Chemical and Microbiological Quality of Runoff Into and Out of Dry Wells; A Case Study in Millburn, NJ

Leila Talebi[1], Robert Pitt[2]

[1] PhD Candidate, Department of Civil, Construction and Environmental Engineering, The University of Alabama, Tuscaloosa, Alabama. Email: ltalebi@crimson.ua.edu

[2] Cudworth Professor, Director of the Environmental Institute, Department of Civil, Construction and Environmental Engineering, The University of Alabama, Tuscaloosa, Alabama Email: rpitt@eng.ua.edu

Abstract

Since 1999, the city of Millburn has required dry wells to accommodate additional flows from newly developed areas. This new requirement to divert the increased roof and impervious surfaces runoff to dry wells was established to mitigate local drainage and water quality problems. One of the components of this EPA funded project was monitoring stormwater flows from individual properties to the dry wells and associated exiting subsurface waters beneath several feet of soil beneath the dry wells. Bacterial and chemical water quality was measured during ten rain events in three different dry wells and at one water storage cistern location. The dry well findings reported in this paper indicate that the dry wells did not significantly change any of the water quality concentrations for the observed stormwater constituents. The cistern system did result in significant reductions in bacteria levels.

Introduction and Background

A dry well is a subsurface discharge location for the disposal of stormwater. Some storage is provided that can usually handle short periods of very intense runoff rates, resulting in decreased overflows. Their main function is to infiltrate stormwater to relatively shallow depths. The decreased surface discharges result in reduced in increased runoff rates and volumes. Typical dry wells in Millburn, NJ, are 4 ft diameter, 6 ft tall perforated concrete chambers surrounded by 2 feet of gravel on all sides, including below the chamber. They also have a surface overflow that directs excess flows into the storm drainage system. The total depth of the dry well system is therefore about 10 ft. State requirements specify that the subgrade soil permeability rate must be sufficient to drain the stored runoff within 72 hours (NJ Stormwater BMP Manual, pp 9.3-1).

The purpose of this EPA supported project was to investigate the hydraulic performance of the dry wells along with any water quality changes associated with the dry well operations. The majority of the dry wells examined during this study received runoff from roofs, while some also received runoff from surrounding paved driveway and parking areas, and from landscaped areas. This paper is mostly excerpted from the recent EPA report: Pitt, R. and Talebi, L. *Evaluation and Demonstration of Stormwater Dry Wells and Cisterns in Milburn Township, New Jersey*, Prepared for the Urban Watershed Management Branch, U.S. Environmental Protection Agency, 2012 (Pitt and Talebi, 2012).

Methods and Materials

Three dry wells were constructed with a shallow monitoring well directly beneath the concrete chamber (sampling water similar to the water in the dry well tank), along with a deep monitoring well located at least 60 cm (2 ft) beneath the deepest depth of the seepage pit gravel. Water samples were collected from these two dry wells for comparison. In addition, a new water storage cistern was also sampled at the inlet and from the outlet for comparisons. Water samples were collected after ten storm events from the monitoring wells beneath the dry wells and from the cistern and all samples were analyzed in duplicate. Many other dry wells were also monitored for hydraulic performance and to measure the infiltration rates, including comparisons to traditional infiltration equations. The full results from these measurements and tests are included in the full EPA report (Pitt and Talebi 2012).

Results and Discussion

The following is a summary of the results for each measured primary constituent group. Only samples from three of the four locations were available for the first event, and cistern samples were not obtained during the second and fourth events. Therefore, seven to ten water quality samples were available from each sampling location for the study. The samples were analyzed in laboratories of the University of Alabama for bacteria (total coliform and *E. coli*), total nitrogen (TN), nitrate plus nitrite (NO_3 plus NO_2), total phosphorus (TP), and chemical oxygen demand (COD). Lead, copper, and zinc were analyzed at a commercial laboratory (Stillbrook Environmental Testing Laboratory in Fairfield, AL). Selected samples were also analyzed for pesticides by the EPA (not reported here). Statistical analyses and plotting of the data were conducted using MINITAB, and MS-Excel software. Table 1 is a summary of concentrations for the influent and effluent samples

Bacteria

The upper detection limit (UDL) of the IDEXX method using straight un-diluted samples is 2,419.2 MPN/100 mL and the lower detection limit (LDL) is 1 MPN/100 mL for both indicator organisms. After completion of the first two rounds of sampling, it was observed that most bacteria levels exceeded the UDL (even with the 24 hr maximum delay, necessitated by sample shipping, that was longer than the desired standard 6 hr holding time). Therefore, one of the samples for each site was also diluted 10 times to increase the UDL to 24,192 MPN/100 mL. For some samples, 20 times dilution was applied to increase the UDL to 48,384 MPN/100 mL. The cistern related sample bacteria levels (especially the outlet samples) were generally lower than for the dry well samples. Total coliform levels are higher in the cistern than the inflow and generally the deep locations at each of the dry well sites had higher levels of total coliforms, possibly indicating some re-growth in the systems.

Nutrients

The total nitrogen concentrations (HACH total nitrogen) (reported as N) ranged from <1 to 16.5 mg/L. The NO_3 plus NO_2 concentrations (HACH Accu-Vac method) ranged from 0.2 to 3.2 mg/L. The total phosphorus concentrations (HACH test-n-tube) ranged from 0.02 to 1.36 mg/L. The median values for most of the locations were very similar for both the shallow and the deeper monitoring well samples and for the inflow and cistern samples, except for one of the sites in which the deeper samples have higher TN median values than for the shallow samples.

However, as shown later, these differences were not significantly different, based on the number of samples available.

Chemical Oxygen Demand (COD)

The COD concentrations (HACH test-n-tube) ranged from 5.0 to 148 mg/L. Also, as shown later, the statistical analyses did not indicate any significant differences between the shallow and deep samples for any location (or for the inflow vs. cistern samples), for the number of samples available.

Metals

Total forms of lead, copper and zinc were analyzed for each sample. The detection limits were 5 μg/L for lead, 20 μg/L for copper, and 20 μg/L for zinc. There were many below detection limit (BDL) values reported. The maximum observed concentration for lead was 380 μg/L which occurred in a deep monitoring well sample under a dry well. The maximum observed concentration of copper was 1,100 μg/L which occurred in a cistern influent sample (possibly due to copper roof gutters on the home). The concentrations of zinc in all samples ranged from BDL to 140 μg/L. The statistical analyses did not detect any significant differences between any of the paired heavy metal values, based on the number of samples available.

Statistical Analyses

A number of complementary statistical analyses of the water quality data were conducted using MINITAB and MS-Excel software, including: log-normal probability plots, Anderson-Darling (AD) p test for normality, group box plots, paired line plots, time series plots, Mann-Whitney comparison tests, and paired sign test (metals only).

Log-normal Probability Plots and Anderson-Darling Test Statistic

Log-normal probability plots were used to identify the range, randomness, and normality of the data and to determine what type of statistical comparison tests can be used for each data set. Figure 1 shows an example of grouped log-normal probability plots (using Minitab) for total nitrogen at one of the locations. A Minitab plot option includes the Anderson-Darling (AD) test statistic. In the AD test, the null hypothesis is that data follow a normal distribution (log-normal for these data as the data are plotted after log transformations). If the p-value is not less than the chosen level of 0.05, there is insufficient evidence to reject the null hypothesis, therefore the data fit the normal distribution. On the other hand, if the p-value is less than the chosen level of 0.05, the hypothesis would be rejected, thus the data do not follow a normal distribution. In this example, the AD p value are both larger than 0.05, indicating that the data likely is normally distributed and parametric statistical tests can be used without losing power.

In this study, the log-normal probability plots are shown for inflow vs. cistern, for the cistern and deep vs. shallow monitoring wells for each dry well sampling site. The log-normal probability plots indicate that most of the data groups are seen to overlap within the limits of the 95% confidence limits, indicating that the data are likely from the same population.

Group Box Plots

The group box plots show the data for all of the sites and (non-metal) constituents. There are no apparent visual trends between any of the paired data. Figure 2 shows a sample group box plot for total nitrogen. The boxes represent the 25th and 75th percentiles on the box ends along with the median in the line within the box. The whiskers represent the 5th and 95th percentiles and stars are outside of these limits. In this example, some of the sampling locations show relatively narrow concentration ranges (135 for example), while others show a much greater variation (18 for example). The paired data rarely show any decreased ranges or concentrations for the deep samples.

Paired Line Plots

Paired line plots (not included in this paper) showed that the concentration values varied with no consistent pattern: in some cases shallow samples may have higher bacteria levels or nutrient levels as well as COD levels, while during other storms, the deep samples may experience higher values.

Mann Whitney Test

The Mann-Whitney test, also called the rank sum test, is a nonparametric test that compares two unpaired groups. Nonparametric tests are preferred when the values are not normally distributed, or the distribution is unknown or mixed (as in this case). We selected this test because not all of the constituents at all sites were normally distributed and the use of a single test is preferred to mixing. The Mann Whitney test was performed using MINITAB to test if the shallow samples have significantly higher or lower concentrations than the deep monitoring well samples (same comparison test for inflow vs. cistern). To make sure that the populations have the same shape, over-laying probability plots were made for the two pairs of data in the previous probability plots. In all the cases, the straight lines were very close to each other and the bandwidths were quite similar. Therefore, the distributions can be reasonably assumed to be the same shape. Table 2 shows the output obtained using MINITAB for comparison between the paired data. Except for the bacteria and COD results for the cistern site, all paired sample sets did not indicate significant differences for these numbers of samples at the 0.05 level. The cistern median total coliform values were greater than the inflow median values, indicating possible re-growth; however, the median *E. coli* and COD cistern values were less than the inflow values.

Paired Sign Test for Metal Analyses

Due to the presence of large numbers of below-detection values for the metal analyses, a simple paired sign test was used to compare each paired set of data. In the paired sign test, the null hypothesis is that the population medians are similar. In each pair of observations, a comparison was made to determine if there is an increase from the shallow sample to the deep sample or if there was a decrease. The advantage of the sign test is that if one part of the pair of data is not detected, while the other is, it is still possible to determine which is larger. However, if both data parts in the pair are not detected, it is not possible to determine which is larger and that pair is ignored in the calculations. If the calculated p value is less than 0.05, then the null hypothesis will be rejected and the data are assumed to originate from different sample populations. Table 3 describes a summary of results from paired sign tests for lead, copper and zinc. No statistically significant differences are seen between the sample sets for these heavy metals for the numbers of samples available.

Conclusion

Shallow and deep samples collected beneath three dry wells and samples at the inflow and in the cistern during ten storm events were analyzed for total coliforms, *E. coli*, total nitrogen, NO_3 plus NO_2, total phosphorus, COD, lead, copper, and zinc. Three samples were also analyzed for pesticides and herbicides. Statistical analyses indicated that the differences in water quality between the shallow and the deep samples were not significant (p values were > 0.05). However, significant differences were found ($p < 0.05$) between the quality of inflow samples and cistern samples for total coliforms (increased values possibly indicating re-growth), *E. coli*, and COD (reduced values).

These findings indicate that the dry wells did not significantly change any of the water quality concentrations for the stormwater constituents observed. If the influent water quality is of good quality, the dry wells can be a safe disposal method for stormwater quality. However, the bacteria and lead concentrations exceeded the groundwater disposal criteria for New Jersey and may require treatment, if the aquifer is critical. The deep monitoring well sample was located at least 1.3 m (4 ft) below the bottom of the dry well (which itself was about 8 ft beneath the ground surface), more than the typical spacing requirement (3 ft) to groundwater. This distance was not sufficient to result in significant or important reductions in the stormwater constituents. It is possible that longer subsurface flow paths would result in reductions.

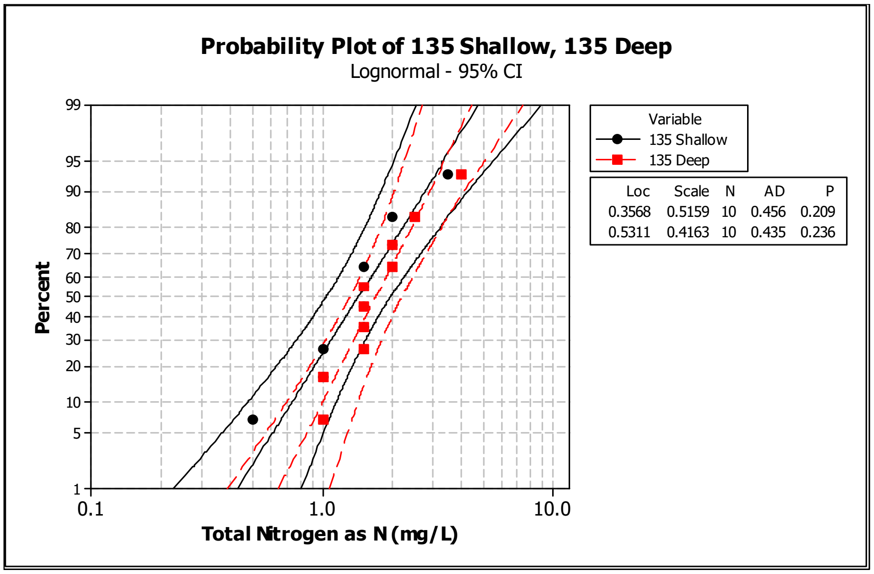

Figure 1. Example log-normal probability plot for TN (135 shallow/135 deep).

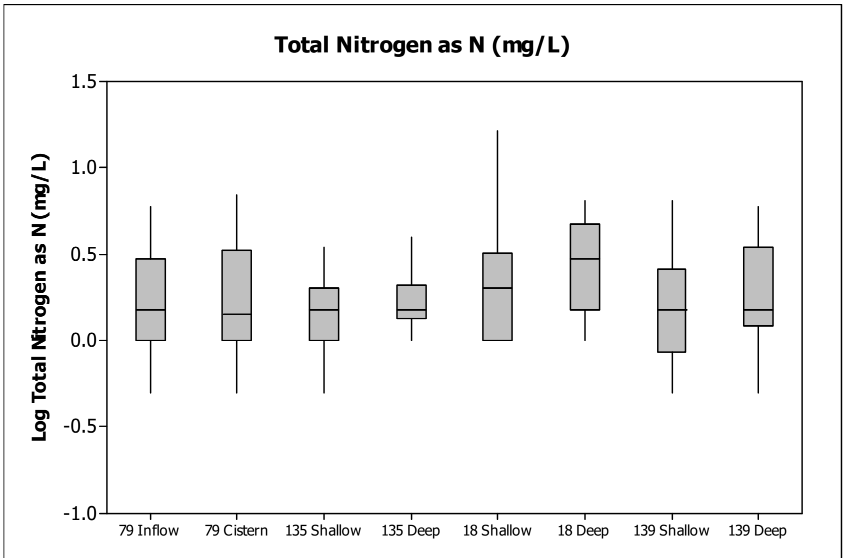

Figure 2. Example group box plot for TN.

Commented [R1]: Please narrow the range of the values (say from -1 to 1.5 (I can't edit it myself!)

L.T.: Sure, I did it but there were two values of -3 that we missed them in the new graph!!!

Table 1. Summary of Concentrations for the Influent and Effluent Samples

	79 Inflow[1]	79 Cistern[2]	135 Shallow[3]	135 Deep[3]	18 Shallow[4]	18 Deep[4]	139 Shallow[4]	139 Deep[4]
Total Coliform Bacteria (MPN/100 mL)								
Number of Samples[5]	7	8	10	10	9	9	9	9
Average[6]	1,349	5,461	10,834	14,632	12,838	20,091	13,111	11,640
St Dev[7]	3,539	8,915	12,040	11,791	12,428	11,060	11,466	10,562
***E. coli* (MPN/100 mL)**								
Number of Samples	2	7	9	9	8	9	9	9
Average	2,545	880	2,049	870	63	57	133	1,190
St Dev	3,595	2,068	2,914	1,178	70	44	91	2,775
Total Nitrogen as N (mg/L)								
Number of Samples	7	8	10	10	9	9	9	9
Average	2.1	2.4	1.6	1.9	3.5	3.2	2.1	2.3
St Dev	1.9	2.2	0.8	0.9	5	1.9	2.1	1.7

	79 Inflow[1]	79 Cistern[2]	135 Shallow[3]	135 Deep[3]	18 Shallow[4]	18 Deep[4]	139 Shallow[4]	139 Deep[4]
NO_3 plus NO_2 as N (mg/L)								
Number of Samples	6	8	10	10	9	8	8	9
Average	1.43	0.75	0.85	0.61	0.86	1.58	0.8	1.03
St Dev	1.1	0.5	0.5	0.2	0.5	1.3	0.6	1
Total Phosphorus as P (mg/L)								
Number of Samples	7	8	10	10	9	9	9	9
Average	0.14	0.19	0.12	0.10	0.21	0.44	0.18	0.18
St Dev	0.13	0.2	0.08	0.04	0.18	0.46	0.11	0.14
COD (mg/L)								
Number of Samples	7	8	10	10	9	9	9	9
Average	30.9	17.1	40.6	31.8	36.7	58.8	42.7	43.1
St Dev	14.2	11.9	13.9	12.8	18.7	49.5	6.5	9.6
Lead (mg/L) (Note: Detection Limit = 0.005 mg/L								
Number of Samples	3	3	2	4	9	9	4	1
Average	0.0063	0.034	0.014	0.021	0.071	0.092	0.01	0.38
St Dev	0.0011	0.048	0.0007	0.0081	0.11	0.11	0.0032	NA
Copper (mg/L) (Note: Note: Detection Limit = 0.02 mg/L)								
Number of Samples	7	8	10	10	3	2	10	1
Average	0.67	0.26	NA	NA	0.03	0.055	NA	0.1
St Dev	0.27	0.36	NA	NA	0.01	0.007	NA	NA
Zinc (mg/L) (Note: Detection Limit = 0.02 mg/L)								
Number of Samples	6	8	5	7	2	3	3	2
Average	0.11	0.046	0.062	0.057	0.045	0.04	0.027	0.065
St Dev	0.032	0.039	0.046	0.031	0.007	0.01	0.012	0.064

[1]The total number of samples for this location is 7.
[2]The total number of samples for this location is 8.
[3]The total number of samples for these locations is 10.
[4]The total number of samples for these locations is 9.
[5]Number of Detected Samples
[6]Average of Detected Samples
[7] Standard Deviation of Detected Samples

Table 2. Summary of Mann-Whitney Test for Paired Data

Parameter		79 Inflow vs. 79 Cistern	135 Shallow vs. 135 Deep	18 Shallow vs. 18 Deep	139 Shallow vs. 139 Deep
Total Coliform	P-value	0.03*	0.40	0.16	0.72
	Significant Difference Observed? (at level of 0.05)	Yes	No	No	No
E. coli	P-value	0.05	0.60	0.69	1
	Significant Difference Observed? (at level of 0.05)	Yes	No	No	No
Total Nitrogen (as N)	P-value	0.86	0.50	0.42	0.64
	Significant Difference Observed? (at level of 0.05)	No	No	No	No
NO_3 plus NO_2 (as N)	P-value	0.14	0.24	0.15	0.77
	Significant Difference Observed? (at level of 0.05)	No	No	No	No
Total Phosphorus (as P)	P-value	0.77	0.94	0.10	0.27
	Significant Difference Observed? (at level of 0.05)	No	No	No	No
COD	P-value	0.04	0.14	0.40	0.83
	Significant Difference Observed? (at level of 0.05)	Yes	No	No	No

* pink high-lighted cells are for constituents and locations where significant differences between the influent and effluent water quality occurred, at a traditional confidence limit of at least 0.05.

Table 3. Summary of Paired Sign Test for Metal analysis

Metal		79 Inflow vs. 79 Cistern	135 Shallow vs. 135 Deep	18 Shallow vs. 18 Deep	139 Shallow vs. 139 Deep
Lead	P-value	> 0.06	> 0.06	0.18	> 0.06
	Significant Difference in Medians?	No	No	No	No
Copper	P-value	0.125	*	>0.06	*
	Significant Difference in Medians?	No	*	No	*
Zinc	P-value	0.45	0.45	>0.06	>0.06
	Significant Difference in Medians?	No	No	No	No

* All the results are below the detection limit (BDL), therefore it is not possible to do a statistical comparison test

References

N.J.A.C. 7:9C. New Jersey Department of Environmental Protection. *Water Monitoring and Standards, Ground Water Quality Standards*, http://www.state.nj.us/dep/wms/bwqsa/docs/njac79C.pdf

New Jersey Department of Environmental Protection. *New Jersey Stormwater Best Management Practices Manual, Chapter 9.3: Standard for Dry Wells, Chapter 9.5: Standard for Infiltration Basins*. February 2004.

Pitt, R., and Talebi, L. *Evaluation and Demonstration of Stormwater Dry Wells and Cisterns in Milburn Township, New Jersey*, Prepared for: Urban Watershed Management Branch, U.S. Environmental Protection Agency. 289 pages (Final Draft Report) January 19, 2012.