Groundwater Research Report
WR02R004

F-TEST FOR NATURAL ATTENUATION IN GROUNDWATER: APPLICATION ON BENZENE

Fe S. Evangelista
Aristeo M. Pelayo

Funding Year: 2002

F-TEST FOR NATURAL ATTENUATION IN GROUNDWATER: APPLICATION ON BENZENE

R/UW-REM-008

Fe S. Evangelista
Mathematical and Computer Sciences Department, UW-Whitewater
evangelf@uww.edu

Aristeo M. Pelayo
Wisconsin Department of Natural Resources
aristeo.pelayo@dnr.state.wi.us

TABLE OF CONTENTS

Appendix B (on separate file): Map, Table with the URLs to View/Download Site Files, Table with Summary of Results, Data, F-Test analysis, Nonparametric statistics

LIST OF FIGURES AND TABLES

PROJECT SUMMARY

Title: **F Test for Natural Attenuation in Groundwater: Application on Benzene**

Project I.D.: R/UW-REM-008

Investigator(s): Principal Investigator(s):
Fe S. Evangelista, Associate Professor, Mathematical and Computer Sciences Department, UW-Whitewater ;
Aristeo M. Pelayo, Hydrogeologist, RR/3, WDNR

Period of Contract: July 1, 2002 – June 30, 2003

Background/Need: Chapter NR 726, Wis Adm. Code, allows closure of a petroleum site contaminated above NR 140 groundwater enforcement standards when natural attenuation (NA) has been demonstrated as an effective remedial option. The primary evidence for NA is contaminant concentration data that show a decreasing trend over time. However, the concentrations may be affected by the fluctuation of the water table such that conclusions from trend analysis and rate of degradation may be premature. A statistical F-test procedure was developed to determine the significance of water-table fluctuations in evaluating NA sites.

Objectives: To investigate the utility of an F-test in analyzing the statistical significance of including both groundwater elevation and time as predictors of benzene concentration. The conclusions of the test would be compared to the conclusions from the Mann-Kendall and Mann-Whitney nonparametric tests.

Methods: Data from the web-accessible WI GIS registry of closed sites were reviewed. Sites were chosen based on the presence of both groundwater elevation and benzene concentration data, and the absence of any active remediation system during the monitoring period. The data was analyzed using the F-test technique, the calculation of the apparent half-life $t_{1/2}$, (*negative* when benzene is increasing) from the slope of the *‘t’*-only regression line; and trends concluded from two nonparametric tests–Mann-Kendall (M-K), and Mann-Whitney U (M-W)–for wells where a negative $t_{1/2}$ was obtained.

Results and Discussion: Thirty wells were chosen from 25 NA sites. Twelve (12) wells were identified where the F-test concluded that the line ‘*t*’-only model is preferred. This implies that straightforward trend analysis of the concentration data is acceptable for these wells. Fifteen (15) sites had wells where at least one of the following was observed: (a) the value of $t_{1/2}$ is negative indicating an increasing trend, (b) the plane model or line *‘z’* model is preferred by the F-test, or (c) the F-test is inconclusive but the line *‘z’* model’s R^2 is larger than that for the line *‘t’* model. The latter two conditions imply that the variable ‘z’ cannot be ignored and points to the influence of groundwater elevation on benzene concentration. Consequently, the time-trend analysis of the data, including conclusions from nonparametric statistical tests may be spurious. The report includes a detailed analysis of four sites to demonstrate the range of results and insights that can be gained in using the F-test technique.

Conclusions/Implications/ Recommendations: The F-test is an analytic tool that could be used to screen sites before the calculation of a degradation rate from the linear regression of concentration *vs.* time or the use of a nonparametric test to show trends. When the F-test shows that groundwater elevation is significant (plane, *‘z’*-only, and inconclusive but *‘z’*-only has larger R^2), then we know that this *invariant* factor is affecting the concentrations; and hence, a nonparametric test is not appropriate. On the other hand, when the F-test shows that *‘z’* can be ignored (*i.e.*, test result of either: *‘t’*-only , or inconclusive but the *‘t’*-only has larger R^2), then nonparametric statistics may be more appropriately used.

Key Words: Benzene, Least-square regression, Natural Attenuation, Statistical F-Test

Funding: University of Wisconsin System

INTRODUCTION

Natural attenuation (NA) describes any or all natural processes – physical, chemical and biological – which contribute to the overall decay and slowed movement of contaminants in the environment. The application of NA as a remedy would require extensive sampling, monitoring, determining a decay rate, and verifying that the conditions conducive to contaminant degradation will be maintained and the cleanup will be effective in the predictable future [*National Research Council*, 1993]. However, in most situations, an abbreviated procedure is followed to assess NA's effectiveness. The normative routine is to estimate a degradation rate via regression of the contaminant concentration observations from monitoring wells with either time [*e.g.*, *McAllister and Chiang*, 1994] or distance from source [*Buscheck and Alcantar*, 1995]. While straightforward, these same studies (and others, *e.g.*, *Odermatt* [1999]) caution that, because other factors at the site are not invariant, conclusions from the regressions may not be sufficient. In particular, they point out that a fluctuating water table can affect the observations regarding concentrations in wells. In this study, we include groundwater elevations as a factor in analyzing the variation in the benzene concentrations from monitoring wells at petroleum-contaminated sites. The state of Wisconsin, through revisions in 1996 of chapter NR 726, Wisconsin Administrative Code, and the promulgation in 2001 of ch. NR 746, allows NA closure of cases involving petroleum contaminants that still exceed ch. NR 140 groundwater enforcement standards. Figure 1 shows the locations of the sites we reviewed from the *WDNR GIS Registry of Closed Remediation Sites* [http://gomapout.dnr.state.wi.us/org/at/et/geo/gwur]. We show how the time-series of concentration data and water-table elevations can be jointly used in the analyses, and show through an F-test whether the inclusion of the groundwater elevation observations is statistically important in explaining the observed concentrations at a monitoring well. We focus on benzene because it is typically the driver for remediation and other decisions at most sites.

PROCEDURES AND METHODS

There are three possible linear regression models that use time and groundwater elevation as predictors of concentration: two (2) line models which we denote as H_t and H_z, where the predictors are time (t) and groundwater elevation (z), respectively; and a plane model $H_{t,z}$, where both variables are used (Figure 2). The respective regression equations for the three models are:

$$H_t : \mathrm{c} \sim \mathrm{c(t)} \text{ or } \textit{'t' - only line} \text{ model :} \quad y_i = \log(c_i) = y_0 + k' t_i + e_i$$

$$H_z : \mathrm{c} \sim \mathrm{c\,(z)} \text{ or } \textit{'z' - only line} \text{ model :} \quad y_j = \log(c_j) = y_0 + y_2 z_j + e_j$$

$$H_{t,z} : \mathrm{c} \sim \mathrm{c(t,z)} \text{ or } \textit{'plane'} \text{ model :} \quad y_k = \log(c_k) = b_0 + b_1 t_k + b_2 z_k + e_k$$

$$i, j, k = 1, 2, \ldots, N$$

where: N = Number of observations consisting of concentration (c) and groundwater elevation (z), both sampled at time (t)

y_i, y_j, y_k = Log-transformed concentration data

e_i, e_j, e_k = Residuals of the line models and plane model

y_o, k'; y_0, y_2; b_0, b_1, b_2 = Model parameters to be determined by "least squares"

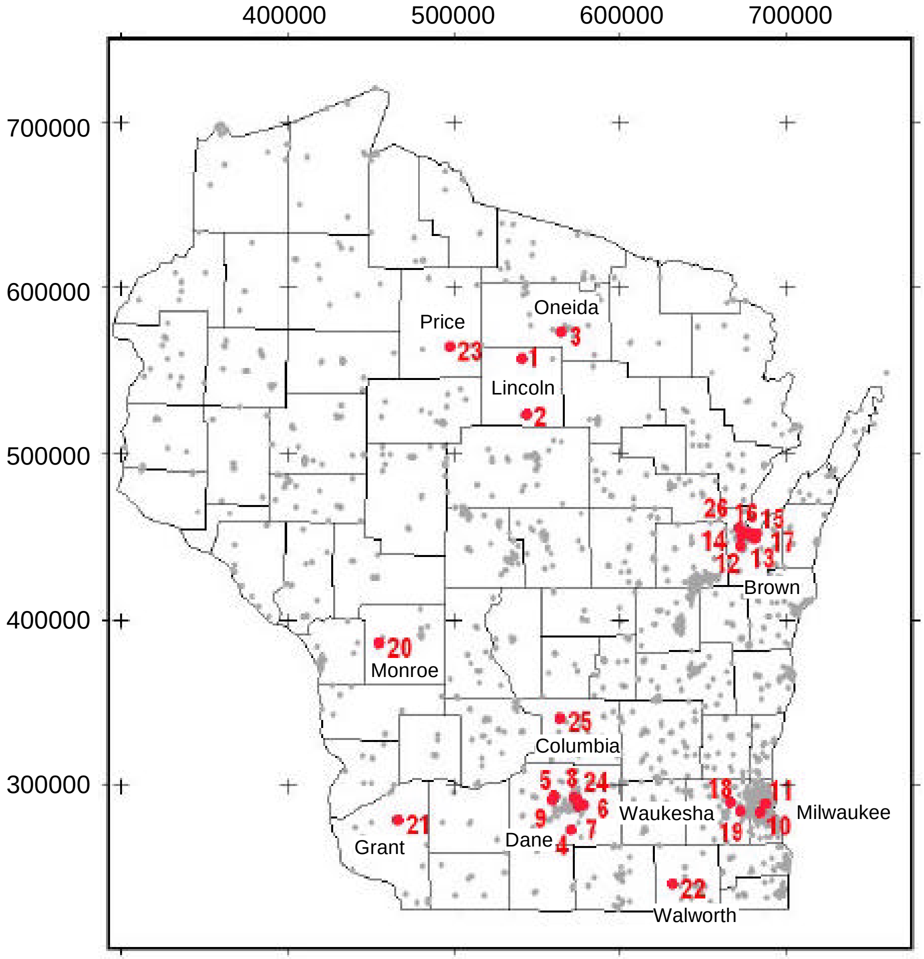


Figure 1. Locations of the 26 sites included in this study. Map projection is in Wisconsin Transverse Mercator (WTM83/91). The counties where the sites are located are labeled. Dots are location of sites in the WDNR GIS Registry of Closed Remediation Sites (*http://gomapout.dnr.state.wi.us/org/at/et/geo/gwur*) that includes over 2,100 UST cases as of September 2003. Milwaukee, Dane and Brown are the counties with the most number of these closed sites in the registry.

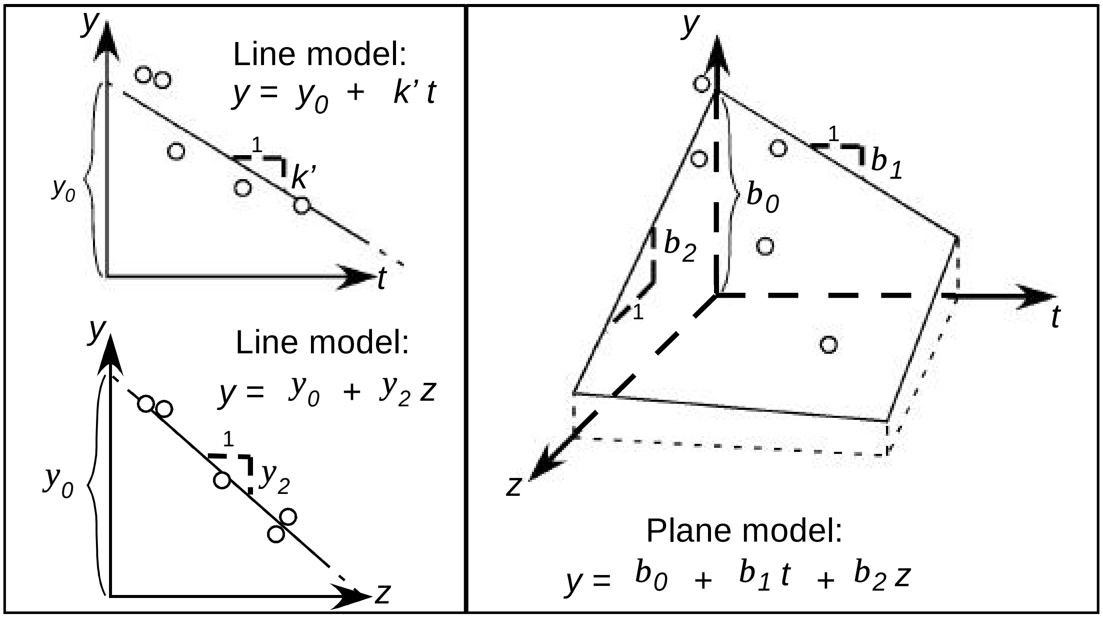


Figure 2. The regression “line” and “plane” parameters. Circles represent the data points. The figure is adapted from *Weisberg* [1985]. The variables t and z are the predictors of y in the regression models.

To determine which of these three models best explains the variation in the concentration data, we devised an F-test to evaluate two hypotheses, as described in *Pelayo and Evangelista* [2003]. Briefly, the conclusion of each hypothesis test is to reject or not to reject the null hypothesis at a given significance level, usually 90%. We summarize the possible results of the F tests in Table 1.

Table 1. Possible Results of the F-Test

H_t vs. $H_{t,z}$ ***H_z vs. $H_{t,z}$***	***Reject H_t***	***Not reject H_t***
Reject H_z	A model with *both* variables *t* and *z* is better than either of the two line models.	Time is a better predictor than groundwater elevation. Adding *t* to a model that has *z* significantly improves the fit, while it is not necessary to add *z* to a model that already has *t*.
Not reject H_z	Groundwater elevation *z* is a better predictor than *t*. Adding *z* to a model that has *t* significantly improves the fit, while it is not necessary to add *t* to a model that already has *z*.	Either of the line models is better than a plane model. (Compare the R^2 value, and statistically test whether the larger R^2 is not zero.)

The implication of the result of the F-tests is that:

1.) When H_t (time, or '*t*'-only model) is the best alternative, then straightforward trend analysis of the concentration data should be acceptable.
2.) When H_z (elevation, or '*z*'-only model), or $H_{t,z}$ (plane model) is the best alternative, the time-trend analysis of the concentration data may be suspect.
3.) When both H_t and H_z are not rejected, the result is *inconclusive* on the plane model's importance, and the line model with the larger R^2 may suffice.

RESULTS AND DISCUSSION

A site was chosen for analysis based on the following criteria: (1) Presence of groundwater elevation data; (2) Benzene concentration data in sufficiently high levels so that an analysis could be made; (3) Absence of any active remediation system that would affect the observations during the monitoring period. Our intention is to show the utility of the F-test technique, and insights we gained by using the technique. Data availability was our foremost consideration. We did not purposely select sites via random sampling, so the sites we included cannot be considered a random sample of closed UST sites in Wisconsin. Table 2 is a summary of the results of this study. It lists: (a) the 26 sites, each identified by its BRRTS Activity Number; (b) the particular site wells we used; (c) the latest benzene data; (d) values of $t_{1/2}$, the apparent half-life (*negative* when benzene is increasing) calculated from the slope of the '*t*'-only regression; (e) the regression's coefficient of determination R^2; (f) trends concluded from two nonparametric – Mann-Kendall (M-K) and Mann-Whitney U (M-W) – statistical tests for wells where we obtained a negative $t_{1/2}$; and (g) results from the F test.

Table 2. Summary of Results

							M-K Test	M-W Test	F Test
Site #	BRRTS#	Well#	Most Recent Benzene (ug/l)	Month / Year of Recent Sample	$t_{1/2}$ (yr)	R^2	If $t_{1/2}$ is *negative*, what is the trend from the nonparametric test?		Result
1	0335000169	MW7A	64	08 / 1997	0.7	0.43			't'-only y(t)
2	0335153171	MW3	70	03 / 2000	2.0	0.58			't'-only y(t)
2	0335153171	MW4	270	03 / 2000	-2.0	0.19	Increasing	Increasing	Elev. 'z'-only y(z)
3	0344001068	MW1	1,300	02 / 2002	3.6	0.24			Inconclusive; y(t) has larger R2.
3	0344001068	MW3	210	02 / 2002	1.0	0.25			Inconclusive; y(z) has larger R2.
4	0313002834	MW1	1,800	04 / 2000	1.0	***0.92***			't'-only y(t)
5	0313104797	MW4	260	09 / 2001	2.5	0.21			Inconclusive; y(t) has larger R2.
6	0313223088	MW2	400	04 / 2002	2.8	0.44			't'-only y(t)
7	0313002020	MW1	160	09 / 1999	-1.0	0.54	Increasing	No trend	't'-only y(t)
8	0313002496	MW4	820	11 / 1999	-9.3	0.00	No trend	No trend	Inconclusive; y(z) has larger R2.
9	0313000520	MW4	1,300	08 / 1998	0.6	***0.77***			Inconclusive; y(t) has larger R2.
10	0341000099	MW4	1,400	09 / 1999	11.0	0.53			't'-only y(t)
11	0341003429	MW4	230	02 / 2002	-6.3	0.47	Increasing	Increasing	't'-only y(t)
12	0305000848	MW8	1,200	04 / 1998	-1.7	0.24	insufficient data		Inconclusive; y(t) has larger R2.
13	0305178601	MW6	4,500	11 / 1999	2.0	0.28			't'-only y(t)
14	0305000233	MW1	27	09 / 1998	1.9	0.03			Inconclusive; y(z) has larger R2.
15	0305216071	MW13	3,000	05 / 2002	-7.7	0.22	insufficient data		Inconclusive; y(t) has larger R2.
16	0305000176	MW3	4,400	05 / 1999	-24.0	0.00	Decreasing	Decreasing	Inconclusive; y(z) has larger R2.
17	0305001435	MW3	0	10 / 2002	0.4	***0.70***			't'-only y(t)
18	0368199644	MW2	170	02 / 2001	3.7	0.17			Inconclusive; y(t) has larger R2.
19	0368004033	MW3	400	06 / 2000	-1.7	0.11	No trend	No trend	Plane y(t, z)
20	0342001502	MW2	45	09 / 2001	-3.8	0.04	No trend	No trend	Inconclusive; y(t) has larger R2.
21	0322001852	MW6	32	01 / 2003	0.4	***0.69***			Inconclusive; y(t) has larger R2.
22	0365176916	MW1	449	03 / 2002	1.2	***0.76***			't'-only y(t)
22	0365176916	MW4	1,620	03 / 2002	-2.4	***0.78***	Increasing	No trend	't'-only y(t)
23	0351000828	MW8	3,000	05 / 2000	2.1	0.46			't'-only y(t)
23	0351000828	PZ1	3,000	05 / 2000	22.9	0.02			Inconclusive; y(z) has larger R2.
23	0351000828	MW7	5,000	05 / 2000	1.4	***0.63***			't'-only y(t)
24	0313001446	MW3	5,500	04 / 2002	-16.1	0.06	Increasing	No trend	Elev. 'z'-only y(z)
25	0311002155	MW1	170	07 / 1999	0.8	***0.98***			Inconclusive; y(t) has larger R2.
26	0305257883	MW3	150	01 / 2003	2.6	0.01			Plane y(t, z)

BRRTS# = DNR Activity Number in *http://www.dnr.state.wi.us/org/aw/rr/brrts/index.htm*

Additional site information is available at *http://gomapout.dnr.state.wi.us/org/at/et/geo/gwur* .

$t_{1/2}$ = Benzene's half-life for t-only [y ~ y(t)] model; negative when benzene is increasing.

R^2 = Coefficient of determination for the t-only model; better fit as R^2 ~1.

M-K Test = Nonparametric α= 0.2 Mann-Kendall Test for well's 8 most recent data

M-W Test = Nonparametric α= 0.1 Mann-Whitney U Test for well's 8 most recent data

Site #1 had an air-sparging system during the monitoring period. Our analysis showed that while the groundwater elevations fluctuated, it (elevation *'z'*) did not seem to have an effect on the decay of the benzene levels. The quick decay (half-life $t_{1/2}$ of less than 1 yr) from a starting benzene concentration of 3,000 ug/l can be attributed, not to NA, but to the active system. We used this site's $t_{1/2}$ to mentally note how the other 25 "NA" sites would fare. In each NA site, we used data from one to three monitoring wells for a total of 30 monitoring wells.

We found eleven (11) of the 25 sites with a well where the value of $t_{1/2}$ is negative, indicating that benzene concentrations are increasing with time. We applied nonparametric tests using Mann-Kendall ($\alpha = 0.2$ level) and Mann-Whitney U ($\alpha = 0.1$) statistics on the eight (8) most recent benzene concentrations from these wells to determine concurrence in concluding that concentrations were increasing from these wells. Only 9 of the 11 sites with negative $t_{1/2}$ wells had 8 or more benzene data. NR 746 prescribes eight (8) as the number of samples when using the Mann-Whitney U test, and requires equal-time intervals between sampling events. The equal-interval requirement would have disqualified all but 2 of the 9 wells, so in our use of the nonparametric tests, we didn't require that the data be equally spaced in time. Of the 9 "negative- $t_{1/2}$" wells with sufficient data, the nonparametric tests (both M-K and M-W) show a *decreasing* trend in one (1) well (Site #16). In looking more closely at the 4-year data from site #16, there is practically a zero slope on the regression of log(benzene) *vs*. time. It is interesting to note that site #16 belongs to a set of sites (including 4 other "negative- $t_{1/2}$" wells from sites # 2, 8, 19 and 24) where our analysis using the F-test technique shows that *'z'* can not be ignored as a factor in the observed benzene concentrations.

The breakdown of the results from the F-test at the 90% confidence level for data from the 30 individual wells (from the 25 NA sites) is as follows: 12 wells – line *'t'*-only model preferred; 2 wells – line *'z'*-only model preferred; 2 wells – plane model is statistically significant; and 14 wells – results were inconclusive. An inconclusive result from the F-test implies that the plane model is not statistically significant, so either the line '*t*-only model or the line *'z'*-only model may suffice, except that the value of R^2 in either of the models is usually quite small. The value of R^2 is between 0 and 1; R^2 is the proportion of the variability in benzene concentration that can be explained by the given variable (either*'t'* or *'z'*). Small R^2 values imply that the variable under consideration is not a good predictor of contaminant concentration. In 11 out of the 14 "inconclusive" wells, the better R^2 value from the line models was less than 0.3, so the decay rate of benzene calculated using the slope of the regression line may be spurious. The 3 other "inconclusive" wells (from sites # 9, 21 and 25) with high R^2 values (0.77, 0.69 and 0.98, respectively) had only 4 data points each, reducing the significance of their R^2.

From the results of our analysis, we identified 15 sites (out of the 25 NA sites) with wells where one of the following holds: the value of $t_{1/2}$ is negative, the plane model or line *'z'* model is preferred by the F-test, or the F-test is inconclusive (like site #16) but the line *'z'* model's R^2 is larger than that for the line *'t'* model. These are sites where the benzene is increasing, or the groundwater elevation *'z'* cannot be ignored in analyzing the benzene concentrations. What follows is our analysis of wells from four (4) of these sites to demonstrate the range of results and further insights that can be gained in using the F-test technique.

Site #2, BRRTS No. 0335153171, MW-4, Merrill (Figure 3). The time-series plots of groundwater elevation and benzene concentration in Figure 3 suggest an inverse relationship between these two quantities – as benzene concentration increases, groundwater elevation decreases and vice-versa. The c^2 error for the *'t'*-only model is 0.523 and is improved to 0.187 by the plane model. In contrast, the c^2 error for the *'z'*-only model is 0.214. The F-test results show that the addition of the variable *z* to a model that already has the variable *t* is statistically significant, with F= 12.5 being larger than the critical-F of 3.59 at a 90% significance level. However, the reverse is not true. The plane's improved c^2 is not statistically significant over the *'z'*-only model (F = 0.99 < 3.59). The conclusion from the F-test is that the observed benzene concentrations from this particular monitoring well are best explained by the fluctuation in the groundwater elevation. In fact, the linear regression line for benzene concentration *vs.* time (solid line in the graph) shows an increasing trend, suggesting the possibility that the contaminant plume is moving downgradient from the site.

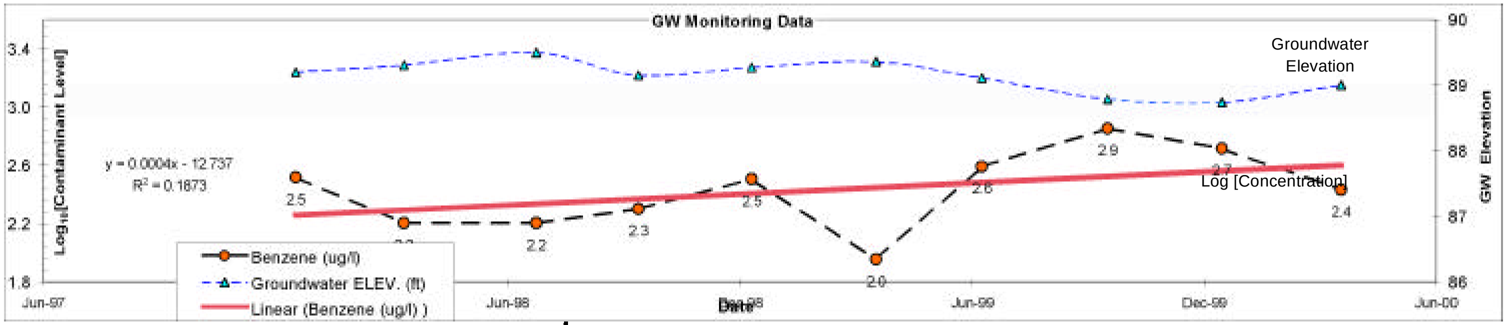
Site #3, BRRTS No. 0344001068, MW-1, Rhinelander (Figure 4). The F-test result for this well typifies what we found for many of the wells we studied, occurring in 14 out of 31 wells. We labeled the result as "Inconclusive" in Table 2. The inconclusive result shows that the plane model is statistically *not* an improvement over either of the line models. However, neither the *'t'*-only model ($R^2 = 0.243$) nor the *'z'*-only model ($R^2 = 0.184$) can explain the observed variation in contaminant concentration convincingly. For this particular well, we conclude that the *'t'*-only model is the better of the two line models. The apparent half-life for the decay of benzene may be just an artifact of the regression since the $t_{1/2}$ is more than 3.6 years, or longer than the 3-year period for the monitoring at this site.

Site #22, BRRTS No. 0365176916, Delavan (Figure 5). For this site, we were able to compare data from a near-source well (MW-1) and a downgradient well (MW-4). The result from MW-1 stresses the dependence of the conclusion on the level of confidence chosen. Using all the data and at a significance level $a = 0.10$, a *'t'*-only model is preferred, while at $a = 0.15$ test, a plane model would be preferred. By excluding the first data point (collected more than 2 years before the rest of the data), we would conclude from the F-test that the plane model is better. We interpreted this to mean that groundwater elevation may be a significant factor, and prompted us to look at other site wells. The groundwater map in the GIS Registry for the site (Figure 5 inset) indicated that the flow direction is to the northeast (long arrow in the map). Closer inspection of the map showed that the previous investigator may have inadvertently contoured ground *surface* elevations from the wells. Our reanalysis of the data showed that the groundwater flow direction is more to the northwest (shorter arrows on the inset map), so MW-4 was a downgradient well. The regression line for MW-4 shows increasing benzene concentration over time – which is what we would expect when the F-test in the near-source well points to a *'z'* effect. Another well downgradient from MW-4 also has increasing benzene concentration over the monitoring period, further confirming that the contaminant plume is moving rather than degrading. The F-test technique may not be able to screen a site like this, especially given only the few data from the source well. For a more robust screening, an $a > 0.1$ test may be needed when fewer data are available, and the additional analysis from a downgradient well can resolve initial ambiguities.

BRRTS No. 0335153171. Additional site information is available at: *http://gomapout.dnr.state.wi.us/org/at/et/geo/gwur/pdf/Nor/Lincoln/03%20LUST/0335153171/0335153171.pdf*

Site 2

MW: **4**										
Groundwater ELEV. (ft)	89.20	89.31	89.50	89.15	89.27	89.36	89.11	88.79	88.74	89.00
Sampling Dates	12/22/1997	03/17/1998	06/29/1998	09/17/1998	12/15/1998	03/23/1999	06/14/1999	09/21/1999	12/20/1999	03/23/2000
Benzene (ug/l)	330.0	160.0	160.0	200.0	320.0	90.0	390.0	710.0	520.0	270.0
Log_{10} [Contaminant (ug/l)]	2.5	2.2	2.2	2.3	2.5	2.0	2.6	2.9	2.7	2.4



DATA			MODELS		
			Line Fit	Plane Fit	Line Fit
y = log(c)	t (Day 1 = 1/1/1900)	z	y(t)	y(t, z)	y(z)
2.5	35786	89.2	2.3	2.5	2.4
2.2	35871	89.31	2.3	2.3	2.3
2.2	35975	89.5	2.3	2.1	2.1
2.3	36055	89.15	2.4	2.5	2.4
2.5	36144	89.27	2.4	2.3	2.3
2.0	36242	89.36	2.4	2.2	2.2
2.6	36325	89.11	2.5	2.4	2.5
2.9	36424	88.79	2.5	2.8	2.7
2.7	36514	88.74	2.6	2.8	2.8
2.4	36608	89	2.6	2.5	2.6

n: 10		c^2 ("t" line)	c^2 (plane)	c^2 ("z" line)
Mean: 2.428				
SYY: 0.644	$S\,(Y_{model} - Y_{data})^2$:	0.52310	0.18744	0.21399

REGRESSION

	Line 't' Model $y = y_0 + k't$	Plane Model $y = b_0 + b_1 t + b_2 z$	Line 'z' Model $y = y_0 + y_2 z$
n (degrees of freedom):	8	7	8
R^2 (Coeff. Determination):	0.187	0.709	0.668
Adjusted R^2:	0.086	0.626	0.626
T-test of R^2: (Acceptable line fit if > 90%)	78.8		99.6
$y_0 \mid b_0 \mid y_0$ (intercept):	-12.73715273	112.5724301	82.65952065
$k' \mid b_1$ (slope for t):	0.000418983	-0.000277528	
$b_2 \mid y_2$ (slope for z):		-1.12291247	-0.900035157
	y(t)	y(t, z)	y(z)
c^2 (Goodness of Fit): smaller is better	0.52310	0.18744	0.21399

Apparent Half-Life		
k	$t_{1/2}$ (d)	$t_{1/2}$ (yr)
0.000964744	-718	-1.968

F-**Test:** *Is the improvement in c^2 by the plane model statistically significant?*

Line Models: -----> H_t ("t" Hypothesis): $y = y_0 + k't$ H_z ("z" Hypothesis): $y = y_0 + y_2 z$

Plane Model: -----> H_A (Alternative): $y = b_0 + b_1 t + b_2 z$

Test Statistics: ---> $F_{ct} = Dc_t^2 / c_n^2$ $F_{cz} = Dc_z^2 / c_n^2$

where: $Dc_t^2 = c^2$ ("t" line) - c^2 (plane) $Dc_z^2 = c^2$ ("z" line) - c^2 (plane)

$c_n^2 = c^2 / n$ = reduced c^2 of the plane model

We can be confident in the relative merit of the plane model if F_c is large.

(If $F_c >$ 3.589, we would favor the plane model over the line model at a 0.1 significance level test.)

(If $F_c <$ 0.506, we have a 50% chance occurrence.)

$F_{ct} = Dc_t^2 / c_n^2 =$ 1.25E+01 $F_{cz} = Dc_z^2 / c_n^2 =$ 9.91E-01

	$a = P_F(F, 1, n)$	Critical F-values				
	0.200	2.002				
--->	0.100	**3.589**	<---			
Plane	0.009	1.25E+01		z	0.353	9.91E-01

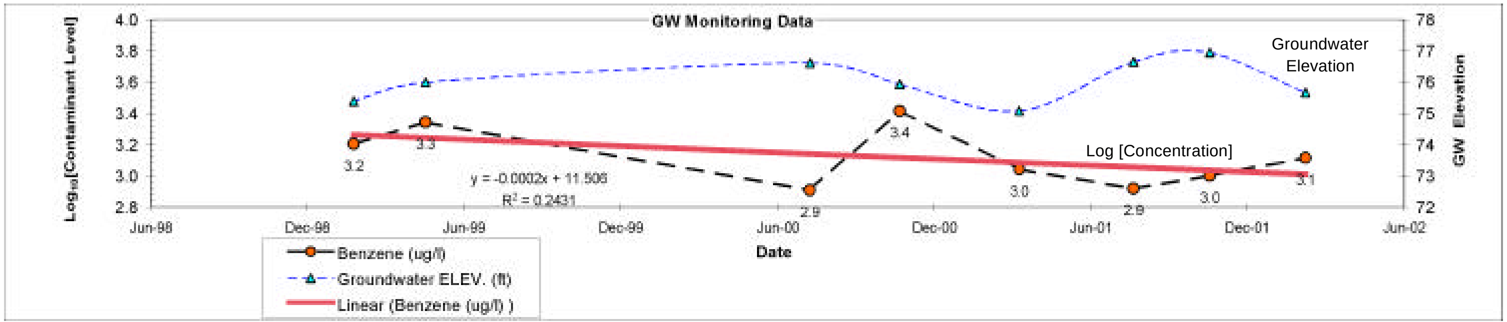
Conclusion: Is y(t, z) better? NO, y(z) - 'z' only line model - is preferred at a 90% level of confidence F test.

Figure 3. Data and results from monitoring well MW-4 at a site in Merrill. The groundwater elevations are plotted as triangles connected by short dashes. The benzene concentrations are shown by filled circles connected by long dashes. The solid line is the regression result of the time-only model. The plots suggest an inverse relationship between groundwater elevation and benzene concentration. The conclusion from the F-tests is that the observed benzene concentrations from this particular monitoring well is best explained by the fluctuation in the groundwater elevation. Incidentally, the benzene concentration data do not show a decay, but rather an increasing trend over time, underscoring the possibility of the contaminant plume moving downgradient from the site.

BRRTS No. 0344001068. Additional site information is available at: *http://gomapout.dnr.state.wi.us/org/at/et/geo/gwur/pdf/Nor/Oneida/03%20LUST/0344001068/0344001068.pdf*

Site 3

MW: *1*								
Groundwater ELEV. (ft)	*75.39*	*75.99*	*76.61*	*75.92*	*75.07*	*76.64*	*76.95*	*75.66*
Sampling Dates	*01/29/1999*	*04/23/1999*	*07/13/2000*	*10/26/2000*	*03/14/2001*	*07/25/2001*	*10/22/2001*	*02/11/2002*
Benzene (ug/l)	*1,600.0*	*2,200.0*	*810.0*	*2,600.0*	*1,100.0*	*830.0*	*1,000.0*	*1,300.0*
Log_{10} [Contaminant (ug/l)]	*3.2*	*3.3*	*2.9*	*3.4*	*3.0*	*2.9*	*3.0*	*3.1*



DATA			*MODELS*		
			Line Fit	*Plane Fit*	*Line Fit*
y = log(c)	*t* *(Day 1 = 1/1/1900)*	*z*	*y(t)*	*y(t, z)*	*y(z)*
3.2	*36189*	*75.39*	*3.3*	*3.3*	*3.2*
3.3	*36273*	*75.99*	*3.2*	*3.2*	*3.1*
2.9	*36720*	*76.61*	*3.1*	*3.1*	*3.0*
3.4	*36825*	*75.92*	*3.1*	*3.1*	*3.1*
3.0	*36964*	*75.07*	*3.1*	*3.2*	*3.2*
2.9	*37097*	*76.64*	*3.1*	*3.0*	*3.0*
3.0	*37186*	*76.95*	*3.0*	*3.0*	*3.0*
3.1	*37298*	*75.66*	*3.0*	*3.1*	*3.2*

			c^2 *("t" line)*	c^2 *(plane)*	c^2 *("z" line)*
n: 8					
Mean: 3.118					
SYY: 0.249		*S* $(Y_{model} - Y_{data})^2$ *:*	*0.18867*	*0.16681*	*0.20340*

REGRESSION

	Line 't' Model $y = y_0 + k't$	*Plane Model* $y = b_0 + b_1 t + b_2 z$	*Line 'z' Model* $y = y_0 + y_2 z$
n (degrees of freedom):	*6*	*5*	*6*
R^2 *(Coeff. Determination):*	*0.243*	*0.331*	*0.184*
Adjusted R^2*:*	*0.117*	*0.063*	*0.048*
T-test of R^2*: (Acceptable line fit if > 90%)*	*78.6*		*71.1*
$y_0 \mid b_0 \mid y_0$ *(intercept):*	*11.50589062*	*16.69674786*	*12.4601932*
$k' \mid b_1$ *(slope for t):*	*-0.000227813*	*-0.000185369*	
$b_2 \mid y_2$ *(slope for z):*		*-0.088829681*	*-0.122876429*
	y(t)	*y(t, z)*	*y(z)*
c^2 *(Goodness of Fit): smaller is better*	*0.18867*	*0.16681*	*0.20340*

Apparent Half-Life		
k	$t_{1/2}$ (d)	$t_{1/2}$ (yr)
-0.000524558	1,321	3.620

F -Test: *Is the improvement in* c^2 *by the plane model statistically significant?*

Line Models: -----> H_t *("t" Hypothesis):* $y = y_0 + k't$ H_z *("z" Hypothesis):* $y = y_0 + y_2 z$

Plane Model: -----> H_A *(Alternative):* $y = b_0 + b_1 t + b_2 z$

Test Statistics: ---> $F_{ct} = Dc_t^2 / c_n^2$ $F_{cz} = Dc_z^2 / c_n^2$

where: $Dc_t^2 = c^2$ *("t" line)* - c^2 *(plane)* $Dc_z^2 = c^2$ *("z" line)* - c^2 *(plane)*

$c_n^2 = c^2 / n$ = *reduced* c^2 *of the plane model*

We can be confident in the relative merit of the plane model if F_c *is large.*

(If $F_c > 4.060$*, we would favor the plane model over the line model at a 0.1 significance level test.)*

(If $F_c < 0.528$*, we have a 50% chance occurrence.)*

$F_{ct} = Dc_t^2 / c_n^2 =$ *6.55E-01* $F_{cz} = Dc_z^2 / c_n^2 =$ *1.10E+00*

	$a = P_F(F, 1, n)$	*Critical F-values*				
	0.200	*2.178*				
--->	*0.100*	***4.060***	<---			
t	*0.455*	*6.55E-01*		*z*	*0.343*	*1.10E+00*

Conclusion: Is y(t, z) better? *Statistically, plane model is not necessary; y(t) with a larger R2 may suffice.*

Confidence level for larger R2 not being zero is 78.6 %.

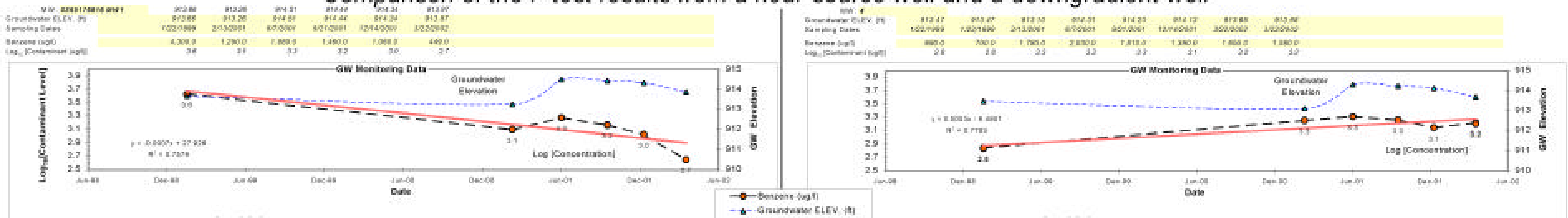
Figure 4. Data and results from monitoring well MW-1 at a site in Rhinelander. The symbols are the same as in Figure 3. This is an example of an “Inconclusive” result from the F test.

BRRTS No. 0365176916. Additional site information is available at: *http://gomapout.dnr.state.wi.us/org/at/et/geo/gwur/pdf/Ser/Walworth/03%20LUST/0365176916/0365176916.pdf*

Site 22

"Is the plane model: y ~ y(t, z) better?"

Comparison of the F test results from a near-source well and a downgradient well



REGRESSION

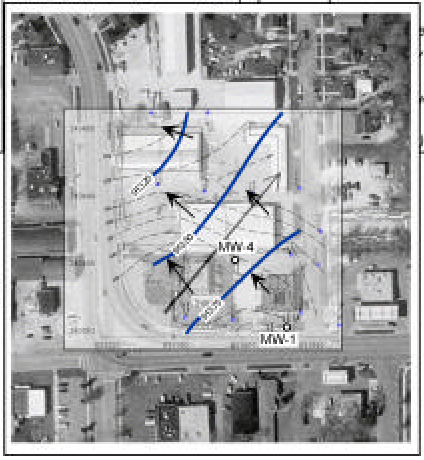
	Line 't' Model $y = y_0 + k't$	*Plane Model* $y = \beta_0 + \beta_1 t + \beta_2 z$	*Line 'z' Model* $y = \psi_0 + \psi_2 z$
ν (degrees of freedom):	4	3	4
R^2 (Coeff. Determination):	0.758	0.903	0.002
Adjusted R^2:	0.697	0.839	-0.248
T-test of R^2: (Acceptable line fit if > 90%)	97.6		6.1
y_0 | β_0 | ψ_0 (intercept):	27.92618541	-216.150058	26.88651031
k' | β_1 (slope for t):	-0.000670179	-0.000814112	
β_2 | ψ_2 (slope for z):		0.272861778	-0.025979311
	y(t)	*y(t, z)*	*y(z)*
χ^2 (Goodness of Fit): smaller is better	0.12484	0.04990	0.51417

Apparent Half-Life		
k	$t_{1/2}$ (d)	$t_{1/2}$ (yr)
-0.001543144	449	1.231

REGRESSION

	Line 't' Model $y = y_0 + k't$	*Plane Model* $y = \beta_0 + \beta_1 t + \beta_2 z$	*Line 'z' Model* $y = \psi_0 + \psi_2 z$
ν (degrees of freedom):	6	5	6
R^2 (Coeff. Determination):	0.779	0.780	0.183
Adjusted R^2:	0.742	0.692	0.047
e line fit if > 90%)	99.6		71.0
$_0$ (intercept):	-9.480073734	-22.79137957	-166.7123748
(slope for t):	0.000341559	0.000335574	
$_2$ (slope for z):		0.014809525	0.165875161
	y(t)	*y(t, z)*	*y(z)*
): smaller is better	0.05322	0.05300	0.19699

Apparent Half-Life		
k	$t_{1/2}$ (d)	$t_{1/2}$ (yr)
0.000786469	-881	-2.415



MW-1 - near-source well

NO, y(t) - time only line model - is preferred at a 90% level of confidence F test.

YES. Plane y(t, z) fit better than either 't' only or 'z' only model BUT only at a 87.6% level of confidence F test.

MW-4 - downgradient well

NO, y(t) - time only line model - is preferred at a 90% level of confidence F test.

Figure 5. Data and results from a near-source well MW-1 and a downgradient well MW-4 from a site in Delavan. The symbols are the same as in Figure 3. The results for MW-1 stress the dependence of the conclusion on the level of confidence imposed on the test. The inset shows a map with groundwater elevation contours from our reanalysis of the most recent groundwater elevation data from the site, showing flow direction (short arrows) to the northwest. The thin dashed contours were from the groundwater map in the GIS Registry file for this site, indicating flow to the northeast (long arrow); however, close inspection showed that the thin contours may be based on ground *surface* elevations. Our reanalysis of the groundwater elevations shows that MW-4 is a downgradient well. The regression line for MW-4 shows an increasing trend, suggesting that the contaminant plume is moving rather than degrading.

Site #23, BRRTS No. 0351000828, MW-8 and PZ-1, Prentice (Figure 6). This next example shows that the F-test results on different wells at a site may not necessarily have the same conclusion. The differing conclusions would need to be reconciled when interpreting the F-test results. In Figure 6, we compare the results from two wells (both near-source) – water-table well MW-8 and adjacent piezometer PZ-1. For either of the wells, the plane-model is not the statistically better alternative. For MW-8, the linear regression of concentration *vs.* time yielded an R^2 of 0.461, while that for the '*z*'-only model is only .019. The addition of the variable '*z*' to obtain a plane model is not statistically significant at the 90% or even 80% level. Hence it seems clear that the line '*t*'-only model is preferred. Another well (MW-7, Table 2) has similarly decreasing benzene, and thus the data from the water-table wells indicate that natural attenuation is taking place. However for PZ-1, the F-test result is "Inconclusive." For PZ-1, both line models have low R^2 value with the line '*z*'-only model being preferred due to its larger R^2 (0.208 vs. 0.017). The trend analysis of the PZ-1 data suggests a much longer half life $t_{1/2}$ of 23 years for the decay of benzene concentration compared to an apparent shorter half life of 2 years in MW-8 (or 1.4 yr in MW-7). To reconcile the results from the piezometer PZ-1 with the results from the water-table wells, it seems likely that there is a downward vertical component to the plume movement. The potential for downward plume migration would need to be evaluated independent of the F-test, and this evaluation comes from the comparison of the groundwater elevations between the wells. While the vertical hydraulic gradient between MW-8 and PZ-1 appears mostly upward, for 3 out of the 8 monitoring rounds, the gradient is downward (indicated by arrows over PZ-1 elevations in Figure 6), which are critical observations in support of the dissimilar F-test results for these wells.

CONCLUSIONS AND RECOMMENDATIONS

Our study has benefited from the web-accessible WDNR Registry of Closed Remediation Sites where site information pertaining to remaining contamination can be downloaded. Whereas in the past, an across-the-board comparison to a set of groundwater quality standards was the key, regulators today decide on the level of residual contamination at *each* site that would not adversely affect a sensitive receptor. In this study, we showed how the application of the statistical F-test technique we developed [*Pelayo and Evangelista,* 2003] as a quick analytic tool may be able to help assess plume behavior. The technique extends the normative procedure of time-trend analysis by including groundwater elevation measurements '*z*' as a factor in explaining the variation in benzene concentrations. Together with the values of $t_{1/2}$ and R^2, the technique is designed to identify sites where '*z*' cannot be ignored as a factor. We started by analyzing only near-source wells from the 26 sites in this study, but then quickly realized that the conclusion from the F-test technique can be made more robust when, not only the near-source well, but also downgradient well(s) and piezometers are included, especially where the remaining benzene is still quite high (>500 ug/l). The inclusion of '*z*' in our analysis provided us an indirect look into how the contaminant plume is changing and how the groundwater movement may be affecting the observed concentration.

One of our original study objectives was to compare the first-order decay rate we determined at '*t*'-only sites to a 1-d batch-flush model. We had presumed (perhaps injudiciously) that the '*t*'-only sites would have benzene decaying, large R^2 (>0.6) on the regression results, and clearly defined groundwater flow for the site. However, a surprising number (11 sites) had wells with

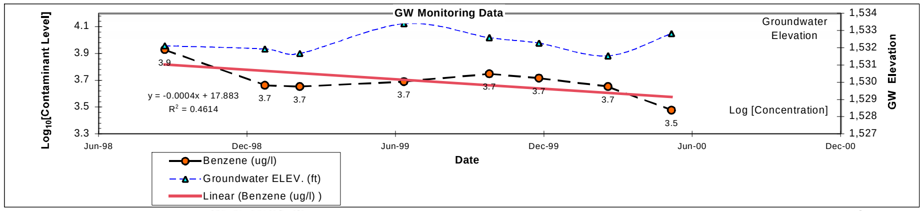
BRRTS No. 0351000828. Additional site information is available at: *http://gomapout.dnr.state.wi.us/org/at/et/geo/gwur/pdf/Nor/Price/03%20LUST/0351000828/0351000828.pdf*

Site 23

"Is the plane model: y ~ y(t, z) better?"
Comparison of the F-test results from 2 adjacent wells

MW: ***8***

Groundwater ELEV. (ft)	*1532.12*	*1531.93*	*1531.67*	*1533.40*	*1532.59*	*1532.27*	*1531.53*	*1532.82*
Sampling Dates	*8/27/1998*	*12/28/1998*	*2/9/1999*	*6/17/1999*	*9/30/1999*	*11/30/1999*	*2/23/2000*	*5/11/2000*
Benzene (ug/l)	*8,500.0*	*4,600.0*	*4,500.0*	*4,900.0*	*5,600.0*	*5,200.0*	*4,500.0*	*3,000.0*
Log_{10} [Contaminant (ug/l)]	*3.9*	*3.7*	*3.7*	*3.7*	*3.7*	*3.7*	*3.7*	*3.5*



STATISTICS

	Line "t" Model	*Plane Model*	*Line "z" Model*
(degrees of freedom):	*6*	*5*	*6*
R^2 (Coeff. Determination):	*0.461*	*0.461*	*0.019*
Adjusted R^2:	*0.372*	*0.246*	*-0.144*
T-test of R^2: (Acceptable line fit if > 90%)	*93.6*		*25.6*
(intercept):	*17.8829586*	*16.0401293*	*46.3382482*
(slope for t):	*-0.00039031*	*-0.000391048*	
(slope for z):		*0.001220229*	*-0.02783217*
	y(t)	*y(t, z)*	*y(z)*
(Goodness of Fit): smaller is better	*0.05932*	*0.05932*	*0.10803*

Apparent Half-Life

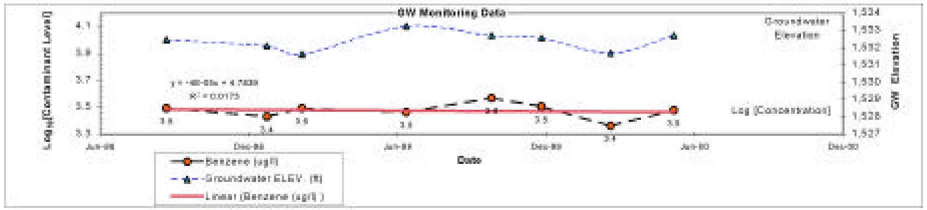
k	$t_{1/2}$ (d)	$t_{1/2}$ (yr)
-0.000898717	771	2.113

MW-8

NO, y(t) - time only line model - is preferred at a 90% level of confidence F test.

MW: ***PZ-1***

Groundwater ELEV. (ft)	*1532.45*	*1532.09*	*1531.63*	*1533.23*	*1532.69*	*1532.56*	*1531.69*	*1532.70*
Sampling Dates	*8/27/1998*	*12/28/1998*	*2/9/1999*	*6/17/1999*	*9/30/1999*	*11/30/1999*	*2/23/2000*	*5/11/2000*
Benzene (ug/l)	*3,100.0*	*2,700.0*	*3,100.0*	*2,900.0*	*3,700.0*	*3,200.0*	*2,300.0*	*3,000.0*
Log_{10} [Contaminant (ug/l)]	*3.5*	*3.4*	*3.5*	*3.5*	*3.6*	*3.5*	*3.4*	*3.5*



STATISTICS

	Line "t" Model	*Plane Model*	*Line "z" Model*
(degrees of freedom):	*6*	*5*	*6*
R^2 (Coeff. Determination):	*0.017*	*0.247*	*0.208*
Adjusted R^2:	*-0.146*	*-0.054*	*0.076*
T-test of R^2: (Acceptable line fit if > 90%)	*24.4*		*74.4*
(intercept):	*4.78386467*	*-75.93366819*	*-73.0866994*
(slope for t):	*-3.6036E-05*	*-5.50735E-05*	
(slope for z):		*0.053126338*	*0.04996299*
	y(t)	*y(t, z)*	*y(z)*
(Goodness of Fit): smaller is better	*0.02458*	*0.01883*	*0.01982*

Apparent Half-Life

k	$t_{1/2}$ (d)	$t_{1/2}$ (yr)
-8.2976E-05	8,354	22.887

PZ-1

Statistically, plane model is not necessary; y(z) with a larger R2 may suffice.

Confidence level for larger R2 not being zero is 74.4 %.

Figure 6. Data from water-table well MW-8 and adjacent piezometer PZ-1 at a site in Prentice. The symbols are the same as in figure 3. The top of the screen of PZ-1 is at least 5 ft deeper than the bottom of MW-8. Comparison between the groundwater elevations from the 2 wells indicates downward vertical hydraulic gradient for 3 of the 8 monitoring rounds (arrows above PZ-1 elevations in the figure). The analysis of MW-8 data yields a conclusion that is different when PZ-1 data are considered. For MW-8, the "t"-only model seems able to explain the decay in the concentrations with a short half-life $t_{1/2}$ of only 2 years. For PZ-1, the F-test "inconclusive" result regarding the plane model, leaves us with the "z"-only model that provides for a better R^2. The time-trend analysis of the PZ-1 concentrations suggest a much longer $t_{1/2}$ of 23 yrs for the decay of benzene in the groundwater.

increasing benzene concentrations. Four wells (from sites #4, 17, 22 and 23) are *'t'*-only wells that show decay in benzene and with $R^2 > 0.6$. These sites would have been good candidates for an additional 1-d batch-flush model, except that each has certain shortcomings. Site #4 had only 5 data points, and no piezometer nor water-table well downgradient of the source well. The groundwater flow map (with mean-sea-level-referenced elevations) in the GIS Registry for site # 4 was based not on UST-investigation wells (with locally referenced elevations), but rather on wells installed more than 200 ft (up- and side-gradient) from the former USTs. Site #17's most recent benzene concentration is very low (< 0.5 ug/l), so the comparison would not have any relevance. For sites #22 and 23, we show our follow up in figures 5 and 6, respectively. So in light of these few sites and limited site-specific parameters (such as hydraulic conductivity and fraction of organic carbon) available, we did not pursue the batch-flush model comparison.

Another study objective was to evaluate sites that were closed on the strength of passing the nonparametric statistics described in ch. NR 746, Wis. Adm. Code, in the light of results of the F-tests to see whether we would arrive at the same or a different conclusion. However, none of the 26 sites in our study were closed under NR 746. Nonetheless, we proceeded with the evaluation, but with a twist. We instead evaluated sites with wells with increasing benzene ("negative-$t_{1/2}$ wells") to see if a contradictory conclusion can be reached when the nonparametric statistical tests are applied. Table 2 has these comparisons. Because of the limited number of data points that can be included in the nonparametric tests (*i.e.,* 8 for the Mann-Whitney U), we can expect differences between the regression and the nonparametric test results, especially when the early concentrations (which are typically large for source wells, but can be the smaller concentrations for the downgradient wells) get omitted. We mentioned a notable contrary result for site #16, where both of the nonparametric tests concluded the benzene to be decreasing. We did a data sensitivity analysis for this site, and found that if we instead used the earliest 8 concentrations (rather than the most recent 8 of the 9 monitoring rounds) from site #16, the verdict between the 2 nonparametric tests would be split, with the Mann-Whitney U (M-W) test failing to show any trend while the Mann-Kendall (M-K) test favoring a decreasing trend.

The M-K test (at $\alpha=0.2$ level) needs only a *minimum* of 4 data points to detect a trend. To assess a possible bias, we used the DNR-supplied M-K spreadsheet, inputting only the most recent 4 concentrations from the negative-$t_{1/2}$ wells. Only one well showed an increasing trend (site # 19); two (2) wells showed decreasing trends (sites #15 and 22), and in 8 (of the 11) negative-$t_{1/2}$ wells, the M-K test fails to detect any trend in the data ("no trend" result). The interesting result was that none of the wells with the highest benzene concentrations was concluded to be "non-stable," but the well with the lowest benzene concentrations was concluded as nonstable. When the M-K test result is a "no trend," the spreadsheet proceeds to estimate the coefficient of variation (CV) of the concentrations. The CV extension would tend to yield different conclusions for the high concentrations than for low concentrations. The CV would favor tagging higher concentrations as stable. This can be easily shown by comparing two "no trend" results: a set of high numbers (4300, 5500, 8000, 5500 from site #24), and a set of low numbers (85, 393, 22, 45 from site #20). The absolute difference among the larger-number set is in the 1000's, and among the smaller numbers, at least an order of magnitude less. However, the CV for the larger-number set (CV = 0.3) is less than the smaller number set's (1.3), so stability when defined as being $CV \leq 1$ would favor the larger-number set. The CV extension can provide the wrong impression when the M-K test has already failed to discern a trend. The M-K procedure can be made more robust by increasing the number of data to 8 (just like for the M-W), and simplifying the procedure by dropping the extension involving CV.

Averting the misapplication of the nonparametric tests was one of our primary purposes in devising the F-test technique. Reliance on nonparametric statistic test using the concentration data alone to assess a plume's stability can be misleading. Doing so would ignore key factors in assessing a plume's behavior, such as a site's hydrogeology (*e.g.,* downward vertical hydraulic gradients) and changes in the horizontal flow direction, and in assessing concentration data in light of well placement and construction within the plume. We recommend that the F-test be a prescreening stage before nonparametric tests are used at sites. When the F-test shows that *'z'* is a factor (plane, *'z'*-only, and inconclusive but *'z'*-only has larger R^2), then we know that this *invariant* factor is affecting the concentrations; and hence, a nonparametric test is not appropriate. When prescreened by its F-test result, site #16's data would not have qualified for nonparametric tests. On the other hand, when the F-test shows that *'z'* can be ignored (*i.e.*, test result of either: *'t'*-only , or inconclusive but the *'t'*-only has larger R2), then nonparametric statistics may be more appropriately used.

We found a wealth of information on closed sites in the WDNR GIS Registry. However, most sites did not have *both* benzene and groundwater elevation data. Commonly the groundwater elevation data are not included. Reviewers would be able to verify if the flow direction remain unchanged with the groundwater elevation data. Our analysis would have benefited had the elevations of the top and bottom of the well's screen been available in the groundwater elevation tables for all sites. For instance, as a means to review (or "QA/QC") the concentration data from a site, the monitoring well screen (top and bottom) elevations would have provided for an estimate of the volume of water purged from the well prior to collecting a sample. For the few wells where well construction information (*i.e.,* elevation of well screen's top and bottom) was available, it became obvious why less benzene was found at certain times, because at those times, the water table had risen above the well screen. Hence inclusion of well construction information in the GIS Registry submittal would improve any future QA/QC of the data for the closed sites.

As a follow-up to this study, we are currently working on a technique to estimate the empirical parameters $\mathbf{t_{1/2}}$ and $\mathbf{x_{1/2}}$, the half-life and the half-distance, respectively, for the groundwater contaminant at a site by using the intercepts and slopes of the regression lines at a near-source and a downgradient well. These empirical parameters, together with the hydrogeological parameters, would help determine whether steady state has been reached. Moreover with $\mathbf{t_{1/2}}$ and $\mathbf{x_{1/2}}$, we can estimate the contaminant level we would expect after a given elapsed time at a particular location downgradient from the source. This technique can be useful for verification purposes at closed sites where monitoring wells have been abandoned, but substantial contamination is still present.

REFERENCES

Buscheck, T. E. and C. M. Alcantar, Regression techniques and analytical solutions to demonstrate intrinsic bioremediation, in: *Intrinsic Bioremediation*, edited by R. E. Hinchee, J. T. Wilson and D. C. Downey, Columbus, OH: Battelle Press, p. 106-116, 1995.

McAllister, P. M., & C. Y. Chiang, A practical approach to evaluating natural attenuation of contaminants in ground water. *Ground Water Monitoring and Remediation*, *14*, p. 161-173, Spring 1994.

McNab, W. W. Jr., B. P. Dooher, A critique of steady-state analytical methods for estimating contaminant degradation rates, *Ground Water*, *36*, p. 983-987, 1998.

National Research Council. *In Situ Bioremediation - When Does It Work?*, Washington, DC: National Academy Press, 207 pp., 1993.

Odermatt, J. R., Remediation by natural attenuation?, *Ground Water Monitoring and Remediation*, *19*, p. 58-60, 1999.

Pelayo, A. M. and F. S. Evangelista, A statistical F test for the natural attenuation of contaminants in groundwater, *Environmental Monitoring and Assessment*, *83*, p. 47-70, 2003.

Wisconsin Department of Natural Resources, Bureau for Remediation and Redevelopment Tracking System (BRRTS), http://www.dnr.state.wi.us/org/aw/rr/brrts/index.htm, visited September 30, 2003.

Wisconsin Department of Natural Resources, GIS Registry of Closed Remediation Sites, http://gomapout.dnr.state.wi.us/org/at/et/geo/gwur, visited September 30, 2003.

Weisberg, S., *Applied Linear Regression*, *2nd E*d., New York: John Wiley & Sons, 324 pp., 1985.

DISCLAIMER