	Record wettest
ME	Record driest	Near normal	Near normal
MI	Much above normal	Near normal	Near normal
MN	Above normal	Above normal	Below normal
NC	Much below normal	Near normal	Record wettest
ND	Near normal	Below normal	Below normal
NH	Much below normal	Near normal	Above normal
NV	Near normal	Much below normal	Near normal
OK	Near normal	Near normal	Below normal
RI	Near normal	Near normal	Above normal
SC	Much below normal	Near normal	Above normal
SD	Above normal	Below normal	Near normal
UT	Near normal	Much below normal	Near normal
WA	Near normal	Below normal	Near normal
WI	Above normal	Above normal	Below normal
WY	Below normal	Much below normal	Below normal

Analysis of Variability

The first data assessment step was to statistically characterize the variability in the parameters used to generate the national estimates. The first parameter evaluated was SSO event discharge volume. A frequency distribution was generated for the reported volume data and goodness-of-fit tests were conducted to determine the type of statistical distribution that the data exhibited. The volume distribution was reasonably described with log-normal distributions.

Three additional parameters related to event frequency were investigated, and each was characterized with a single value for each state. The parameters were: 1) number of systems with events; 2) population served; and 3) number of events per year. Frequency distributions were generated for each of these parameters and goodness-of-

fit tests were conducted to determine the type of statistical distribution. Log-normal distributions were found to adequately describe each parameter.

Regional Characterization

The data for SSO frequency, volume, and cause were stratified by EPA region to determine if geographical differences in SSO characteristics existed. The cause of SSO events was found to vary significantly by region. Box and whisker plots were initially generated to provide a visual depiction of differences in volume by region. Analysis of variance (ANOVA) testing was then conducted to verify that significant differences existed between regions. Once these differences were confirmed, additional Bonferroni-adjusted ANOVA testing was conducted to identify individual differences between regions. Analyses were conducted on log-transformed data, consistent with the determination of log-normality discussed earlier. Based on these results, all subsequent analyses for cause category were conducted on a region-specific basis. State-specific cause information was used for the 25 states in EPA's data management system. Distributions for SSO cause were developed and applied by region to states without state-specific information in EPA Regions 1, 3, 4, 5, 9, and 10. One average cause distribution was used for Regions 6, 7, and 8 because the cause of SSO events reported by the states in these regions was very similar. Finally, none of the 25 states in EPA's data management system are in EPA Region 2, so an average cause distribution was developed and applied from the reporting states in Regions 1, 3 and 5. The cause distributions, by region, for non-reporting states are summarized in Table G.4.

Table G.4 Percentage of SSO Events, by Cause, by Region

Region	Blockage	Line Break/Misc.	Mechanical/Power Failure	Wet Weather – I/I	Unknown
1	41%	13%	10%	30%	6%
2	25%	12%	11%	36%	17%
3	36%	13%	7%	13%	30%
4	34%	11%	7%	12%	35%
5	8%	10%	16%	58%	8%
6	48%	9%	7%	21%	15%
7	48%	9%	7%	21%	15%
8	48%	9%	7%	21%	15%
9	69%	15%	6%	1%	9%
10	22%	13%	16%	6%	43%

Less significant regional differences were observed for SSO event frequencies and the volume of individual spills; therefore, subsequent analyses regarding SSO event frequencies and volumes were not stratified on a regional basis.

Analysis of Volume Discharged by Cause Category

The next data assessment step was to determine whether a relationship existed between volume discharged and cause of the SSO event. SSO volumes were found to vary significantly across most cause categories. Box-and-whisker plots were generated to provide a visual depiction of differences in volume by cause category. No significant differences in volume discharged were observed across the cause categories of Line Break and Miscellaneous. These two categories were therefore combined into a single category, as shown in Table G.4. ANOVA testing was conducted to verify that significant differences existed between the remaining categories. Once these differences were confirmed, additional Bonferroni-adjusted ANOVA testing was conducted to identify

individual differences between cause categories. Analyses were conducted on log-transformed data, consistent with the determination of log-normality discussed earlier.

Statistical Regression to Estimate SSO Frequency

In order to estimate national SSO frequency, the frequency data from the reporting states were extrapolated to estimate the non-reporting states. For this final data assessment step, a series of linear regressions were developed allowing event frequencies to be estimated for each state. Several regressions (both in linear and log-log space) were conducted to evaluate potential predictors of event frequency. The independent variables evaluated were:

- Number of SSSs in the state reporting at least one SSO event
- Total number of SSSs in the state
- Population served by SSSs in the state

Based upon these analyses, the two best predictors of event frequency were:

1) Log-log regression of number of SSO events per year as a function of the number of SSSs in the state with events, linked to a second regression of number of SSSs in the state reporting at least one SSO event with events, as a function of total number of SSSs in the state; and
2) Log-log regression of number of events per year as a function of total population.

Both methods provided a similar level of accuracy, explaining approximately 40 percent of the variability in the observed frequency data; therefore, both the system-based and population-based methods were used in generating the national estimates.

G.4 Assumptions

Estimating the national SSO frequency and volume from available data required a number of assumptions. The two primary assumptions that have the greatest potential to affect the estimates are:

- The degree to which reported SSO events in a specific time period represent all SSO events that occurred statewide during the same period; and
- The degree to which the number of SSSs in a state serves as a predictor of SSO frequency and volume, compared to population served.

To account for the uncertainty caused by these assumptions, separate analyses were conducted using a range of values. The range of results obtained from these alternative analyses helps to bound the uncertainty in the estimates generated.

Scenarios Evaluated

Different assumptions can be made regarding the degree to which the reported data represent statewide conditions. It could be assumed that each state's reporting data reflect all SSO events that occurred in that state. Alternatively, it could be assumed that the data reflect SSO events that occurred only for those communities that chose to report SSO events. It is not clear at this time which of these assumptions is most appropriate. To account for the uncertainty caused by these alternate assumptions, two separate scenarios were simulated:

- Scenario 1: Available reporting data reflect all SSO events that occured statewide
- Scenario 2: Available reporting data reflect events that occured for only those communities that chose to report

These scenarios were further evaluated using two different predictors of SSO frequency for states that do not track SSOs electronically:

- Predictor a: Event frequency based on total number of separate sewer systems in state
- Predictor b: Event frequency based on population served by separate sewers in state

The two pairs of assumptions discussed above result in four possible combinations of scenarios: 1a, 1b, 2a, and 2b. Scenario 2b, however, could not be evaluated as it requires data on population served by separate sewers in each municipal jurisdiction reporting at least one SSO event. These data were available only on a statewide basis. Consequently, only three scenarios were evaluated: 1a, 1b, and 2a.

G.5 Calculation Procedure

National estimates of SSO frequency and volume were generated by extrapolating the available data across the remaining 25 states (and Washington, D.C.) through the following five-step procedure:

1. Tabulate the total number of SSO events and the SSO volume for each of the reporting states.
2. Estimate the total number of SSO events per year for each non-reporting state.
3. Divide the total number of SSO events in each non-reporting state into different categories describing the cause of the SSO event.
4. Calculate the SSO volume for each cause category in each non-reporting state.
5. Calculate national estimates by summing across all states.

Step 1. Data Tabulation
Data tabulation was described earlier.

Step 2. Estimate total number of SSO events per year for each state
The total number of SSO events per state was calculated for all three scenarios. Scenarios 1a and 1b assume that the available reporting data reflect all SSO events statewide, so the reported frequencies were not adjusted when calculating frequencies for the non-reporting states. Scenario 2a assumed that the available reporting data reflect only those communities that choose to report. For Scenario 2a estimates, the expected frequencies were scaled upward to represent the ratio of separate sewer systems reporting SSO events to the total number of sewer systems. Non-reporting communities in each state were assumed to experience SSOs in a frequency distribution that matched the reporting communities.

Step 3. Divide total number of SSO events into respective cause categories
The initial data assessment calculated the relative frequency of the cause of SSO events by EPA region. Region-specific ratios are applied in Step 3 to define the number of events by cause category for each non-reporting state, as presented in Table G.4.

Step 4. Calculate SSO volume for each cause category
The initial data assessment defined SSO event volume as a function of cause. These region-specific cause and cause-specific volumes were applied in Step 4 to the frequency of events to define the total volume of SSO by cause category for each of the non-reporting states.

Step 5. Calculate national estimates

SSO event frequency and cause-specific volumes for each non-reporting state were generated in Steps 2 through 4. These estimates were combined with the data from the reporting states and were summed to provide a national estimate of SSO frequency and volume for each of the three scenarios examined.

G.6 Results

The results of the analyses for each of the three scenarios are summarized in Table G.5. These results indicate that the annual frequency of SSO events is between 23,000 and 75,000 events per year, with a corresponding volume of 3 to 10 billion gallons per year. The methodology used in developing this estimate results in an average volume per spill of approximately 125,000 gallons, while the average volume per spill in EPA's data management system is approximately 94,000 gallons per spill. This occurs as a direct result of a methodology that accounts for regional differences in the cause of SSO events. Of the six states with the highest populations and numbers of systems that did not provide data for this analysis, five (IL, NJ, NJ, OH, and PA) are in areas of the country where higher volume wet weather SSO events are more common.

The relatively large range is due to the uncertainty regarding the extent to which communities reporting at least one SSO event reflect all communities that had an SSO event, as the differences between Scenario 1a and Scenario 2a are much greater than the differences between Scenario 1a and Scenario 1b. Absence of a Scenario 2b does not seem to affect the overall range. As seen for Scenario 1, the population-based estimates are lower than the systems-based estimates. This suggests that Scenario 2a would be greater than Scenario 2b. It is important to note that the ranges provided in Table G.5 are not necessarily all-encompassing, as the estimates contain additional assumptions that could either raise or lower the values provided. For example, this estimate assumes that reporting communities report every SSO event that occurred within their community; that is, no SSO event went unnoticed and unreported. If reporting communities failed to report any number of their SSO events, the estimates provided in Table G.5 would underestimate the true frequency and volume of SSOs. The estimate also assumes that non-reporting communities have SSOs in the same frequency distribution as reporting communities; this assumption could over-estimate the frequency and volume of SSOs for non-reporting communities. There is no way to quantify the significance of these types of assumptions, but the uncertainty introduced by these assumptions is anticipated to be small compared to the variability already considered in the analysis. Further, EPA believes that the alternative assumptions are more likely to affect SSO event frequency, rather than volume estimates.

Table G.5 National Estimates of SSO Frequency and Volume

Scenario	Estimated Number of Events per Year	Estimated SSO Volume (billion gallons)
Scenario 1a– systems based	24,564	3.06
Scenario 1b– population based	23,103	2.85
Scenario 2a– systems based	74,813	9.74