TABLE 6. SLOPES CORRESPONDING TO THE INTERVALS SHOWN ON FIGURE 11

Interval	Slope Limit of Range
1	0.000100
2	0.003311
3	0.006521
4	0.009732
5	0.012943
6	0.016154
7	0.019364
8	0.022575
9	0.025786
10	0.028996
11	0.032207
12	0.035418
13	0.038629
14	0.041839
15	0.045050
16	0.048261
17	0.051471
18	0.054682
19	0.057893
20	0.061104
21	0.064314
22	0.067525
23	0.070736
24	0.073946
25	0.077157
26	0.080368
27	0.083578
28	0.086789
29	0.090000
30	> 0.090000

A final justification for the assumption that the pipe slopes are exponentially distributed is the following. The mean and standard deviation of the coefficients of variation of the slope for all 75 basins are 0.87 and 0.25, respectively. The mean value of 0.87 for the coefficients of variation very closely approximates the theoretical value of 1.0, characteristic of the exponential distribution.

This finding has significant importance in the practical application of the regression model developed in this study. This is especially true for cases where detailed statistical analysis of pipe slopes information is not available and only the mean value of the pipe slope is required to completely define with reasonable accuracy the pipe slope distribution.

The special slope parameters S_{PD} and $S_{PD/4}$ can then be estimated through this pipe approximation. Definition of the pipe slopes distribution using the exponential model is also important in delineating the extent and geographic location of the deposition loads in the system. This topic will be further discussed in Section 5.3.2*.

4.4.4 Average Pipe Diameter

The average pipe diameter of all segments within each collection system was computed by weighing circular diameters of each segments by its corresponding lenghts. An equivalent circular pipe diameter was first determined for non-circular sections before the weighted average was computed. The fraction of non-circular pipes was a small percentage of the total pipe length. All pipes are circular in the WRNDB system. The number of non-circular pipes represent about 5% of the total for the Dorchester system, whereas in Fitchburg they represent less than 1% of the total pipe lengths.

The formulas used for the equivalent circular diameter for the non-linear sections are presented in Table 7 and have been derived to yield roughly the same hydraulic radius at low depths of flow. For rectangular and U-shaped pipes a simple equivalence of total areas are indicated in Table 7. These forms are non-existent in the data for this study. The average pipe diameters determined for all 75 basins are presented in Table 8.

---

* It should be stressed here that in the generation of the slope data for the regression analysis described in Section 4.5 the assumption that the pipe slopes are exponentially distributed was not used. The probability distributions of slopes for all basins were determined from their respective slope data.

TABLE 7. FORMULAS FOR EQUIVALENT CIRCULAR DIAMETERS USED IN COMPUTING THE BASIN AVERAGE DIAMETER

Pipe Shape	Dimensions	Equivalent Circular D
Circular	D	D
Rectangular	X,Y	$(1.273\ X\ Y)^{1/2}$
Ellipse	X,Y	$(X\ Y)^{1/2}$
Egg	X,Y	$5.51X^{5.38}/(X+Y)^{4.39}$
Horseshoe	X,Y	$(X+Y)/2.0$
Oval	X,Y	$0.67\ X$
Ovoid	X,Y	$5.51X^{5.38}/(X+Y)^{4.39}$
Modified Circle	X,Y	$(X\ Y)^{1/2}$
Arch	X,Y	Y
U	X,Y	$(1.273\ X\ Y)^{1/2}$

TABLE 8.– SUMMARY DATA ON DERIVED LENGTHS, SLOPES AND PIPE DIAMETERS

BASIN NO.	80% DEPOS. LENGTH(FT) PLD	AVERAGE DIAM(IN) $\bar{D}$	AVERAGE SLOPE $\bar{S}$	SLOPE SPD	SLOPE SPD/4
1	15050.	10.5	0.027901	0.009965	0.003870
2	2080.	10.0	0.036746	0.022575	0.007320
3	434.	10.0	0.046065	0.058940	0.024900
4	450.	10.0	0.009118	0.008010	0.004290
5	1502.	10.7	0.021963	0.017470	0.006230
6	180.	10.0	0.008611	0.009732	0.009030
7	741.	10.8	0.016157	0.005849	0.001900
8	1019.	10.0	0.015280	0.014440	0.005050
9	36.	10.0	0.006400	0.006118	0.002200
10	560.	10.0	0.017163	0.019364	0.004560
11	304.	10.0	0.012293	0.004222	0.003480
12	685.	10.0	0.013985	0.009732	0.004290
13	6056.	10.9	0.013979	0.005996	0.002040
14	634.	10.8	0.019666	0.007681	0.003660
15	405.	10.0	0.030103	0.021642	0.010600
16	715.	10.0	0.004528	0.005510	0.000089
17	34879.	12.1	0.014539	0.006521	0.001830
18	1983.	9.5	0.009127	0.005119	0.001230
19	5938.	8.6	0.017344	0.005765	0.003430
20	83.	8.0	0.017852	0.017050	0.006270
21	171.	8.0	0.026175	0.032207	0.010400
22	648.	8.0	0.019257	0.014923	0.004780
23	1839.	8.0	0.021631	0.012943	0.004460
24	9228.	8.6	0.027730	0.009732	0.003660
25	4517.	8.7	0.019560	0.004989	0.001850
26	1379.	11.1	0.016939	0.009732	0.003810
27	5914.	11.0	0.016298	0.009732	0.001840
28	480.	12.0	0.004485	0.005012	0.002130
29	929.	10.2	0.025510	0.003311	0.001060
30	6005.	12.7	0.013617	0.005064	0.001910
31	526.	10.8	0.016448	0.006521	0.003280
32	7967.	8.7	0.016991	0.006344	0.004070
33	6903.	10.0	0.015655	0.006870	0.002180
34	3956.	11.6	0.017855	0.008449	0.002240
35	10954.	12.6	0.018203	0.006143	0.003460
36	2591.	13.3	0.014155	0.002892	0.000798
37	1461.	10.3	0.034669	0.004463	0.001100
38	936.	11.7	0.005660	0.005396	0.003380
39	1100.	12.0	0.030361	0.011499	0.004090
40	1913.	12.3	0.005365	0.004528	0.001230
41	1445.	9.3	0.027941	0.025786	0.007380
42	786.	12.0	0.037994	0.013749	0.004240
43	2540.	12.7	0.002575	0.002542	0.007100
44	1695.	12.0	0.029394	0.019364	0.004800
45	4869.	12.0	0.006341	0.004945	0.003330
46	5279.	14.1	0.024465	0.007763	0.004180
47	2893.	19.9	0.011737	0.003065	0.000841
48	2647.	12.4	0.025934	0.016154	0.006410
49	5293.	12.2	0.022566	0.001077	0.000344
50	571.	11.7	0.038857	0.047799	0.002560
51	1040.	11.4	0.035336	0.007467	0.002350
52	2295.	11.8	0.012463	0.011354	0.007200
53	1545.	14.6	0.015249	0.002514	0.000704
54	2530.	13.9	0.011539	0.007625	0.004020
55	929.	10.6	0.005995	0.005086	0.002170
56	1972.	11.2	0.041951	0.012943	0.000788
57	1461.	14.1	0.015366	0.002852	0.001370
58	1373.	13.1	0.010369	0.004675	0.001780
59	4847.	14.3	0.017349	0.009732	0.003930
60	2613.	12.7	0.024400	0.006521	0.003930
61	4934.	12.2	0.025037	0.005449	0.002820
62	6016.	12.6	0.055506	0.002802	0.000776
63	642.	12.1	0.022864	0.002143	0.000611
64	3408.	12.5	0.027835	0.009732	0.003800
65	1815.	12.9	0.029568	0.003311	0.001180
66	3434.	13.1	0.034241	0.012943	0.004610
67	2726.	14.3	0.079921	0.008435	0.001790
68	6498.	12.7	0.011681	0.008106	0.003280
69	2106.	12.0	0.018102	0.011068	0.004420
70	13312.	15.6	0.019405	0.006521	0.003020
71	26905.	13.3	0.015667	0.002787	0.000772
72	18773.	12.4	0.026287	0.008287	0.000310
73	13588.	12.4	0.002375	0.009732	0.004270
74	1306.	10.7	0.033799	0.006521	0.004020
75	4751.	13.6	0.029285	0.016154	0.005560

4.4.5 Distribution of Solids Deposited by Pipe Length

The cumulative distribution of solids deposited in the collection system versus the cumulative length of sewer pipe where the deposition occurs is the only exogenous information that the user will have to accept in applying this methodology. In other words, this distribution is the only information that cannot be derived from local data and can only be modified in a minor way by input from local conditions. These modifications will be covered later.

Some simplifications had to be made in deriving these distributions especially considering the great number of pipe segments (around 6000) that were numerically considered. The procedure established for each basin, as a function of the maximum and minimum estimated loads, a series of 200 intervals where loads of approximately equal values were accumulated, together with their corresponding pipe lengths. Whenever the basin had less then 200 pipe segments the number of intervals would be made equal to the number of pipe segments of the basin. The cumulative values of loads versus lengths were then computed for each of these intervals.

Plots of several cumulative probability functions for a few basins are presented in Figures 12 and 13. An overall curve computed from all 75 basins is also presented in Figure 13. A total of 5000 intervals were used, in computing the overall curve, yielding a smooth curve as can be seen from Figure 13. Plots in semi-log paper of the complementary values of the load probabilities, (1.-Fload) (in the log axis) versus the cumulative probability of pipe lengths, resulted in nearly straight lines, indicating that this distribution is also exponential.

It was observed by associating the curves in Figures 12 and 13 with their respective mean slopes that the average basin slope increases moving from the lower to the upper curves. This explains comparatively higher percentages of loads depositing in lower percentages of pipe lengths. Based on this observation, Figure 14 was prepared. In its preparation a smooth curve was drawn at approximately the middle of the range of curves from Figures 12 and 13 with mean slopes corresponding approximately to the range limits in the legend of Figure 14. With the mean pipe slope known for a given basin, Figure 14 can be used to estimate the percentages of pipe lengths corresponding to given percentages of total solids deposited. Further discussion on the applications of Figure 14 will presented in Section 5.3.3.

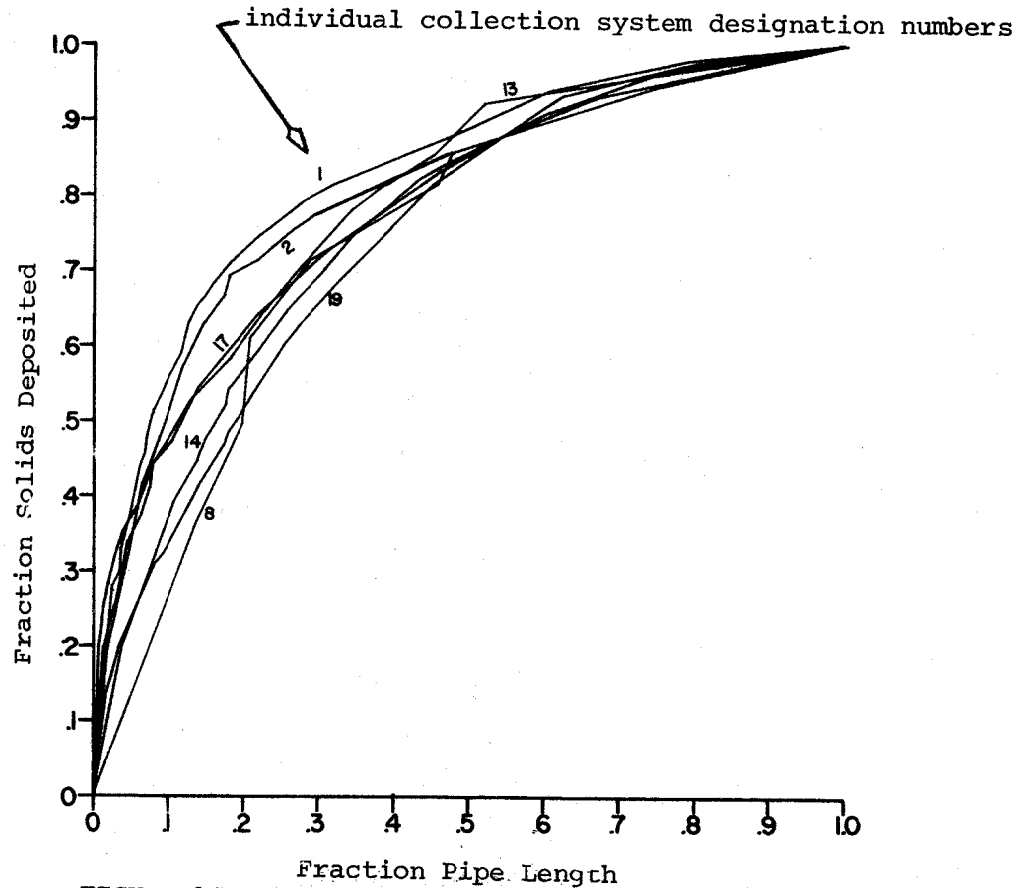

FIGURE 12. DISTRIBUTION OF SOLIDS DEPOSITED BY PIPE LENGTH-WRNDB SYSTEM

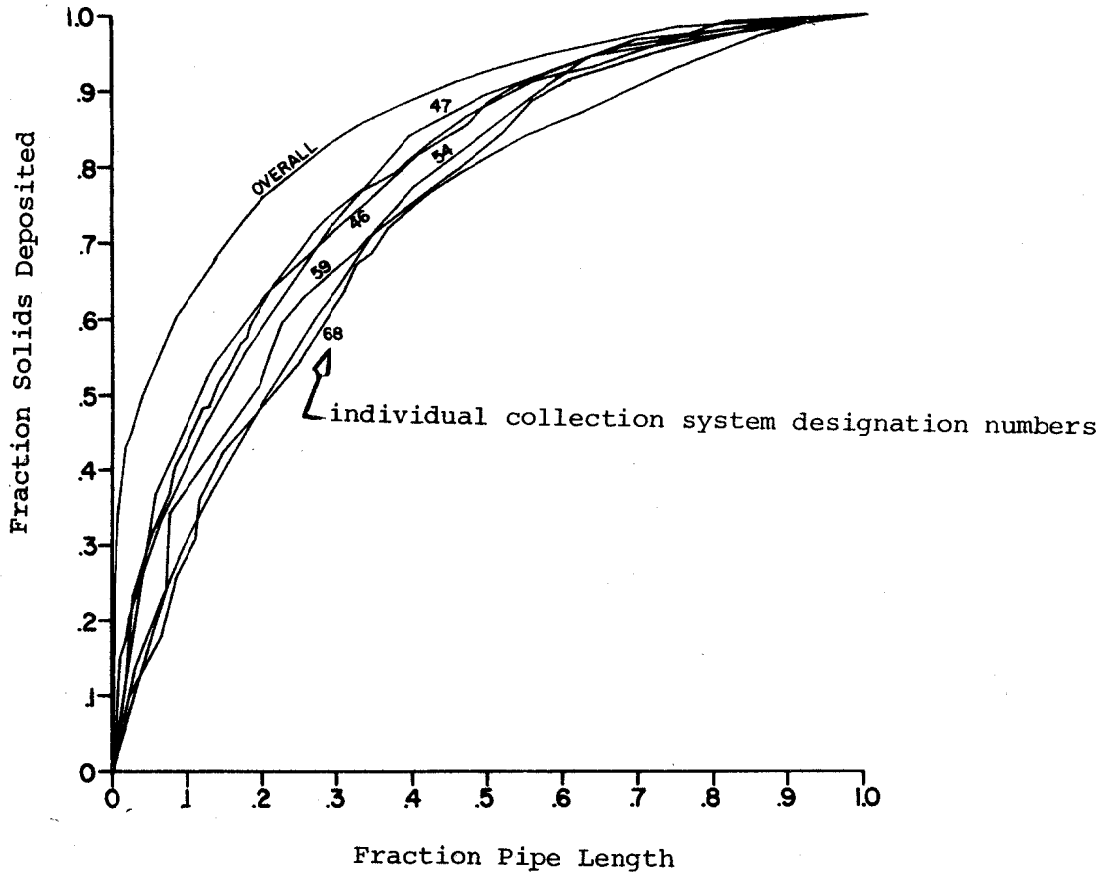

FIGURE 13. DISTRIBUTION OF SOLIDS DEPOSITED BY PIPE LENGTHS-DORCHESTER SYSTEM.

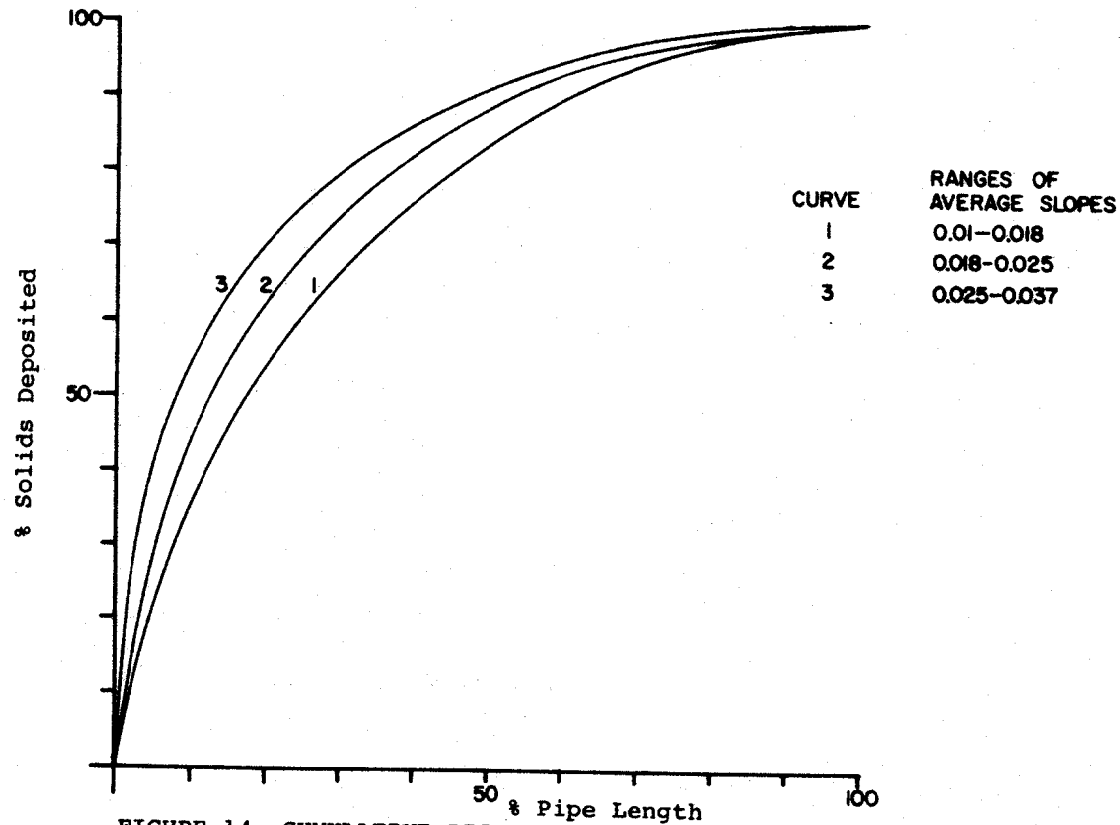

FIGURE 14. CUMULATIVE DISTRIBUTION OF SOLIDS DEPOSITED VS PIPE LENGTH

The choice of the pipe length corresponding to 80% of the loads deposited as an independent variable in the regression analysis discussed in section 4.2.1, resulted solely from the observation of higher separation of the curves around the value of 80%. Other percentages of mass deposited could have given the same results. It should be noted that the loads estimated by the regression equations presented in Section 4.5 correspond to 100% of the loads deposited. The length L_{PD} corresponding to 80% of the loads is only one regressor devised to explain the variation in the total loads.

4.4.6 Pipe Lengths Corresponding to 80% of the Loads Deposited - L_{PD}

The cumulative distribution of loads deposited versus the cumulative lengths of pipe for all basins were prepared. The values of L_{PD} to be used in the regression analysis were determined in the following manner. The computer printouts of the distributions were scanned for the percentages of pipe length (PL_D) corresponding to 80% of the total loads deposited and either read directly or more often they were interpolated. Those percentages were then applied to the total pipe lengths in the basins resulting in the L_{PD} values for all basins presented in Table 8.

4.4.7 Slope Corresponding to PL_D(S_{PD})

The determination of S_{PD}, the slope corresponding to the percentage of pipe (PL_D) where 80% of the total loads deposits is illustrated in Figure 2. The value of PL_D in the cumulative distribution of pipe slopes for a given basin is established and the corresponding slope value S_{PD} is determined. This was performed for all basins by reading directly or interpolating values in the tables of the cumulative distributions of the pipe slopes of these basins. The values of S_{PD} for all the 75 basins considered are given in Table 8.

There is no theoretical justification for the choice of S_{PD} as a regressor. As a matter of fact the average of the slopes ($\bar{S}_{PD}$) below S_{PD} had originally been thought of as a better regressor than S_{PD} itself, and was the first to be included in the regression analysis. A high correlation between S_{PD} and $\bar{S}_{PD}$ was observed and S_{PD} was included in the regression since its reduction of the sum of squares was slightly better than $\bar{S}_{PD}$. The variable $\bar{S}_{PD}$ was dropped from further consideration since its determination required far more effort than that for S_{PD}.

4.4.8 Slope Corresponding to $PL_D/4$ - ($S_{PD}/4$)

The determination of $S_{PD}/4$ is illustrated in Figure 2. First of all, multiply by 1/4 the percentage , PL_D, corresponding

to 80% of the loads deposited. Next enter that value in the cumulative distribution of pipe slopes for a given basin and determine the corresponding value, $S_{PD/4}$. This step was done by reading directly or interpolating values in the tables of the cumulative distributions of the pipe slopes for the basins. The values of $S_{PD/4}$ for all 75 basins are presented in Table 8.

The choice of $S_{PD/4}$ as an independent variable in the analysis regression did not involve any theoretical consideration. It's choice resulted from an "a priori" belief that representing the lower ranges of the pipe slope distribution would be significant in the regression analysis. It is shown in Section 4.5 that this assertion is true.

4.4.9 Summary of Results

Descriptions of how the data for the regression analysis was generated was presented in Sections 4.4.1 through 4.4.8 and of how the values of the dependent variable, TS, were generated by the use of a deposition model. The independent variables of the regression were described in Sections 4.4.2 through 4.4.8. The only remaining independent variable not described in those sections is the waste flow rate contribution in gpcd. This variable, as described in Section 4.3 and mentioned again in Section 4.3.2, was fixed at the values of 40., 110., 190., and 260. gpcd, covering, therefore, a wide range of flows.

To summarize, the values obtained for the independent variables are given in Tables 6 and 8. The tabular values of L and A are given in Table 6 and the values of L_{PD}, $\bar{D}$, $\bar{S}$, and S_{PD} and $S_{PD/4}$ determined for all 75 basins are given in Table 8. The results from the deposition model are lengthy and are not presented in this report.

4.5 Regression Analysis

4.5.1 Regression Procedures

The linear regression program used to empirically establish the relationships of the total daily suspended solids (TS) deposition within a sewerage collection system with the independent variables described in previous sections, is one that operates in a step forward manner. At each step in the analysis the particular variable entered into the regression equation accounts for the greatest amount of variance between it and the dependent variable, i.e., the variable with the highest partial correlation with the dependent variable. The program is flexible to allow any independent variable to be: (1) left free to enter the regression equation by a criteria of the sum of squares reduction; (2) forced into the regression equation; or (3) be kept definitely out of the regression equation in one given selection. The procedure permits examination of several alternative considerations of the independent variables by optional selections of variables to be forced

in and out of the regression equation or to be simply left free to enter the equation using variance reduction criteria. The relative importance of each variable in the regression equation can be ascertained using this procedure.

4.5.2 Regression Results

In this analysis various predictive models are analyzed relating total suspended solids deposition within a collection system with the aforementioned independent variables under the assumption of clean pipe conditions. These relationships are therefore applicable for situations in which the sewer piping system is properly maintained. The effects of age and improper maintenance on collection deposition loadings were examined and the results are presented in section 4.5.4. The degree of increased daily deposition resulting from improperly maintained systems was crudely simulated using several assumed depths of bottom sediments.

In this analysis both linear additive and multiplicative models were investigated. The untransformed observed values of the dependent and independent variables were initially used, leading to a strictly linear regression equation. In the second case the observed values of both the dependent and independent variables were transformed by taking their natural logarithms, leading to a linear equation in the logarithmic domain which can then be put into a non-linear multiplicative form.

The results of the regression analysis are presented in Tables 9 through 12 noting only those selections considered more significant. Tables 9 and 10 refer to the linear model regression results and Tables 11 and 12 refer to the multiplicative model logarithmic regression results. The sequence of independent variable inclusions for several selections using the linear form are given in Table 9. The symbols in the table are explained as follows. The letters DP under the variable TS indicates that TS is the dependent variable. The other letters under the independent variables for any selection indicate:

AV - the variable is available to enter the regression equation subject to the threshold variance reduction limit fixed at 0.0001 for all selections and all runs, implying that a variable will enter the regresson if it can further reduce the total sum of squares be at least 0.01%.

FI - the variable is to be forced in the regression although it is also subject to the 0.0001 limit noted above. FI implies priority over the AV variables in entering the regression.

TABLE 9. LINEAR REGRESSION RESULTS: MULTIPLE CORRELATION COEFFICIENT AND STANDARD ERROR OF THE ESTIMATE

Selection No.	TS	L	A	$\bar{S}$	$\bar{D}$	L_{PD}	S_{PD}	$S_{PD/4}$	Q	R
1		(1)*	(3)	(5)	(6)	(2)			(4)	0.906
	DP	AV	AV	AV	AV	AV	AV	AV	AV	
		0.776**	0.901	0.906	0.906	0.892			0.905	
		252.4***	170.9	167.4	167.6	177.9			167.8	
2		(1)	(2)		(4)				(3)	0.814
	DP	AV	AV	AV	AV	FO	FO	FO	AV	
		0.766	0.810		0.813				0.814	
		252.4	230.9		229.7				229.0	
3		(1)	(3)		(6)	(2)	(5)		(4)	0.905
	DP	AV	AV	FO	AV	AV	AV	FO	AV	
		0.766	0.901		0.905	0.892	0.905		0.905	
		252.4	170.9		168.3	177.9	167.9		167.8	
4		(1)	(3)		(5)	(2)		(6)	(4)	0.905
	DP	AV	AV	FO	AV	AV	FO	AV	AV	
		0.766	0.901		0.905	0.892		0.905	0.905	
		252.4	170.9		168.0	177.9		168.5	167.8	
5		(2)	(4)	(1)	(6)	(3)			(5)	0.906
	DP	AV	AV	FI	AV	AV	AV	AV	AV	
		0.765	0.902	0.013	0.906	0.892			0.906	
		253.2	170.5	392.7	167.6	177.9			167.4	
6		(3)	(5)	(2)	(7)	(4)		(1)	(6)	0.906
	DP	AV	AV	FI	AV	AV	AV	FI	AV	
		0.765	0.902	0.228	0.906	0.892		0.228	0.906	
		253.9	171.0	383.0	168.2	178.4		382.3	167.9	

* Order of entrance of the variable in the regression equation
** Multiple correlation coefficient
*** Standard error of the estimate
DP Dependent variable
AV Available to enter the regression equation
FI Forced in the regression equation
FO Forced out of the regression equation
R Multiple correlation coefficient

TABLE 10. COEFFICIENTS OF THE LINEAR REGRESSION EQUATIONS

Selection No.	TS	a	b	c	d	e	f	g	h	i
1	DP	42.030	0.022	1.317	-1428.22	5.472	-0.067			-0.421
2	DP	42.235	0.019	1.690		2.199				-0.421
3	DP	44.838	0.022	1.269		3.733	-0.066	-1168.40		-0.421
4	DP	32.760	0.022	1.279		4.309	-0.066		-1863.57	-0.421
5	DP	42.030	0.022	1.317	-1428.22	5.472	-0.067			-0.421
6	DP	47.612	0.022	1.314	-1392.30	5.172	-0.067		- 695.85	-0.421

DP = dependent variable

General equation: $TS=a+bL+cA+d\bar{S}+e\bar{D}+fL_{PD}+gS_{PD}+hS_{PD/4}+iq$

Units:

- TS - lbs/day,
- L - ft,
- A - acres
- $\bar{S}$ - ft/ft,
- $\bar{D}$ - inches,
- L_{PD} - ft,
- S_{PD} - ft/ft,
- $S_{PD/4}$ - ft/ft, and
- q - gpcd.

TABLE 11. LOG REGRESSION RESULTS: MULTIPLE CORRELATION COEFFICIENT AND STANDARD ERROR OF ESTIMATE

Selection No.	TS†	L†	A†	$\bar{S}$†	$\bar{D}$†	L_{PD}†	S_{PD}†	$S_{PD/4}$†	Q†	R†
1		(1)	(5)	(7)		(6)	(2)	(4)	(3)	0.974
	DP	AV	AV	AV	AV	AV	AV	AV	AV	
		0.873**	0.973	0.974		0.974	0.946	0.974	0.973	
		0.777***	0.363	0.364		0.364	0.518	0.363	0.369	
2		(1)	(4)	(3)	(5)				(2)	0.923
	DP	AV	AV	AV	AV	FO	FO	FO	AV	
		0.873	0.921	0.919	0.923				0.902	
		0.777	0.622	0.629	0.617				0.689	
3		(1)	(6)	(3)	(5)	(4)			(2)	0.939
	DP	AV	AV	AV	AV	AV	FO	FO	AV	
		0.873	0.939	0.919	0.938	0.936			0.902	
		0.777	0.554	0.629	0.555	0.564			0.689	
4		(1)		(4)	(6)	(5)		(2)	(3)	0.954
	DP	AV	AV	AV	AV	AV	FO	AV	AV	
		0.873		0.947	0.954	0.954		0.910	0.938	
		0.777		0.512	0.479	0.481		0.660	0.552	
5		(1)	(4)			(5)	(2)		(3)	0.973
	DP	AV	AV	FO	AV	AV	AV	FO	AV	
		0.873	0.973			0.973	0.916		0.973	
		0.777	0.367			0.368	0.518		0.369	
6		(1)			(5)	(4)		(2)	(3)	0.942
	DP	AV	AV	FO	AV	AV	FO	AV	AV	
		0.873			0.942	0.941		0.910	0.938	
		0.777			0.537	0.541		0.660	0.552	
7		(1)	(4)	(6)		(5)	(2)		(3)	0.973
	DP	AV	AV	AV	AV	AV	AV	FO	AV	
		0.873	0.973	0.973		0.973	0.946		0.973	
		0.777	0.367	0.369		0.368	0.518		0.369	
8		(2)	(7)	(1)		(6)	(3)	(5)	(4)	0.974
	DP	AV	AV	FI	AV	AV	AV	AV	AV	
		0.890	0.974	0.023		0.974	0.946	0.974	0.973	
		0.725	0.364	0.591		0.364	0.520	0.364	0.370	

continued

TABLE 11. (continued)

Selection No.	TS†	L†	A†	$\bar{S}$†	$\bar{D}$†	L_{PD}†	S_{PD}†	$S_{PD/4}$†	Q†	R†
9		(2)	(5)	(7)		(6)	(3)	(1)	(4)	0.974
	DP	AV	AV	AV	AV	AV	AV	FI	AV	
		0.910	0.974	0.974		0.974	0.947	0.566	0.974	
		0.620	0.363	0.364		0.364	0.515	1.311	0.363	
10		(3)	(7)	(2)		(6)	(5)	(1)	(4)	0.974
	DP	AV	AV	FI	AV	AV	AV	FI	AV	
		0.919	0.974	0.566		0.974	0.974	0.566	0.947	
		0.628	0.364	1.314		0.364	0.364	1.311	0.512	

* Order of entrance of the variable in the regression equation
** Multiple correlation coefficient
*** Standard error of the estimate
† Regression carried out with the natural logs of the variables
DP Dependent variable
AV Available to enter the regression equation
FI Forced in the regression equation
FO Forced out of the regression equation
R Multiple correlation coefficient

TABLE 12. COEFFICIENTS OF THE NATURAL LOG REGRESSION EQUATIONS

Selection No.	TS	a	b	c	d	e	f	g	h	i
1	DP	0.0028	0.932	-0.055	-0.059		-0.064	-0.770	-0.104	-0.510
2	DP	0.0013	1.180	-0.178	-0.418	0.604				-0.510
3	DP	0.0011	1.560	-0.109	-0.499	0.687	-0.508			-0.510
4	DP	0.0011	1.252		-0.396	0.398	-0.355		-0.368	-0.510
5	DP	0.0036	0.927	-0.077			-0.043	-0.884		-0.510
6	DP	0.0044	1.096			0.499	-0.239		-0.443	-0.510
7	DP	0.0032	0.952	-0.077	-0.040		-0.062	-0.859		-0.510
8	DP	0.0028	0.932	-0.055	-0.059		-0.064	-0.770	-0.104	-0.510
9	DP	0.0028	0.932	-0.055	-0.059		-0.064	-0.770	-0.104	-0.510
10	DP	0.0028	0.932	-0.055	-0.059		-0.064	-0.770	-0.104	-0.510

DP = dependent variable

General equation: $TS = aL^{b}A^{e}\bar{S}^{d}\bar{D}^{e}L_{PD}^{f}S_{PD}^{g}S_{PD/4}^{h}q^{i}$

Units:

- TS - lbs/day,
- L - ft,
- A - acres,
- $\bar{S}$ - ft/ft,
- $\bar{D}$ - inches,
- L_{PD} - ft,
- S_{PD} - ft/ft,
- $S_{PD/4}$ - ft/ft, and
- q - gpcd.

FO - means that the variable is to be kept out of the regression equation.

In selection number 1 shown in Table 9 all variables were free to enter the regression equation. The numbers above the selection symbols, AV, FO, and FI indicate the order in which the variables stepped into the equation. They reflect, therefore, the relative marginal importance of each variable present in the final equation of a given selection. Variables available to enter the regression shown without this ordering number are the ones that did not meet the 0.0001 variance reduction requirement.

The first number below the AV, FO and FI symbols is the multiple regression coefficient (adjusted for the degrees of freedom) achieved with the entrance in the regression of the present variable. The square of that number is the R^2 value at that step. For example, referring to selection 1 in Table 9, the first variable to enter the regression is the total pipe length in the basin L. At the first step the regression coefficient, R, is 0.776 implying $R^2 = 0.602$. When the second variable, L_{PD}, enters the regression the multiple regression coefficient improves to 0.892, explaining therefore 79.57% of the total sum of squares ($R^2 = 0.7957$). The successive reductions in the total sum of squares and the marginal contribution of each addition variable can be assessed by following through the order of the variables entering the regression. The second number below the AV, FI and FO symbols represent the standard error of the estimate, adjusted for the degrees of freedom. The square of that number is the variance of the estimate. The reduction in the uncertainty of the estimate can be noted by following that number throughout the sequence of variables introduced to the regression. An increase of the standard error at a given step indicates that the additional information realized by introducing the variable is off-set by the loss in degrees of freedom, implying that the regression equation is better off without that particular variable.

Other selections shown in Table 9 were designed to: (1) compare regression equations based on primary variables only, such as total length of pipe, L, total area, A, average pipe slope, $\bar{S}$, etc. and equations containing more refined variables, derived from primary data, such as L_{PD}, S_{PD}, $S_{PD}/4$; or (2) to assess the importance of a given variable or variables given that others were forced into the regression. Although several selections may have the same final equation, the order in which the variables entered the regression is useful to assess the importance of the individual variables. The last column in Table 9 indicates the maximum multiple correlation coefficient for a given selection. The linear regression is very good for all selections, with correlation coefficients ranging from 0.814 to 0.906 (R^2 from 0.663 to

0.821, respectively). The coefficients of the variables selected in Table 9 which entered the regression equation are given in Table 10. For each of the selections shown on Table 9 these coefficients define the corresponding regression equations using the general equation presented at the bottom of Table 10. A blank in Table 10 indicates that the corresponding variable does not belong to the regression equation of the given selection.

Similar results for the logarithmic regression are contained within Tables 11 and 12. The last column in Table 11 shows excellent results for the logarithmic regression in all selections, with a minimum multiple regression coefficient of 0.923 and a maximum of 0.974. This implies values of R^2 ranging from 0.852 to 0.949. The multiplicative model is superior to the linear forms explaining roughly 95 percent of the total variability of the dependent variable.

4.5.3 Alternative Model Selections

In this section several of the regression models described in Section 4.5.2 with all the relevant information presented in Tables 9 through 12 will be recommended for user application. Alternative forms reflecting the availability of data and/or user resources will be presented. The simplest forms require little data and have the least predictive reliability whereas the more complicated models, requiring greater user resources and data availability, provide estimates with extremely high reliability.

Comparison of the multiple correlation coefficients for the linear and the logarithmic regressions in Tables 9 and 11, respectively, indicates that the logarithmic multiplicative models are superior. Therefore, all the models selected below will derive from the logarithmic regression results.

4.5.3.1 The Elaborate Model

The highest multiple correlation coefficient, 0.974 (R^2 = 0.949) was obtained in selection 1 in Table 11. Inspection of that table indicates that this maximum correlation value had already been attained at the fourth regression step (under $S_{PD/4}$). The addition of a fifth variable, A, in the regression does not improve the multiple correlation coefficient nor the standard error of the estimate. Beyond the fifth variable the standard error increases. Therefore, only the first four variables should be retained in the regression equation yielding:

$$TS = 0.0038\ L^{0.8142}\ S_{PD}^{-0.8187}\ S_{PD/4}^{-0.1078}\ q^{-0.5098}\ (R^2=0.949) \qquad (2)$$

where TS is deposited solids loading in lbs/day, L is total length of sewer system in feet, S_{PD} and $S_{PD/4}$ are slope parameters defined in Section 4.4 and q is the per capita waste rate in gpcd. The above coefficients are slightly different from those shown on Table 12, which are for the full equation, that is, for all seven variables entered into the equation.

Utilization of equation (2) requires knowledge of total pipe length, the per capita contribution and the two pipe slope parameters, S_{PD} and $S_{PD/4}$. These slope parameters in turn are a function of the percentage of pipe where 80% of the total mass (TS) deposits (PL_D) and the probability distribution of the pipe slopes. The value PL_D is assumed given for the regression analysis, derived from the extensive computer analyses performed during this study and reported in Section 4.4.5. The probability distribution of the pipe slopes can either be derived from histograms computed from local pipe slope data or it can be defined with reasonable approximation from the mean pipe slope ($\bar{S}$) only, as described in Section 5.2.5.2. If the pipe slopes are not available a regression of mean ground slope and mean pipe slope, as also described in Section 5.2.3.2, could be performed.

4.5.3.2 An Intermediate Model

A simpler model than that given by equation (2), including only the primary variables, can be derived from selection 2 in Table 11 and is given by:

$$TS = 0.001303\ L^{1.18}\ A^{-0.178}\ (\bar{S})^{-0.418}\ (\bar{D})^{0.604}\ \bar{q}^{0.51}\ (R^2=.852) \qquad (3)$$

where $\bar{S}$ is average pipe slope, and $\bar{D}$ is the average equivalent diameter in inches; and A is service area (acres).

In this case the coefficients above are the same as shown in Table 12, since all variables that entered the regression were retained. All five independent variables present in equation (3) can be defined with good precision in any practical application. Neither the distribution of deposited loads versus pipe length nor the probability distribution of the pipe slopes are required. The mean pipe slope $\bar{S}$, again, can be correlated to the mean ground slope as described in Section 5.2.3.2, if no information on pipe slopes is available. A fitted equation for L as a function of A was derived from the data used in this study and is presented in Section 5.2.2

If the mean pipe diameter, $\bar{D}$, is eliminated from this regression the loss in precision of the estimate is not significant, resulting in the expression:

$$TS = 0.00389\ L^{1.2195}\ A^{-0.1866}\ (\bar{S})^{-0.4343}\ q^{-0.51}\ (R^2=0.848) \qquad (3)$$

4.5.3.3 The Simplest Model

The highest R^2 value that can be obtained with the least number of independent variables is also derived from selection 2 in Table 11. That is given by the regression equation:

$$TS = 0.0076\ L^{1.063}\ (\bar{S})^{-0.4375}\ q^{-0.51}\ (R^2=0.845) \qquad (5)$$

4.5.3.4 Final Comments on the Selection of the Model

The selection of the regression equations was meant to summarize the 16 possibilities presented in Tables 9 through 12. Other choices can be made from other selections presented, that may be more compatible with particular applications, although at lower values of R^2. It should be noted that the coefficients presented in Tables 10 and 12 apply only to the complete regression equations, and if one variable is deleted from the equation those coefficients do not remain the same.

The user should note that the parameters of the regression equation estimated by the least squares procedure are a function of the data used in their estimation. These parameters are not known with certainty and represent only estimates of the true parameters of the model. The estimation of the parameter is improved as the number of data points increase and as the spread or variance of the independent variables also increases.

It is known from regression theory that the procedure provides the best estimates (least variance) around the means of the independent variables used in computing the parameters of the regression equation. As the values of the independent variables depart from the mean the uncertainty of the estimation given by the equation increases and can become large outside the range of values used as data in the regression. In other words. the uncertainty of the estimate given by the regression equation may be large when the model is used for conditions requiring extrapolation beyond the range of data used to develop the model.

The values of the independent variables used in the regression analysis for all 75 basins were presented in Tables 6 and 8. The range, mean and standard deviations of the independent variables were computed from the data given in Tables 6 and 8 and are presented in Table 13. A total of 300 observations were used in the regression analysis since the deposition loadings for each of the 75 basins were computed for four per capita waste generation rates.

4.5.4 Effects of Age and Maintenance

The regression equations presented in Section 4.5.3 were derived from deposition data computed under the assumption of

TABLE 13. RANGES, MEANS AND STANDARD DEVIATIONS OF THE INDEPENDENT VARIABLES USED IN THE REGRESSION

Variable	Range	Mean	Standard Deviation
L (ft)	263.-134528.	13702.	22867.
A (acre)	3.-641.	76.	102.
$\bar{S}$ (ft/ft)	0.002375-0.079921	0.021	0.0126
$\bar{D}$ (in)	8.-20.	11.5	1.97
L_{PD} (ft)	36.-34879.	4026.	5811.
S_{PD} (ft/ft)	0.001077-0.058940	0.0101	0.0093
$S_{PD/4}$ (ft/ft)	0.000089-0.024900	0.0037	0.0033
q (gpcd)	40.-260.	*	*

* q is not a random variable

clean pipes, with no bottom sediments. In this section the impact of poorly maintained systems was crudely simulated by arbitrarily assuming various levels of prior sediment accumulation in the pipes. These sediment levels would change the bottom cross-sectional shape of the pipe channel and hence the depth of flow, the hydraulic radius and the shear stress characteristics.

Two cases simulating different degrees of maintenance other than perfect clean pipe conditions were considered. In the first case, or the intermediate maintenance category, sediment beds ranging from 1 to 3 inches in depth were assumed for all pipes with slopes less than 0.0075. A sediment bed of 3 inches was assumed for all pipes with slopes less than 0.0005. The bed depths then ranged linearly starting at 3 inches for a pipe slope of 0.0005 up to one inch for a pipe slope of 0.0075. This range was established using judgment and also based on visual inspection of numerous combined sewer laterals in eastern Massachusetts sewerage systems. In the second category of maintenance, the zero maintenance case, sediment beds ranging from 3 to 6 inches for the same range of slopes was considered.

Considering the two age and maintenance criteria mentioned here, the deposition model was used to estimate total deposition loadings for each of the 75 sewer systems for each of the four per capita waste generation rates of 40, 110, 190 and 260 gpcd. Before similar regression computations were performed on the deposition results obtained for pipes with bottom deposits, a comparison was made of the total deposited loads computed under the assumptions of clean and sedimented pipes.

For each basin the ratios of TS computed for sedimented pipes with sediment beds of 1-3 inches and 3-6 inches and the TS values for clean pipes were calculated for all four per capita waste rates considered, i.e., 40, 110, 190 and 260 gpcd. The resulting ratios were very stable for a given per capita waste rate for both cases of sediment deposits. The mean and coefficient of variation of these ratios are presented in Table 14 for both conditions of bottom deposits.

The results shown on Table 14 suggest that the prediction of TS in sedimented pipes could be accomplished by a simple functional multiplicative correction of the results given by any of the regression equations for clean pipes. An equation was fitted using the data of Table 14 for each of the bed deposit conditions.

These equations are:

- For a system with deposits ranging from 1 to 3 inches;

$$TS_{1-3\ inches} = 1.68\ q^{-0.076}\ TS(clean) \quad (R^2=0.988) \qquad (6)$$

TABLE 14. AVERAGE VALUES OF THE RATIOS OF COMPUTED LOADS IN DEPOSITED PIPES OVER CLEAN PIPES

Ratios	Average Values of Ratios for Per Capita Waste Rates			
	40	110	190	260
$TS_{1-3\,inches}/TS(clean)$	1.263	1.186	1.128	1.094
	(0.18)	(0.14)	(0.07)	(0.12)
$TS_{3-6\,inches}/TS(clean)$	1.312	1.211	1.151	1.121
	(0.14)	(0.11)	(0.09)	(0.09)

The numbers in parenthesis indicate the coefficient of variation of the ratios.

where q = flow per capita, and TS(clean) = load of total solids computed from any of the regression equations presented in Section 4.5.2

- For a system with deposits ranging from 3 to 6 inches:

$$TS_{3-6\ inches} = 1.79\ q^{-0.084}\ TS(clean)\quad (R^2=0.999) \qquad (7)$$

The R^2 values indicated above refer to the regression of the ratios of TS on the values of flow per capita.The small difference found between the two conditions of bottom deposits may well be the result of an inappropriate accounting of these factors by the deposition model. On the other hand it may simply have resulted from the particular combination of pipe diameters and sediment depths used as data, which may have led to actually small differences in flow depths above the sediment levels, and therefore small differences in shear stress between the cases.

4.5.5 Predictions of Other Parameters

In Appendix A some of the results from the first phase of the sewer flushing experiments are presented. In Table A-1 of the appendix the results of 83 flushing experiments performed at 4 different sites in the Boston area are presented indicating the mass removals per flush of various pollutants including BOD, COD, TKN, NH3, P, TS and VSS. The values represent estimates of the various pollutant masses removed by flushing as well as the total mass of depositing pollutants. Those results can serve as a basis for the estimation of other parameters associated with the total solids that deposit in the system.

A regression was performed between TS and each one of the other 6 indicators presented in Table A-1. The resulting regression equations are presented in Table 15, with their associated correlation coefficients.

Estimates of the total daily BOD, COD, TKN, NH3, P and VSS depositing loads within a given collection system can be made using the regression equations in Table 15 with the predicted TS loadings calculated from any of the regression equations presented for clean pipe conditions in Section 4.5.3 and the bias correction factors for pipes with sediment beds given in Section 4.5.4.

TABLE 15. REGRESSION OF DIFFERENT POLLUTANTS ON TS

Regression Equation*	Correlation Coefficient
$BOD = 0.344TS^{1.308}$	0.80
$COD = 0.875TS^{1.04}$	0.77
$TKN = 0.039TS^{1.135}$	0.67
$NH_3 = -0.0336 + .017TS$	0.44
$P = -0.006 + .00768TS$	0.67
$VSS = 0.689TS^{1.033}$	0.97

* Units are lb/day

SECTION 5

MODEL UTILIZATION

5.1 Introduction

This section summarizes the steps required by the user to achieve the objectives set forth in Section 4.1.1, which are to define on a preliminary basis: a) the total amounts of solids and other pollutants that deposit in the sewerage system; and b) the extent of the collection system over which the deposition takes place.

The nature of the available data and/or the degree of resources the user can commit will define different work programs which in turn will lead to different levels of confidence in the results. In order to avoid references to past sections, all necessary formulas will be summarized in Sections 5.2.1 to 5.2.6.

5.2 Summary of Formulas for the Estimation of TS

This section summarizes all necessary formulas for the estimation of TS, separating formulas already presented in Section 4 and introducing a few new formulas. Only variable symbols introduced here will be defined in the following sections.

5.2.1 Formulas for Estimation of Total Loads Deposited - TS*

5.2.1.1 The Elaborate Model

The elaborate model derived in Section 4.5.3.1 for clean pipe conditions is given by:

$$TS = 0.0038\ L^{0.8142}\ S_{PD}^{-0.8187}\ S_{PD/4}^{-0.1078}\ q^{-0.5098}\quad (R^2=0.949) \qquad (8)$$

---

* These equations are for predicting deposited solids for clean pipe conditions. The user is referred to Sections 4.5.3 and 4.5.4 for age and maintenance considerations and for the prediction of other pollutants.

5.2.1.2 Intermediate Model

The intermediate models in Sections 4.5.3.2 for clean pipe conditions are given by:

$$TS=0.001303\ L^{1.18}\ A^{-0.178}\ (\bar{S})^{-0.4343}\ q^{-0.51}\ (R^2=0.852) \quad (9)$$

$$TS=0.00389\ L^{1.2195}\ A^{-0.1866}\ (\bar{S})^{-0.4343}\ q^{-0.51}\ (R^2=0.848) \quad (10)$$

The difference between the two models is that the second neglects the average pipe diameter $\bar{D}$, with a small loss in precision.

5.2.1.3 The Simplest Model

The simplest model defined in Section 4.5.3.3 for clean pipe conditions is given by:

$$TS=0.0076\ L^{1.063}\ (\bar{S})^{-0.4375}\ q^{-0.51}\ (R^2=0.845) \quad (11)$$

5.2.2 Estimation of Total Pipe Length

The total pipe length of the system, L, and its corresponding collection area, A, are generally assumed to be known. In cases where this information is not known and where crude estimates will suffice, the total pipe length can be estimated from the total basin area using the expressions:

$L = 168.95\ A^{0.928}$ (R^2= 0.821)-low population density (10-20 people/acre)

$L = 239.41\ A^{0.928}$ (R^2= 0.821)-moderate-high population density (30-60 people/acre)

These expressions were derived from the data for the 75 basins used in this study, yielding a correlation coefficient of 0.906. Since detailed population density estimates for each collection system in the data set were not available, the following approach was used to account for the effect of population density on length of sewer. The data for the 75 basins were separated into two distinct data sets representing low population densities (the 35 basins from the West Roxbury-Newton-Dedham-Brookline sewerage system) and moderate to high population densities (the latter 40 basins from the Dorchester and Fitchburg sewerage systems). A dummy variable approach was used in the regression analysis to obtain the two expressions given in equation (12). A more explicit treatment of population density would be possible if detailed population density estimates per collection system were available.

5.2.3 Estimation of Mean Pipe Slope $\bar{S}$

5.2.3.1 Pipe Slope Data is Available

In this case:

$$\bar{S} = \frac{\sum_{i=1}^{n} S_i l_i}{\sum l_i} \qquad (13)$$

where: S = slope of segment i

l = length of segment i

n = total number of pipe segments

5.2.3.2 Pipe Slope Data Not Available

If data on pipe slope is not available the user will determine a mean value for the ground slope using any procedure, such as the techniques of uniform grid sampling or random sampling. With the mean ground slope value, determine the mean pipe slope by:

$$\bar{S} = 0.348\ (\bar{S}_G)^{0.818} (R^2=0.96) \qquad (14)$$

where: $\bar{S}_G$ = mean ground slope (ft/ft)

The expression above resulted from a regression of mean ground versus mean pipe slope for all 75 basins of this study. The resulting correlation coefficient was 0.98.

This relation was derived in the following manner. Topographic mylar overlays with ground contour at ten feet intervals were developed for the WRNDB and the Fitchburg sewer system service areas. Similar maps existed for the Dorchester and South Boston sewer systems. Using the topographic maps the collection system areas were subdivided into smaller fractions considered to have uniform slopes. The ground slopes were computed for each of the subareas. Those ground slopes were then associated with the percentages of the collection system area they covered. The basin average ground slope was then computed by weighing the individual slopes by their associate subarea percentages. Ground slopes in the smaller basins were determined by using one or two subareas whereas the ground slopes in the large basins were determined using 10 to 21 subareas.

5.2.4 Distribution of Total Solids Deposited by Pipe Segment - Determination of PL_D

The distribution of total solids deposited by pipe length is presented in Figure 14. The user can determine PL_D corresponding

to 80% of the solids deposited by entering the graph with "% of mean deposited" equal to 80% and, from the curve that best appropriate the average pipe slope $\bar{S}$, read in the horizontal axis, "% pipe length", the value of PL_D. This value will be required to determine S_{PD} and $S_{PD/4}$ in Section 5.2.5.2

5.2.5 Determination of Slopes S_{PD} and $S_{PD/4}$

5.2.5.1 Pipe Slope Data is Available

The collection system pipe slopes can be directly used to define the cumulative pipe slope distribution. In this study the pipe lengths (as integer values) associated with the observed slopes were taken as their frequencies. The user should adopt the same approach for compatability with the equations for TS derived in this study. Simplified methods may also be used, with less accurate results. If, for example, the slopes are not weighted by their lengths, this is equivalent to assuming that all pipe segments have the same length, which in some cases may be a reasonable assumption.

5.2.5.2 Pipe Slope Information Not Available

The average slope $\bar{S}$ can be computed using either equation (13) or equation (14) to define the pipe slope cumulative density function (CDF) as follows:

$$FS = 1-e^{-s/\bar{S}} \qquad (15)$$

If the probability F_S is known, with a fixed $\bar{S}$ the corresponding slopes can be computed, and vice-versa. For the basins used in this study, equation (15) was verified with excellent results.

To determine the slope S_{PD}, use the value of PL_D determined in Section 5.2.4, and make:

$$F_S=1-e^{-s/\bar{S}}=PL_D$$

and determine s. Then:

$$S_{PD}=s$$

To determine the slope $S_{PD/4}$, make:

$F_S=1=e^{-s/S}=1/4PL_D$ and determine s. Then:

$$S_{PD/4}=s.$$

5.2.6 Formula for the Average Pipe Diameter

If the system contains non-circular shapes, first compute their equivalent circular diameters using the equations for

equivalent pipe diameters presented in Table 7. Then, compute the average diameter by:

$$D = \frac{\sum_{i=1}^{n} D_i l_i}{\Sigma l_i} \qquad (16)$$

where: D_i = diameter of segment i

l_i = length of segment i

n = total number of pipe segments.

5.3 General Description of User's Steps

5.3.1 Determination of the Total Solids Deposited

The generalized procedure for obtaining total solids deposited estimates is presented in Figure 15 and should be referred to in the ensuing discussion.

The first question to be answered at the onset is whether pipe slope data are available. If the answer is positive the user should use equation (13) to determine the average pipe slope $\bar{S}$; otherwise, a mean ground slope should be determined as described in Section 5.2.3.2 and then equation (14) should be used to determine $\bar{S}$.

The next question that arises is whether the total pipe length of the collection system is known. In the negative case the total area of the collection system is assumed known and the total pipe length is estimated by equation (12). The per capita waste rate including infiltration is then established.

The selection of the equation for the estimation of the solids deposited, TS, follows. If the simplest model is desired the user has at this point all the elements required by equation (11) to compute TS. If the intermediate model is chosen and the pipe diameter information is available the mean pipe diameter is computed by equation (16). All information is computed to estimate TS from equation (9). If pipe diameters are not available then use equation (10). If the elaborate model is chosen, the value of PL_D must be determined using Figure 14 as described in Section 5.2.4. The next variables to be determined are S_{PD} and $S_{PD/4}$.

If no pipe slope information is available the user has no alternative but to use the exponential approximation for the pipe slope cumulative distribution. If the pipe data are available but the user considers the exponential approximation satisfactory he may use equation (15), associated with PL_D and $\bar{S}$