MANAGEMENT OF DATA PRIOR TO PERFORMING AN ANALYSIS OF PLUME STABILITY

(And One Other Assorted Topic)

David Powell, RG, CHMM
Sunbelt Environmental Services, Inc.

Pretty Much How I Feel Every Time I Submit A Plume Stability Analysis

Your CV’s Are High Your R2 #’s Are Low
Your Conclusions Mean Nothing

And You Should Feel Bad

Types of Datasets

- Perfect Datasets
- Less Than Perfect Datasets
- Far From Perfect Datasets

What Makes Less Than or Far From Perfect Datasets

- Results of creating datasets with issues
- How to manage these issues prior to starting the plume stability analysis
- Non-Detect Data

Other Assorted Topic – Using a Ricker Method Plume Stability Analysis For Sites With Free Product Left At the Site

Perfect Datasets

- All data collected within the past 2 years
- Quarterly data with no missing quarters
- At least 6 to 8 results
- All non-detects have the same reporting limits
- Monitoring wells are evenly spaced without large gaps in coverage (for isopleths or Ricker)

How to Handle Perfect Datasets:

Step 1 - Congratulate yourself on your hard work and good fortune.

Boom – Done

Proceed to COC concentration stability analysis.

However,

Perfect Datasets are about as common as:

BANK ERRORS IN YOUR FAVOR

UNI-KITTIES

TRANSPARANCY IN GOVERNMENT

LONG-OVERDUE HOGWARTS LETTER

Less Than, or Far From Perfect Datasets

- Quarterly Data, but some quarters were skipped
- Less than 6 events
- Oldest data is more than 3 years old
- One or two gaps in monitoring well network (for Isopleths or Ricker)
- Many “Non-Detect” Results

Quarterly Data, but some quarters were skipped:

- Will not have a significant impact on the results of Linear Regression (or their variants), Mann-Kendall, or a Ricker Analysis.

– However –

- It is important to have seasonal representation.

- Unknown impacts might be able to be observed and rectified before significant impact can occur. e.g. a broken well flush mount on a monitoring well in an area that gets runoff from a gas station lot.

- Off-site, or additional on-site impacts can be identified closer to when they occur, rather than long after the fact.

Less Than, or Far From Perfect Datasets

- Missing Data

The well was dry

The lid was stuck

Couldn’t be found

Well contained a dangerous spider

Lid has been covered with:

- Asphalt
- Gravel
- Concrete
- Grass
- Debris

Vicious dog on the property

Missing Data:

Cause of Many PSA Headaches.

- Missing data will keep some wells from being included in your analysis.
- Missing data contributes to the possibility of:
  - Low R^2 numbers on trendlines on linear regression charts
  - Potentially false trend lines on linear regression charts
  - High CV values in Mann/Kendall

Can be dealt with by interpolation or extrapolation, which will be covered later.

Less Than 6 Events:

- Simple Linear Regression requires a minimum of three events to be completed.

- MRBCA guidance requires a minimum of four events for plume stability analysis.

- ITRC guidance on groundwater statistics calls for a minimum of eight events for linear regression.

More Data = Greater Reliability.

- "Not-So-Simple" Linear Regression – Based on the slope of the confidence intervals around the regression line and can be performed in Excel.

Less Than 6 Events:

- Mann-Kendall requires a minimum of four events to be completed.

- 6 Events are required by MDNR when the result is no-trend and the CV test for stability is applied.

More Data = Greater Reliability.

Less Than 6 Events:

- Ricker only needs a minimum of three because the trends of plume characteristics are evaluated using Linear Regression.

- MRBCA guidance requires a minimum of four events for plume stability analysis.

More Data = Greater Reliability.

- Plume Statistics Generated by the Ricker Method Can Be Analyzed With Mann-Kendall. (Requiring 4 events).

Oldest Data is More Than 3 Years Old:

The statistical methods we use for plume stability analysis do not care when your samples were collected.

- Linear Regression will still make reliable trends regardless of the age of the samples, the x-axis scale will just be larger.

- Mann-Kendall is a non-parametric method so dates really don't matter at all.

- Since the Ricker uses Linear Regression for the analysis of plume trends, the same rule applies.

HOWEVER:

Large gaps in between sample dates create what I am going to call:
Qualitative Uncertainty

- Unknown impacts to the groundwater you are evaluating could have taken place at your site during the gap in monitoring.

- Large gaps in sampling dates make it more difficult to identify seasonal influences on groundwater surface elevations. This is more significant if LNAPL is present at your site. You may be missing sampling events where static water levels were low enough to allow LNAPL to flow into screened areas of wells.

- Monitoring wells may become damaged and allow for the infiltration of surface runoff for long periods of time unnoticed.

MORE IMPORTANTLY

- **MRBCA guidance calls for 1 to 3 years of quarterly data, with 2 years being the expectation.**

- The use of data older than 3 years may be (and often is) accepted by MDNR.

  - If your data is older than 3 years, send an argument to MDNR first before even starting the plume stability analysis.

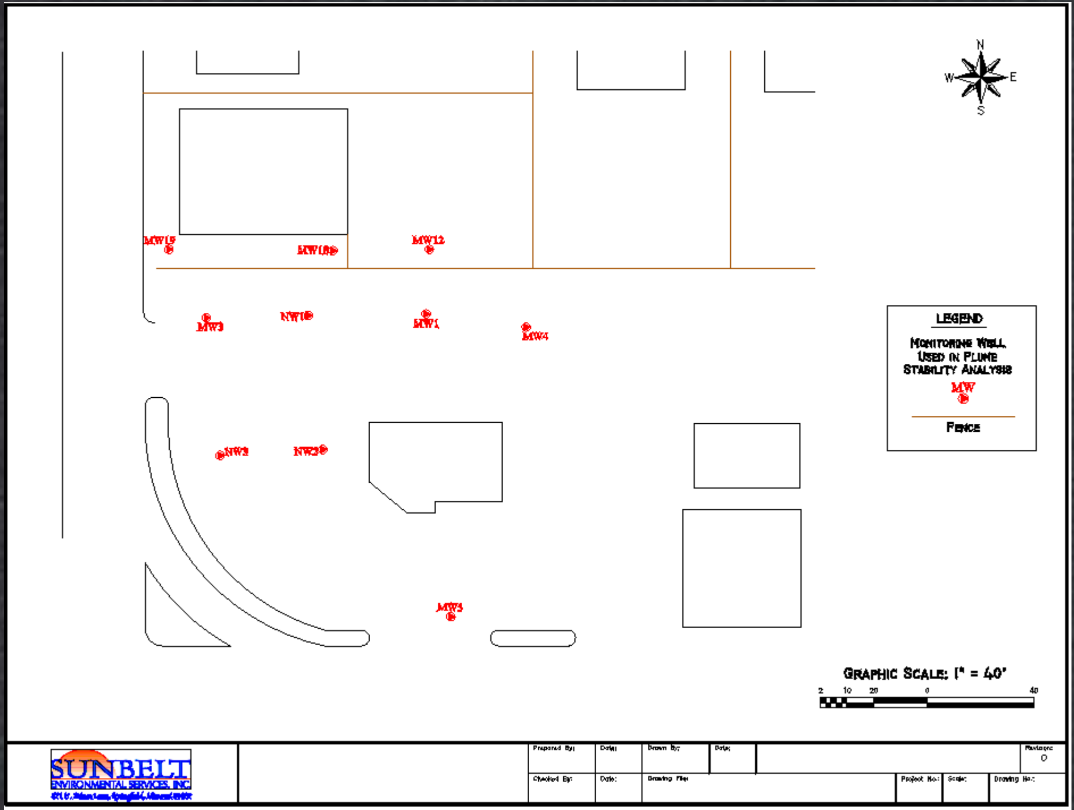
  - Applies to both Mann-Kendall and Linear Regression.

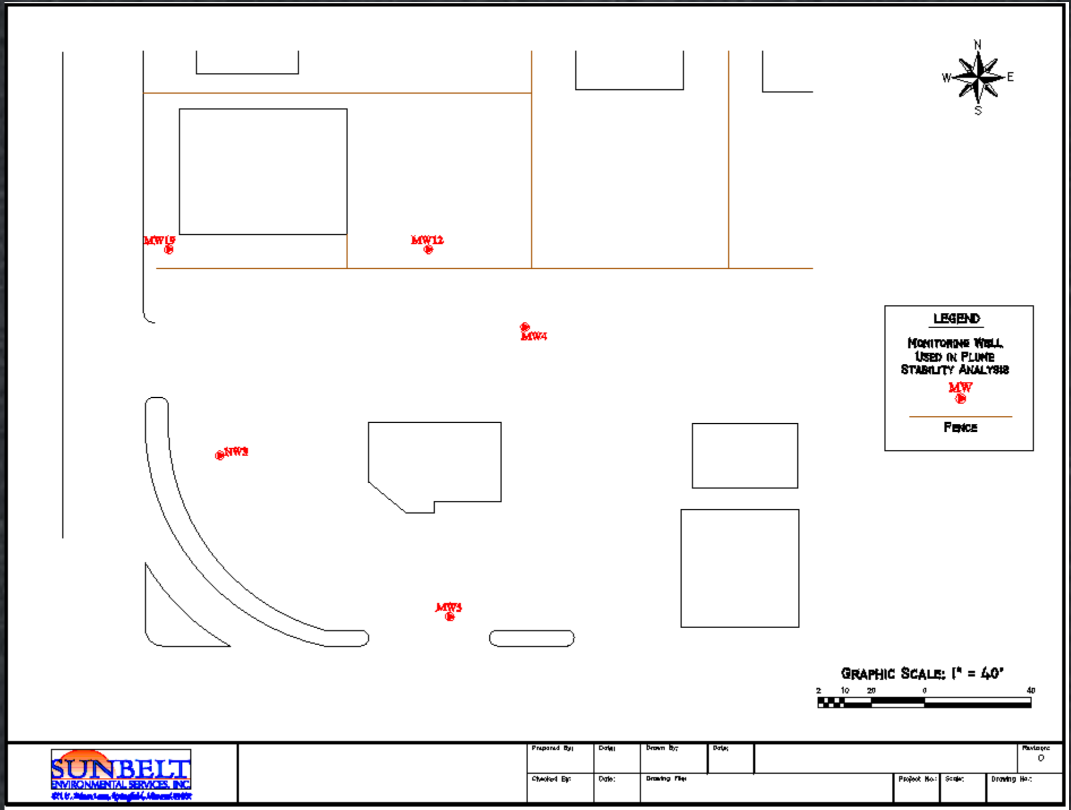
  - POST-REMEDIATION DATA ONLY!

Gaps in Monitoring Well Network (For Ricker Method Analysis)

- Surfer© calculates plume parameters using x/y/z coordinates. The more x/y/z data points, the less the software needs to interpolate/extrapolate.

This Site Should Have More Reliable Results Than: This Site





Gaps in Monitoring Well Network (For Ricker Method Analysis)

- Outlier wells can be left out of the Ricker analysis if COCs are consistently Non-Detects or below DTLs.
  - HOWEVER, outlier wells can be important to defining the plume boundary. Only exclude if there are multiple non-detects between your plume edge and the outlier wells.
- If a monitoring well must be included in the analysis, but there is a significant gap between it and other wells, x/y coordinates and concentrations of "ghost points" can be interpolated between wells.
  - HOWEVER, this entails a lot of extra work, and will likely not be accepted by MDNR.

Strategies For Managing Problems With Datasets

Really only two problems you can do anything about **after** the fact.

- Missing Data
- Non-Detect Data

You Can Replace Missing Data

Interpolation/Extrapolation of Missing Data

You can use your Non-Detect Data

Using non-detects in plume stability analysis

INTERPOLATION

A general guide for when it is acceptable to interpolate data:

If there are at least the same number of good data points on either side of the missing data as there are missing data points, it is ok to interpolate the missing data.

9.84E-01	3.46E-01	NA	4.55E-01	7.46E-02	7.20E-02	4.10E-02			YES
9.84E-01	3.46E-01	NA	NA	7.46E-02	7.20E-02	4.10E-02			YES
4.10E-02	9.84E-01	3.46E-01	NA	NA	NA	7.20E-02	4.10E-02	4.10E-02	YES
9.84E-01	NA	NA	NA	NA	7.20E-02	4.10E-02			NO
9.84E-01	3.46E-01	NA	4.55E-01	7.46E-02	NA	4.10E-02	4.10E-02		YES
9.84E-01	3.46E-01	NA	NA	7.46E-02	NA	4.10E-02			NO

NA	=NO DATA

INTERPOLATION FORMULA

= DATA PRECEEDING MISSING DATA + { DATA FOLLOWING MISSING DATA - DATA PRECEEDING MISSING DATA } / NUMBER OF CELLS TO REACH DATA FOLLOWING MISSING DATA FROM DATA PRECEEDING MISSING DATA

Example 1:
1 Missing Data Point

C3 | =B3+(D3-B3)/2

	A	B	C	D	E	F	G
1					**Extrapolation**		
2							
3	9.84E-01	3.46E-01	**4.01E-01**	4.55E-01	7.46E-02	7.20E-02	4.10E-02

=B3+(D3-B3)/2

Example 2:
2 Missing Data Points

	A	B	C	D	E	F	G	H
1								
2	9.84E-01	3.46E-01	2.56E-01	1.65E-01	7.46E-02	7.20E-02	4.10E-02	
3								
4								
5								

2.56E-01	=B2+(E2-B2)/3
1.65E-01	=C2+(E2-C2)/2

!! HANDY EXCEL TIP !!

Putting a $ after a row or column letter/number anchors that cell in the formula.
For Example:

=C2+(E2-C2)/2

Cell E2 will not change when you copy the formula.

Example 3:
3 Missing Data Points

	A	B	C	D	E	F	G	H	I	J
1										
2	4.10E-02	9.84E-01	3.46E-01	2.78E-01	2.09E-01	1.41E-01	7.20E-02	4.10E-02	4.10E-02	
3										
4										

2.56E-01	=C2+(G2-C2)/4
2.10E-01	=D2+(G2-D2)/3
1.41E-01	=E2+(G2-E2)/2

Note: I only anchor the data point following the missing data (G2).
The second cell in the parenthesis is the new "data point preceding the missing data" calculated by the interpolation formula (C2, D2, and E2).

INTERPOLATION OF DATA BETWEEN MONITORING WELLS

Some reasons to possibly interpolate data between monitoring wells:

1. Your plume goes under a street and you have data points on either side.
2. There is a large structure on your site and your plume is under it with data points surrounding it.
3. If it is important to know approximate location of where concentrations are delineated.
4. You would like to increase the amount of data to use in a Ricker to increase precision.
5. More useful for larger sites.

I have not, and likely will not use this method for my sites.

While not specifically called out in the guidance, discussion with MDNR project managers has indicated that "Ghost Points" would not likely be considered.

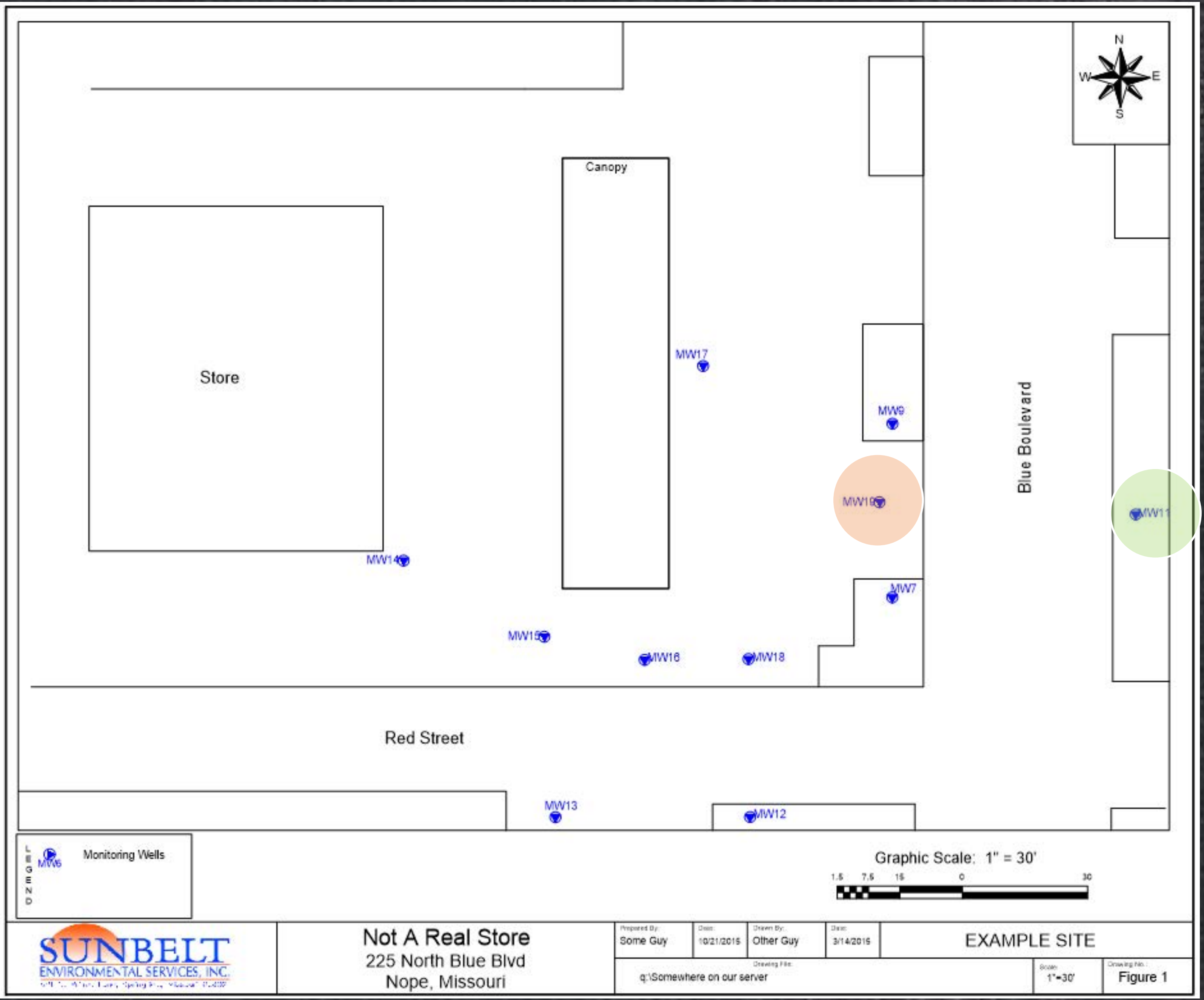
INTERPOLATION FORMULA #2

= (DATA ONE - DATA TWO) / NUMBER OF CELLS TO REACH NEXT DATA INCLUCING INSERTED CELLS + PREVIOUS DATA

USE $ Between Before Row Number to Anchor Data One In The Formula

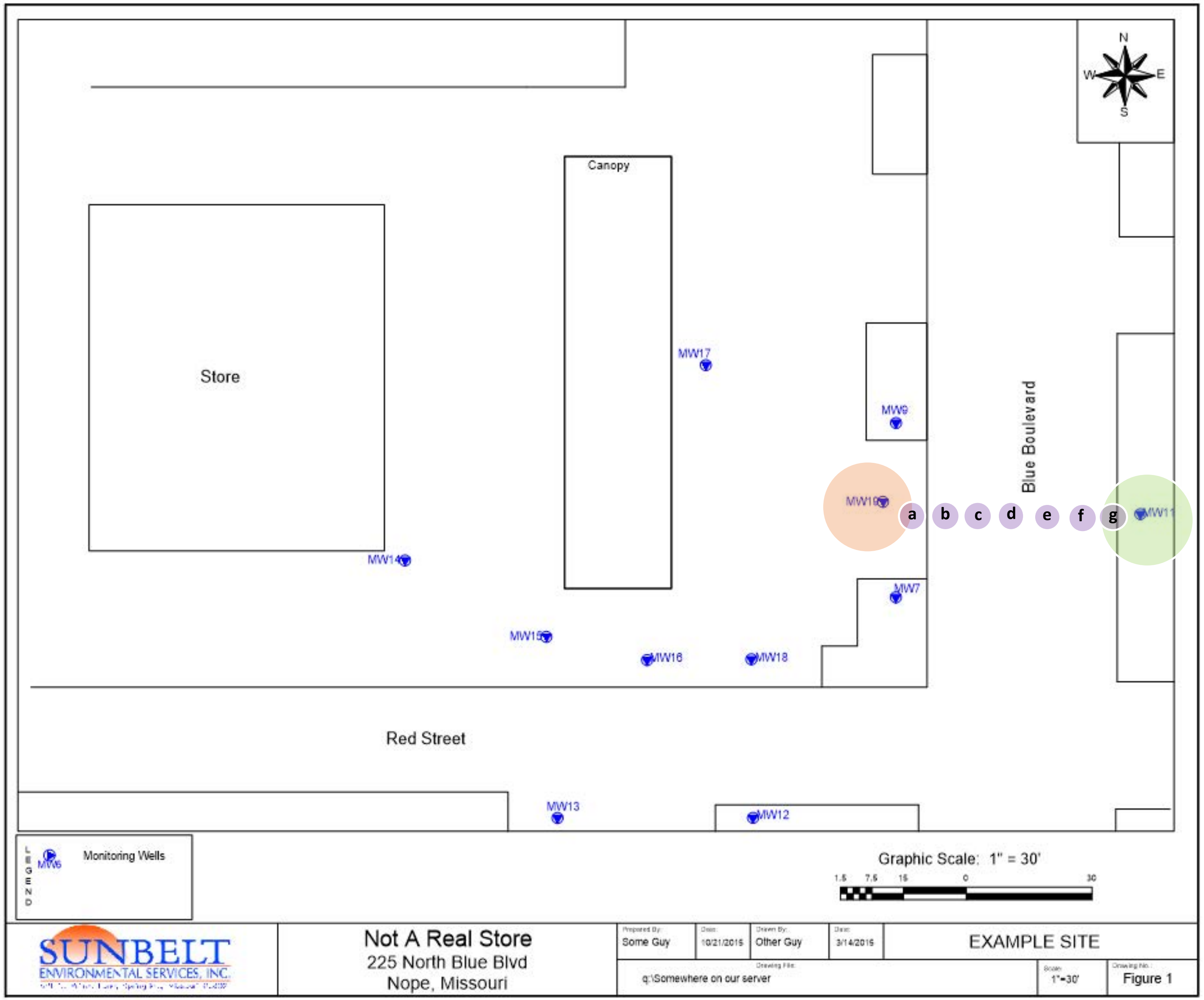
Use $ Before Row AND Column Numbers ex. C3

Example:



N
W
E
S
Canopy
Store
MW17
MW9
Blue Boulevard
MW10
MW11
MW14
MW7
MW15
MW16
MW18
Red Street
MW13
MW12
Monitoring Wells
Graphic Scale: 1" = 30'
SUNBELT
ENVIRONMENTAL SERVICES, INC.
Not A Real Store
225 North Blue Blvd
Nope, Missouri
Some Guy
10/21/2015
Other Guy
3/14/2015
EXAMPLE SITE
q:\Somewhere on our server
1"=30'
Figure 1

Example:



N
W
E
S
Canopy
Store
MW17
MW9
Blue Boulevard
MW10
a
b
c
d
e
f
g
MW11
MW14
MW7
MW15
MW16
MW18
Red Street
MW13
MW12
LEGEND
MW6
Monitoring Wells
Graphic Scale: 1" = 30'
SUNBELT
ENVIRONMENTAL SERVICES, INC.
Not A Real Store
225 North Blue Blvd
Nope, Missouri
Some Guy
10/21/2015
Other Guy
3/14/2015
EXAMPLE SITE
q:\Somewhere on our server
1"=30'
Figure 1

Example:

=(B10-B2)/8+B2

=(C10-C2)/8+C4

=(D10-D2)/8+D7

	A	B	C	D
1	**Well**	**X**	**Y**	**Concentration (mg/L)**
2	MW19	4216694.532	4098282.318	1.500
3	a	4216724.577	4098278.899	1.375
4	b	4216754.622	4098275.48	1.250
5	c	4216784.666	4098272.061	1.125
6	d	4216814.711	4098268.643	1.000
7	e	4216844.756	4098265.224	0.875
8	f	4216874.801	4098261.805	0.750
9	g	4216904.845	4098258.386	0.625
10	MW11	4216934.89	4098254.967	0.500

EXTRAPOLATION

When it is acceptable to extrapolate data:

FORWARD

To replace a missing data point in you most recent event only.

No future predicting for plume stability except possibly to demonstrate the likelihood of eventually reaching statistically reliable results.

Talk to MDNR Prior to Doing It.

BACKWARD

To replace missing data for a well because it was installed later in the site characterization process.

Example:

FORWARD EXTRAPOLATION

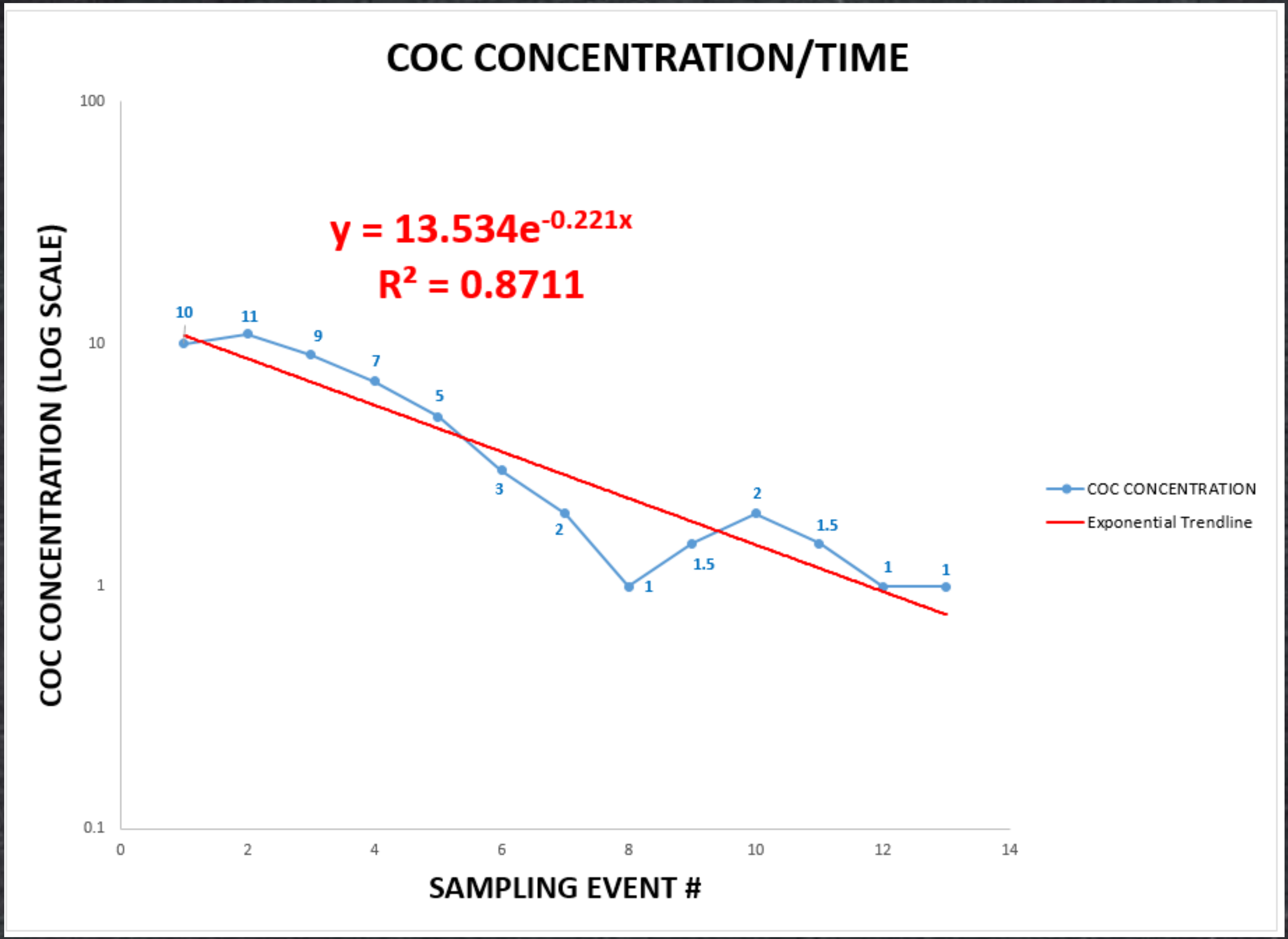
DATASET

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	**Sampling Event Number**												
2	1	2	3	4	5	6	7	8	9	10	11	12	13
3	**Concentration**												
4	10	11	9	7	5	3	2	1	1.5	2	1.5	1	1
5													
6													

!! Dates are changed to Sampling Event Numbers !!

Example:

FORWARD EXTRAPOLATION



LINEAR REGRESSION GRAPH WITH LOGARYTHM Y-AXIS AND EXPONENTIAL TRENDLINE, EQUATION DISPLAYED

The equation has the information we want.

I have just formatted the chart for this demonstration.

Example:

TRENDLINE EQUATION

$$y = 13.534e^{-0.221x}$$

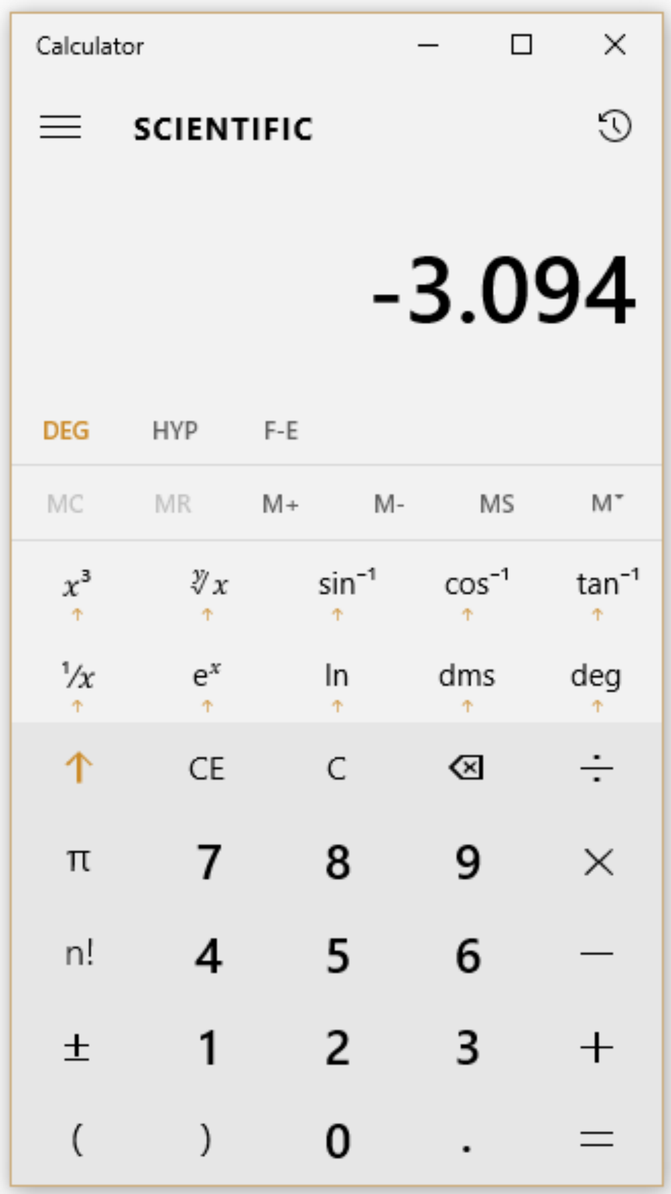
The last sampling event is #13.

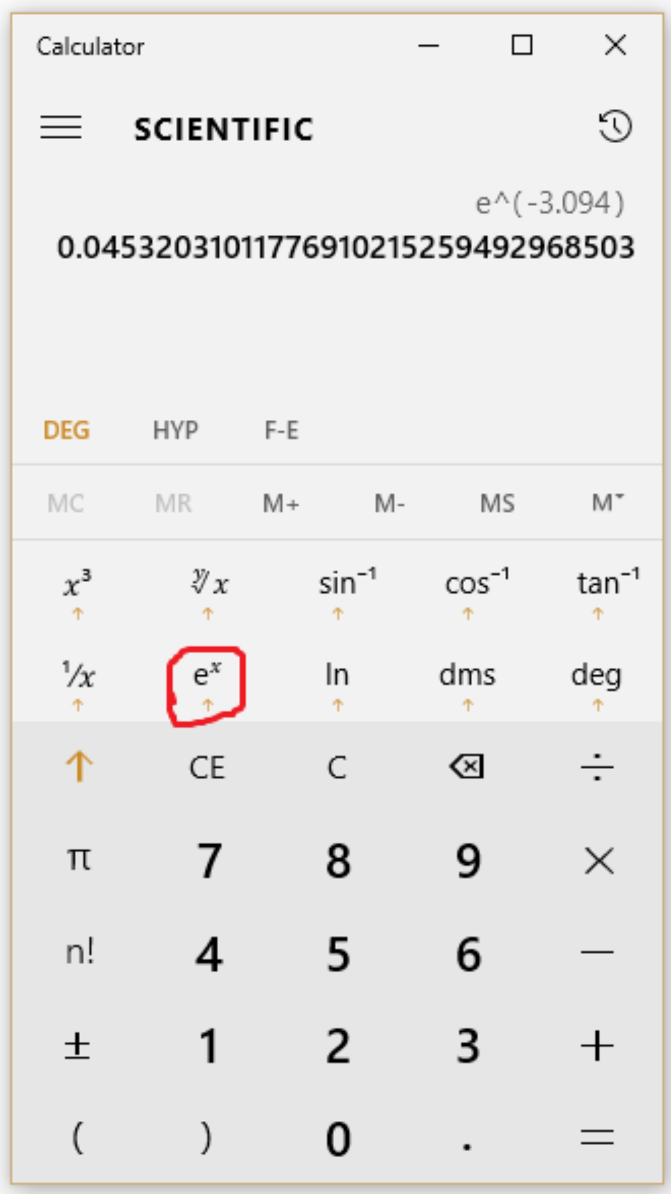
We will extrapolate 2 numbers to demonstrate how it works.

! TRY TO AVOID FORWARD EXTRAPOLATION !

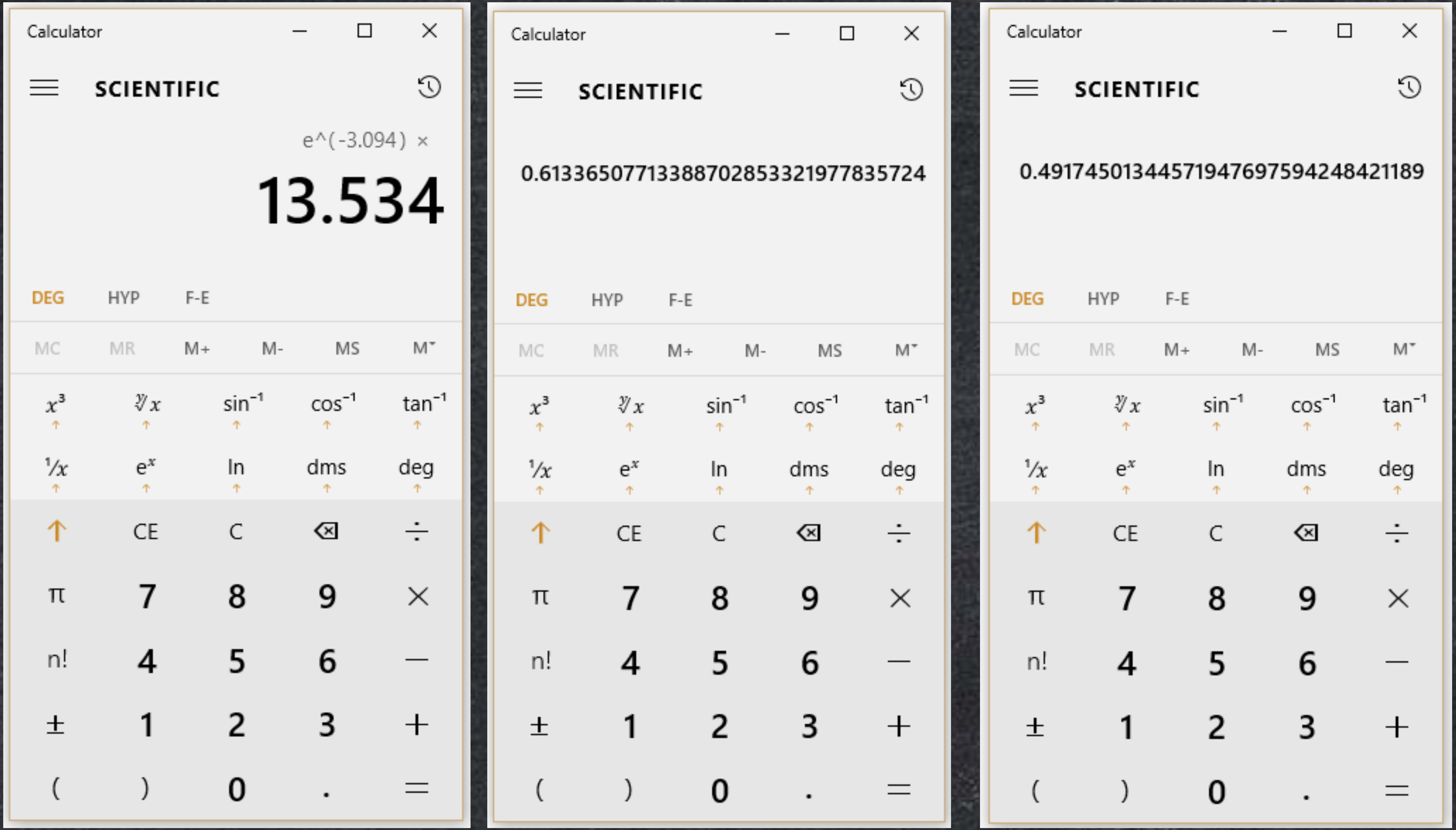
I would not suggest trying to actually use more than one extrapolated value for a PSA, and even then, talk to MDNR before doing it.

Solve the equation using your next two sampling events, #'s 14 & 15.





Done on a Windows 10 Calculator


Calculator
SCIENTIFIC
e^(-3.094) ×
13.534
Calculator
SCIENTIFIC
0.61336507713388702853321977835724
Calculator
SCIENTIFIC
0.49174501344571947697594248421189

BACKWARD EXTRAPOLATION

TRENDLINE EQUATION

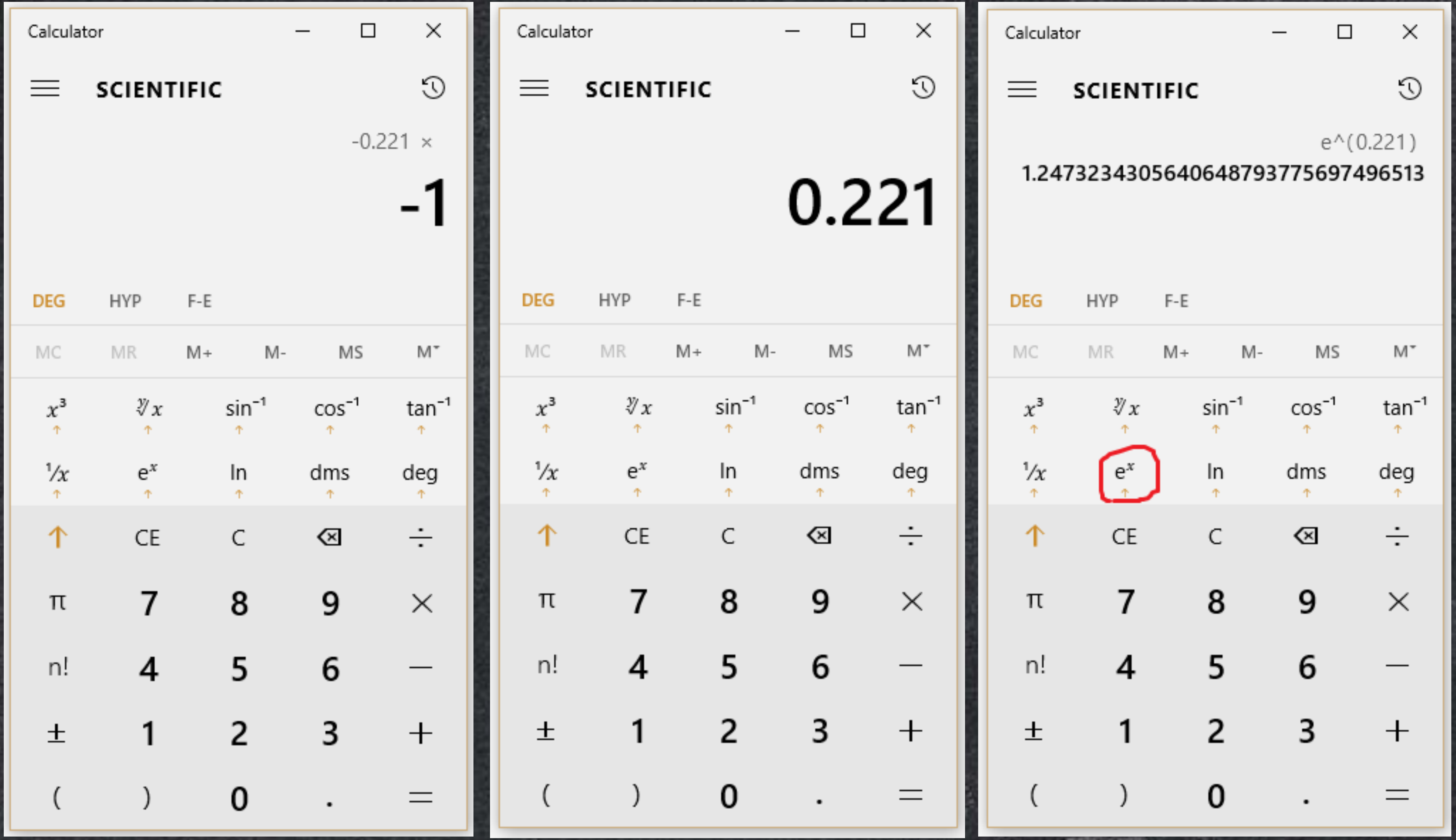
$$y = 13.534e^{-0.221x}$$

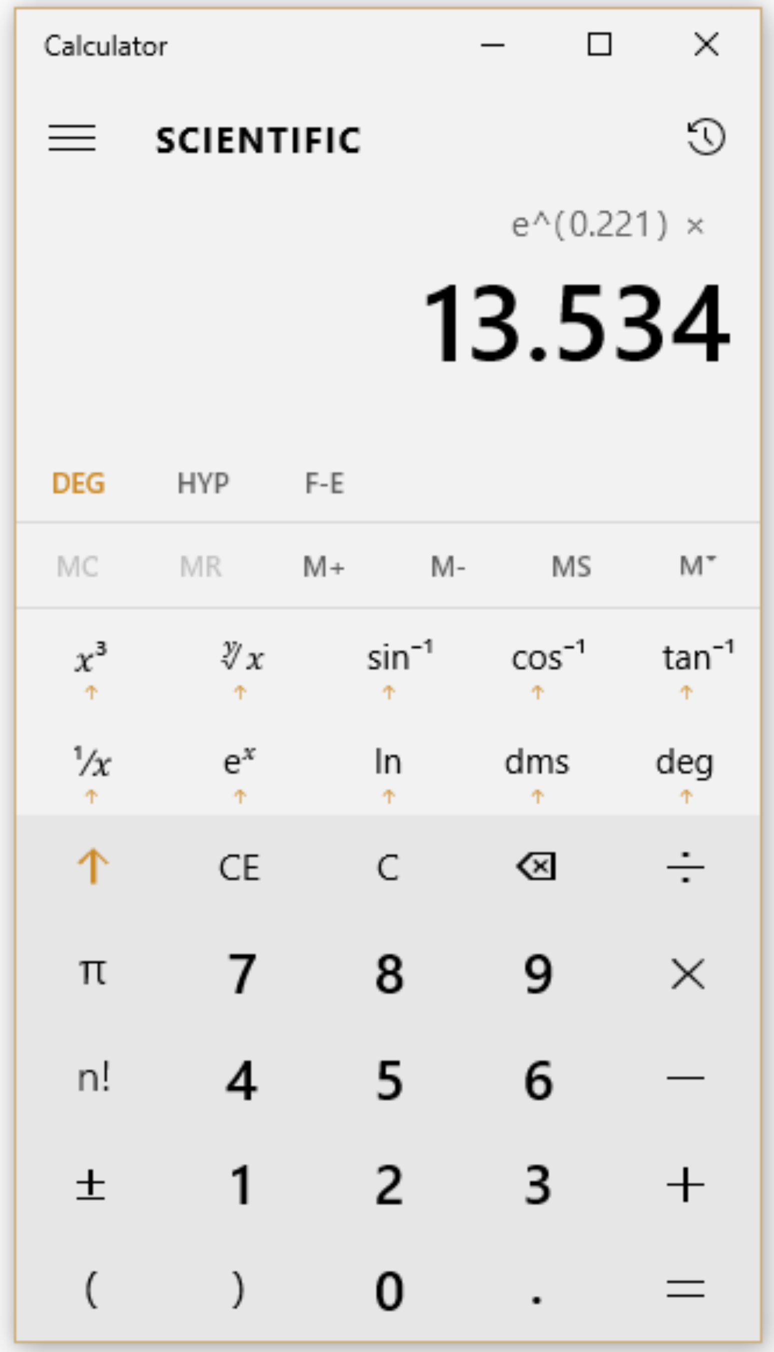
The first sampling event is #1.

We will back-extrapolate 2 numbers to demonstrate how it works.

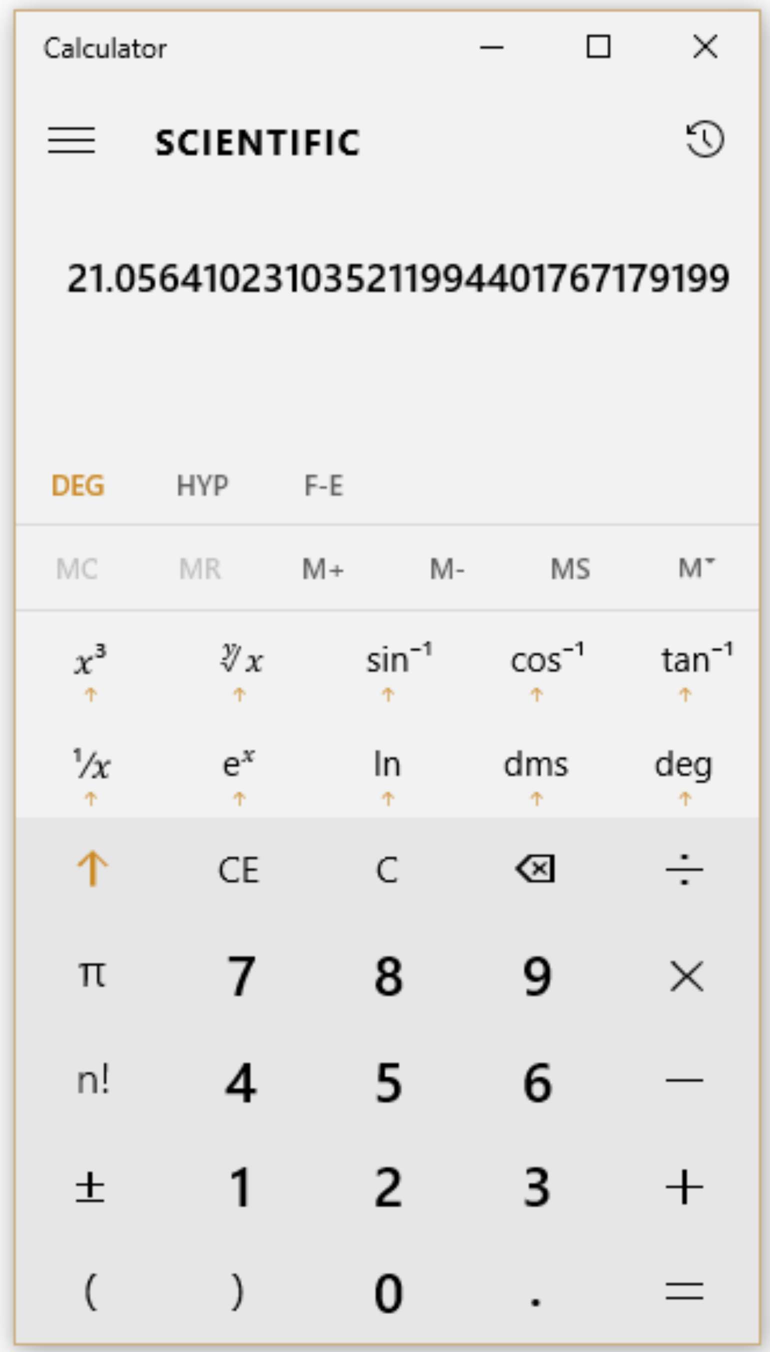
If you are planning to do more that a couple of wells it would be a good idea to talk to MDNR before doing it.

Solve the equation using the previous two sampling events, #'s -1 & -2.

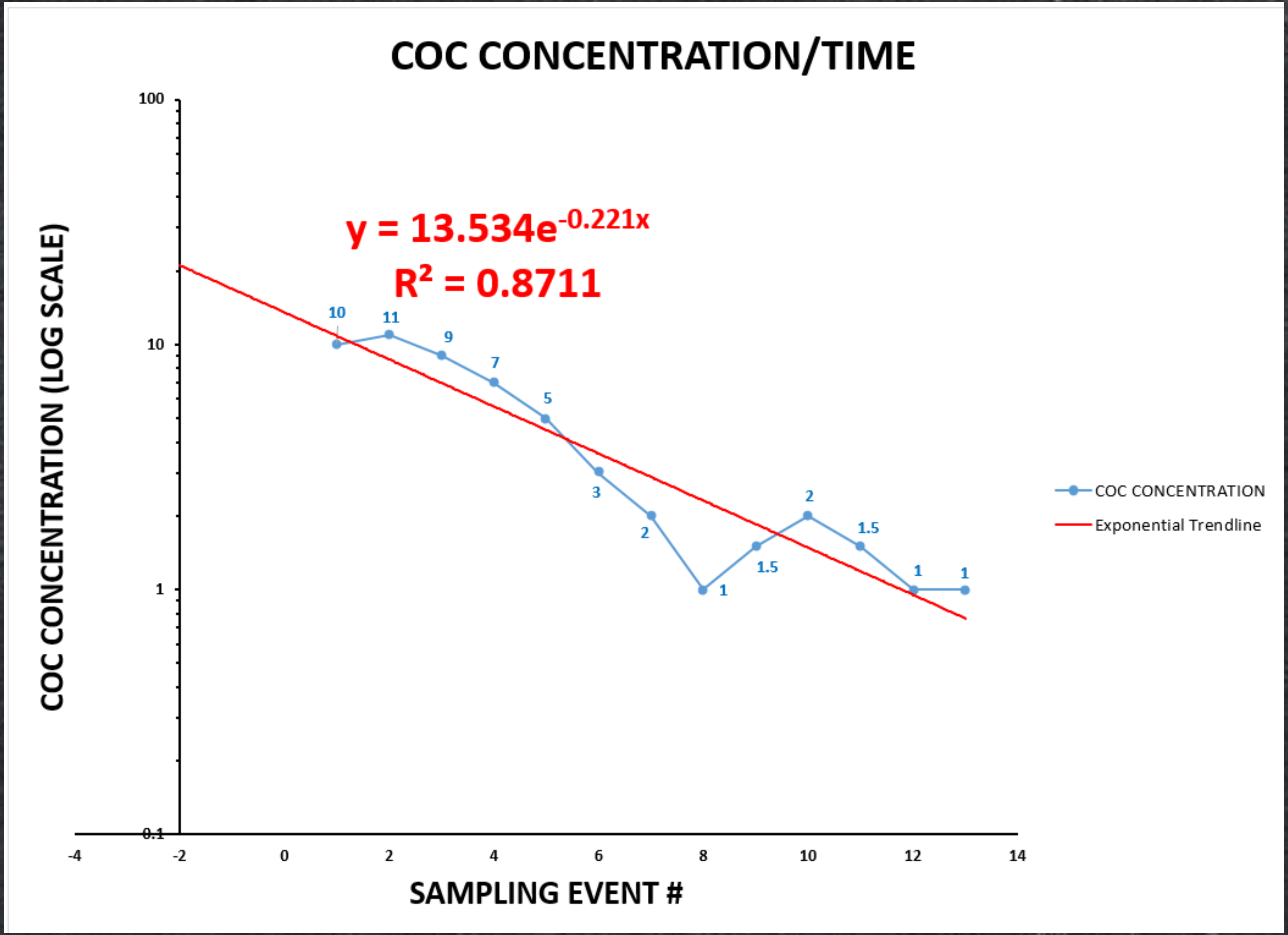

Calculator
SCIENTIFIC
-0.221 ×
-1
Calculator
SCIENTIFIC
0.221
Calculator
SCIENTIFIC
e^(0.221)
1.2473234305640648793775697496513


Calculator
SCIENTIFIC
e^(0.221) ×
13.534


Calculator
SCIENTIFIC
16.88127530925405407749602899178


Calculator
SCIENTIFIC
21.056410231035211994401767179199

So 16 and 21 for the previous 2 events.



Now that we know how to interpolate and extrapolate, what is an acceptable percentage of interpolated/extrapolated data?

Ask the regulatory agency prior to getting your data ready.

My guess would be <20% interpolated and an extremely limited amount of extrapolated data.

NON-DETECT DATA

During Site Characterization & Groundwater Monitoring:

All these concentrations are below detection limits!

During Risk Assessment:

All these concentrations are below detection limits!

During Plume Sability:

All these Non-Detects are messing up my PSA!

PROBLEMS NON-DETECTS CAUSE WITH PLUME STABILITY

Instructions: To use the spreadsheet, provide at least four rounds and up to ten round concentration units. All non-detect values should be assigned a single value, less than data entry error may cause "DATA ERR" to be displayed. Dates that are not consecuti and decreasing trends at 80% and 90% confidence levels. If an increasing or decreasi proposed by Wiedemeier, et al (2000), *Designing Monitoring Programs to Effectively Eva* will print both the data analysis sheet and the plot of concentration trends.

This spreadsheet is adapted from State of Wisconsin DNR, Remediation and Redevelop

Site Name =		City =
	Monitoirng Well & Compound	MW1 Benzene
Event Number	Sampling Date (most recent last)	Concentration (leave blank if no data)
1	29-Mar-14	0.00010
2	24-Oct-14	0.00030
3	28-May-15	0.00010
4	13-Nov-15	0.00029
5	12-Feb-16	0.00050
6	29-May-16	0.00031
7		
8		
9		
10		
	Mann Kendall Statistic S =	8
	Number of Rounds n =	6
	Average =	0.00
	Standard Deviation =	0.00
	Coefficient of Variation (CV) =	0.57
Trend ≥ 80% Confidence Level		INCREASING
Trend ≥ 90% Confidence Level		No Trend
Stability Test, If No Trend Exists at 80% Confidence Level		NA
Error Check, Blank If no Errors Detected		
	Data Entry By =	D. Powell
	Concentration Units =	mg/L

Problem Non-Detects Cause for Mann-Kendall Analysis.

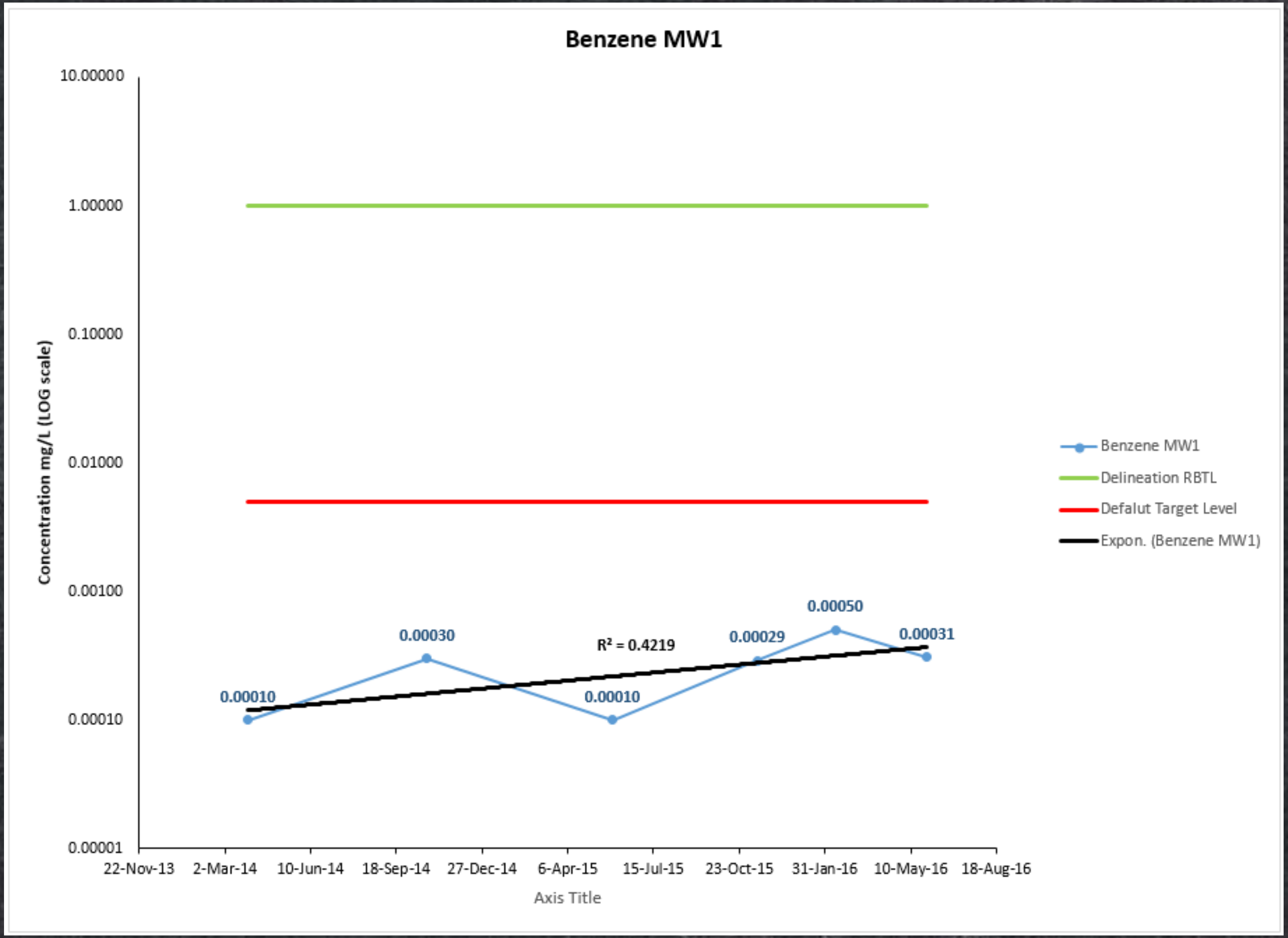
Example:

Data for detected concentrations of benzene in MW1 are **In Green** and non-detect results have been replaced with ½ of reporting limits.

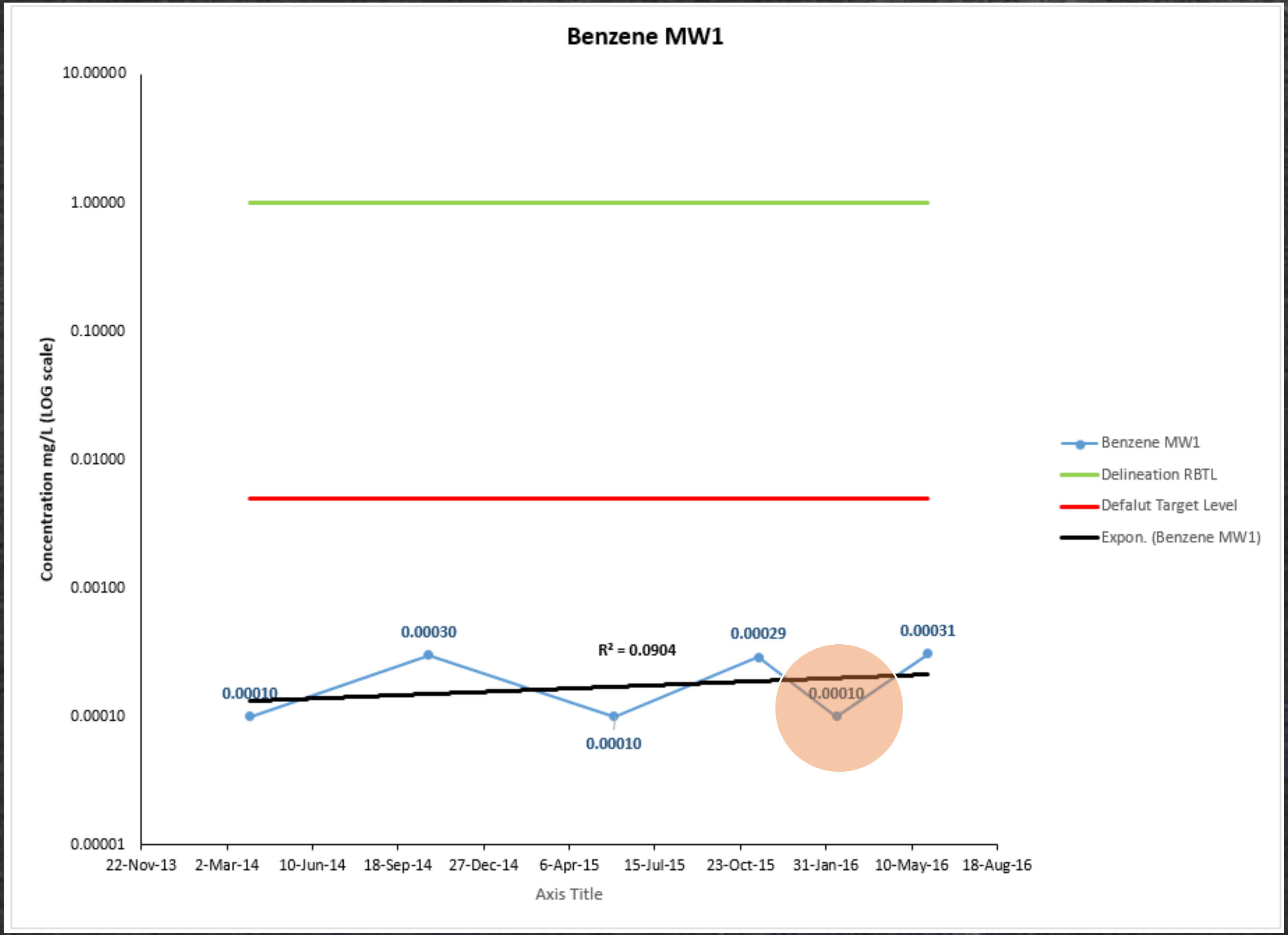
The reporting limits for Benzene were different on 2/12/16 due to a higher concentration of TPH-GRO during that sampling event.

Because the reporting limits during the first events are lower, an increasing trend result is returned even though the actual concentrations are extremely low.

Same Data Used In A Linear Regression Chart



Same Data Used In A Linear Regression Chart



Changing the non-detect result to the typical value lowers the angle of the trendline significantly & reduces the R^2 Number

Guidance Available for Handling Non-Detects

ITRC (Interstate Technology & Regulatory Council). 2013. Groundwater Statistics and Monitoring Compliance, Statistical Tools for the Project Life Cycle. GSMC-1. Washington, D.C.: Interstate Technology & Regulatory Council, Groundwater Statistics and Monitoring Compliance Team

http://www.itrcweb.org/gsmc-1/

The IRTC Document has copious amounts of information on handling non-detects in Section 5.7 Managing Nondetects in Statistical Analyses

Most of Section 5.7 is for using non-detects in more complicated statistical analysis.

For the Majority of My Projects – Simple Substitution

Typically use ½ of the Laboratory Detection Limit

Using maximum laboratory detection limits is good for the calculation of representative concentrations, but should not be used for plume stability analysis.

– From the IRTC Guidance –

"EPA's Unified Guidance suggests that the substitution method can be acceptable when only a small portion of the data set (10-15 percent) consists of nondetects. When the nondetect proportion is quite low, statistical results based on using simple substitution are not likely to vary substantially from other methods."

OTHER ASSORTED TOPIC

Ricker Method Plume Stability At Sites With LNAPL/Free Product in Wells

! This is new information !

I am including it in my presentation because it is pertinent to Plume Stability Analysis and I have recently discussed it with MDNR.

Ricker Method Plume Stability At Sites With LNAPL/Free Product in Wells

Condition 1 of MRBCA Sections 7.6, 8.4, and 9.5 (confirmation that the plume is stable or decreasing) may be demonstrated in a Ricker Method Analysis of Plume Stability in the following manner:

OPTION #1

If free product is limited to a maximum of 2 wells, and there is a measurable dissolved-phase concentration plume:

Run a single Ricker Method Plume Stability Analysis that COMPLETELY excludes the monitoring wells with free product, whether consistently or intermittently.

Ricker Method Plume Stability At Sites With LNAPL/Free Product in Wells

OPTION #2

If free product is present in greater than two wells, and/or a measureable dissolved-phase concentration plume is limited to one or zero monitoring wells:

Consult MDNR on how to conduct a site-specific plume stability demonstration.

Additional plume stability efforts (at sites where LNAPL is present) might be required if any of the following are true:

- Plume Stability Analysis Fails To Demonstrate A Stable/Decreasing Plume
- Active Remediation Is Needed
- The Evaluator Has Not Demonstrated That All 5 Objectives of MRBCA Section 6.8 are met

1. **No Explosive Conditions**
2. **Extent of Free Product Is Delineated**
3. **LNAPL/Free Product Is Not Producing Unacceptable Risk To Human/Ecological Exposure Pathways**
4. **Free Product Is Not Migrating**
5. **Free Product Shall Be Removed To The Maximum Extent Practicable**

For Assistance in Demonstrating Delineation And Lack Of Migration of LNAPL/Free Product
Refer to Section 6.8.2 of the
2013 MRBCA Guidance Document.

"The occurrence of free product petroleum must be documented and investigations must be conducted to determine the extent of the free product and whether and to what extent it is migrating. This determination will require the installation of a number of borings and monitoring wells sufficient to fully define the free product and periodic measurements of free product in these wells. ***The resulting data must be sufficient to demonstrate spatial and temporal trends in free product thickness. Note that free product thickness is critically affected by water table fluctuations.***"

Section 6.82 Of The 2004-2005 MRBCA Guidance Document States Similar Requirements.

SHAMLESS PLUG SLIDE

That will remain up during the question/answer period

SUNBELT
ENVIRONMENTAL SERVICES, INC.

www.sunbeltenv.com

www.facebook.com/SunbeltEnvironmentalServices/

dpowell@sunbeltenv.com