March 2020

Making Better Use of Short-Term Wetland Hydrology Data:

Correlation with Nearby Long-Term Daily Data and Probability-Duration Analysis

Purpose: To explain and provide a step-by-step approach to a useful technique from the NRCS publication Hydrology Tools for Wetland Identification and Analysis called **probability-duration analysis**, which can greatly increase the value of short-term wetland hydrology monitoring data if longer-term data are available from nearby sites.

Audience: Wetland professionals interested in squeezing the most value out of short-term wetland hydrology data. Some familiarity with hydrology data and with spreadsheet formulas is helpful.

Intended use: In situations where only short-term monitoring data are available, the technique can be used to increase the value of those data where they correlate with nearby longer-term data. **The technique is NOT meant to be used to bypass monitoring protocols put in place through wetland banking and other programs**.

Introduction

In determining hydrology on a wetland site, we often find ourselves in the uncomfortable position of making decisions based on short-term hydrologic monitoring – perhaps only one growing season's worth or less. The data on their own do not provide a statistically significant basis for answering the kinds of questions that are typically asked about depth, duration, and frequency – questions such as "*Is the water table within 12 inches of the ground surface for 14 consecutive days in most years?*" We can increase the value of short-term data by interpreting them in the context of antecedent precipitation conditions, but even so, we find ourselves making educated guesses about depth, duration, and frequency.

However, an under-utilized tool exists that can be extremely useful for making better use of limited data. If a **nearby** location has **longer-term daily hydrologic monitoring data**, and if those data are **correlated** with our shorter-term site data, we can greatly increase the value of our limited measurements.

Probability-duration analysis is a standard technique from the publication Hydrology Tools for Wetland Identification and Analysis that can be used to increase the value of limited data from a site **if the following are true**:

- **Longer-term daily data** are available from a nearby monitoring site, such as stage or flow data from a stream gauge, lake or wetland water-level data, or observation well water-level data.
- Enough longer-term data are available from the nearby site – ideally there should be at least 10 years of daily data.
- The limited data from the site **correlate** with the long-term data.

If these conditions are met, then we can greatly increase the value of limited data by using the statistics from the longer-term data to better interpret the limited site data.

Example 1: The Basic Idea

The first example will illustrate the basics of correlating site data with nearby longer-term daily data. The second example will give details on how to do probability-duration analysis using spreadsheet methods.

Suppose we have only one growing season's water-level data from a shallow monitoring well on a wetland site (**Figure 1**). We want to know whether the location meets the technical criterion for wetland hydrology. We have dutifully downloaded and analyzed precipitation data from the State Climatology Office in order to try to interpret the data in the context of antecedent precipitation (**Figure 1**). (For more details on evaluating antecedent precipitation, see the guidance document Evaluating Antecedent Precipitation Conditions Using Climate Data Available in Minnesota on the Board of Water and Soil Resources website.) We suspect that the well's location does meet the technical criterion for wetland hydrology ("water table within 12 inches of the surface for 14 consecutive days in most years") as all the measured water levels from mid-April through mid-June are within 12 inches of ground surface. However, the "most years" part is problematic – we only have data over one growing season, and our precipitation analysis shows that April through mid-May were wetter than normal.

Fortunately, a stream gauge on a nearby ditch has 10 years of daily stage data, including the same year we collected data from our site. If the data from our site are correlated with the stream gauge data, we can then "borrow" the statistics from the stream gauge to better interpret our site data.

Figure 2 shows the measured water levels from the well on our site plotted against the daily average water level at the stream gauge for the same days as the well measurements. A respectable correlation emerges, which was estimated and sketched visually in **Figure 2**. A linear or logarithmic best-fit curve could also be be calculated with spreadsheet software. A good correlation does not necessarily represent a cause-and-effect relationship, but indicates that the water levels at both locations respond to local environmental conditions in similar ways that can be correlated. How good a correlation is "good enough"? If we have calculated the correlation curve, we can use a statistical measure of correlation such as $\mathbf{R^2}$. But the human eye is a good judge. For example, **Figure 3** shows an example of poorly correlated data – not good enough to proceed with this method.

Because our example (**Figure 2**) displays a correlation, we can use **probability-duration analysis** with the daily stream gauge data to estimate the stage that would be attained or exceeded for **14 consecutive days** in **most years** at the stream gauge – the **50% probability 14-day-duration stage**. Doing this analysis uses spreadsheet tools, which will be described in detail in the next example. For this first example we will just assume we have that value in hand. Using that value derived from the stream-gauge data, we plot the 50% probability 14-day-duration stage of **891.4** feet (the vertical black dashed line in **Figure 4**).

We see that for all days when the ditch water level was greater than the 50% probability 14-day-duration elevation (>891.4 feet), the water level in the well was within 12 inches of the surface. Because the correlation is good, we can reasonably infer that the durations of high water table measured in the well were the same as the ditch's water-level durations, so we conclude the well experiences a water table within 12 inches of the surface for 14 consecutive days with a greater than **50% annual probability** – that is in **most years**. We conclude that the well location **does** meet the technical criterion for wetland hydrology.

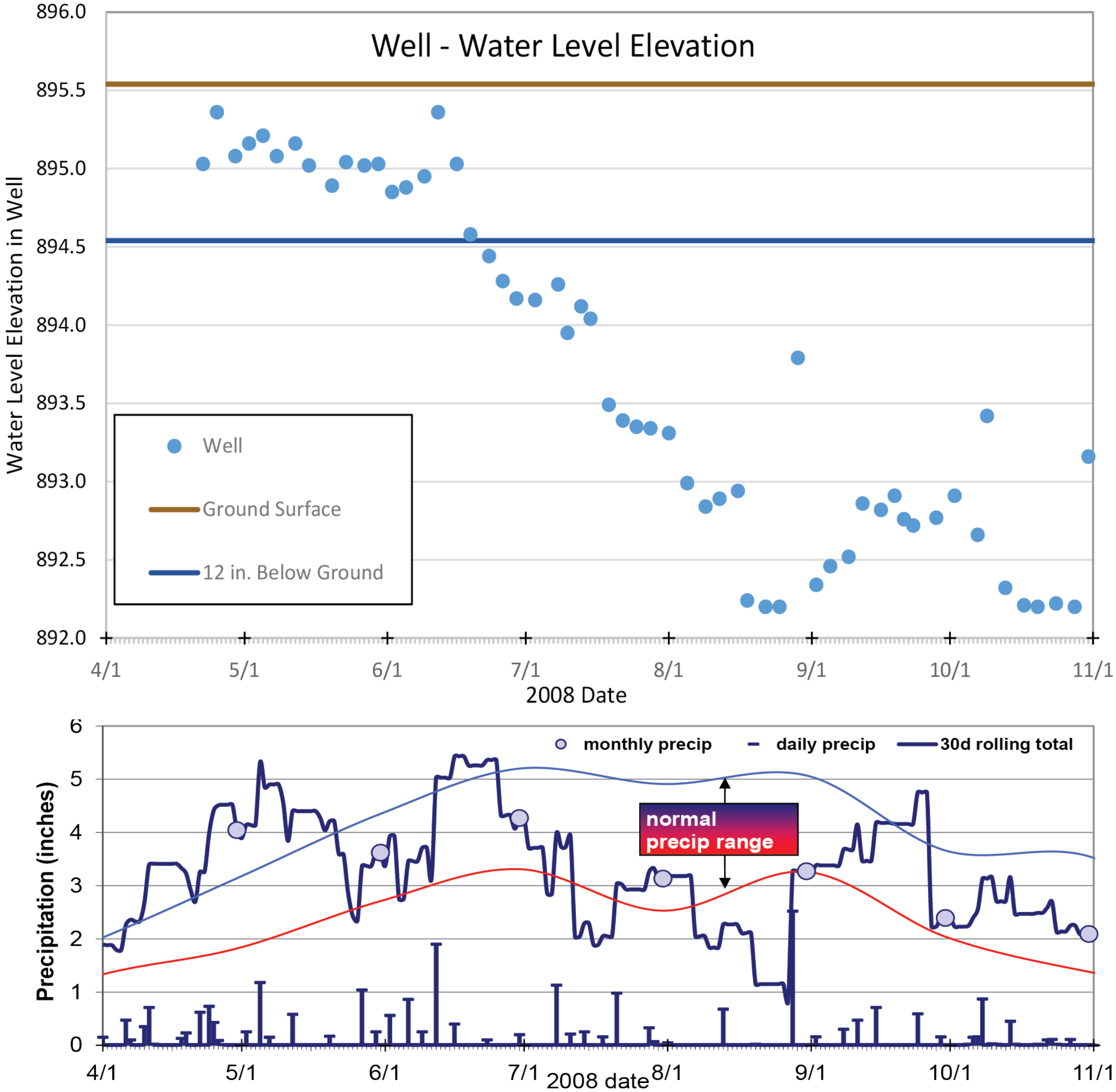


Figure 1

Above: *Water level data from a shallow monitoring well.* ***Below:*** *Daily and monthly precipitation, normal precipitation range, and 30-day rolling total. We suspect that the location of the well does meet the technical criterion for wetland hydrology as all the measured water levels from mid-April through mid-June are within 12 inches of ground surface. But we only have one growing season's worth of data, and our precipitation analysis shows April through mid-May being wetter than normal.*

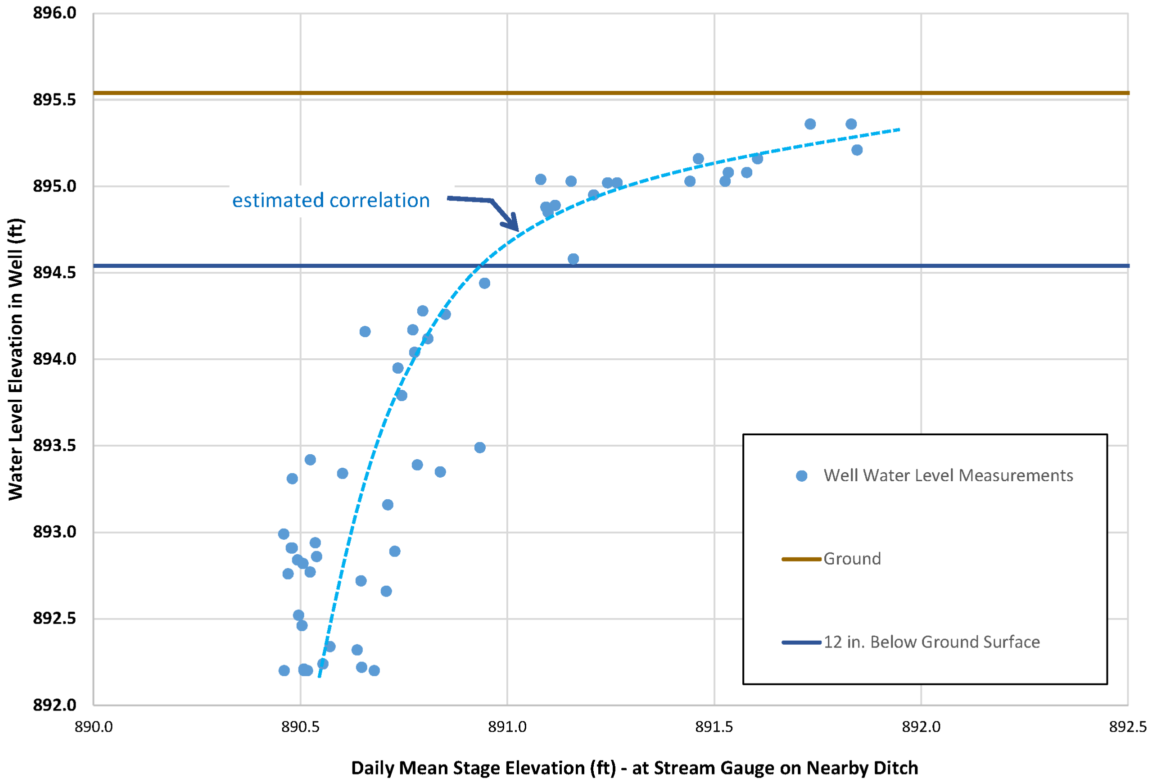


Figure 2
Water level elevations from the monitoring well plotted against the daily average water level at the stream gauge on the same days as the well measurements. In this case, the estimated correlation curve (light blue dashed line) was sketched visually, but often a linear or logarithmic equation can be fit to the data.

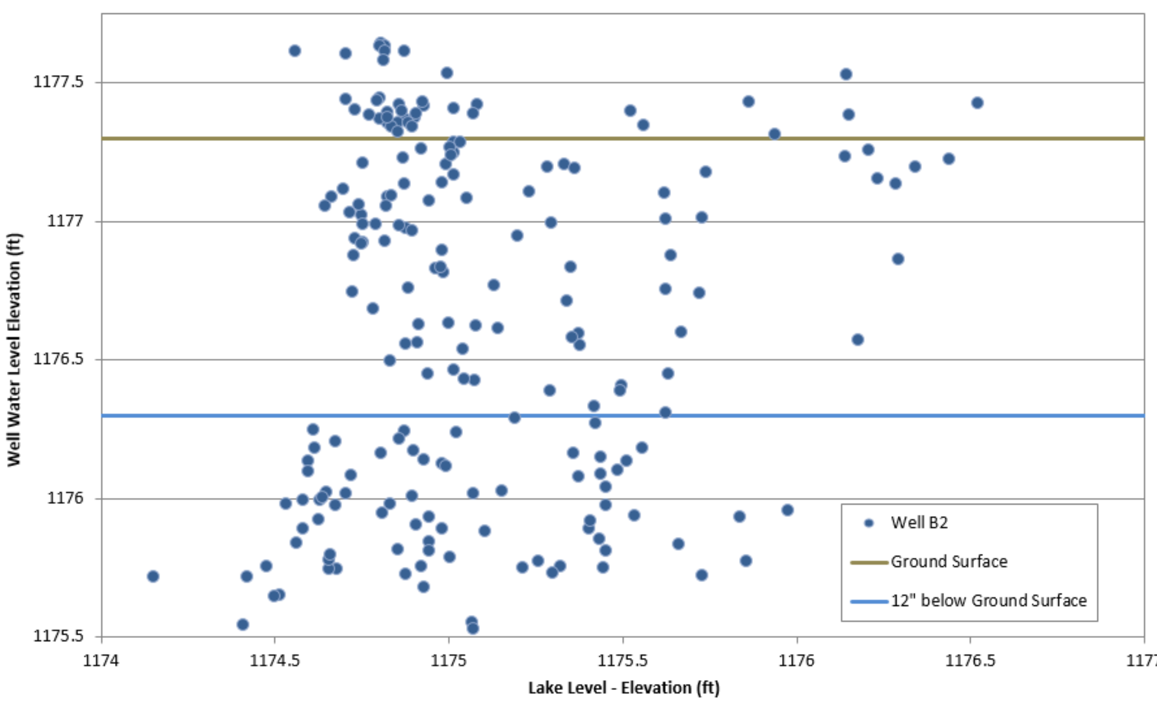


Figure 3
*Daily average water level elevations from a shallow monitoring well plotted against the daily average water level elevation at a nearby lake on the same days as the well measurements. In this case, the data here are **poorly, if at all, correlated**.*

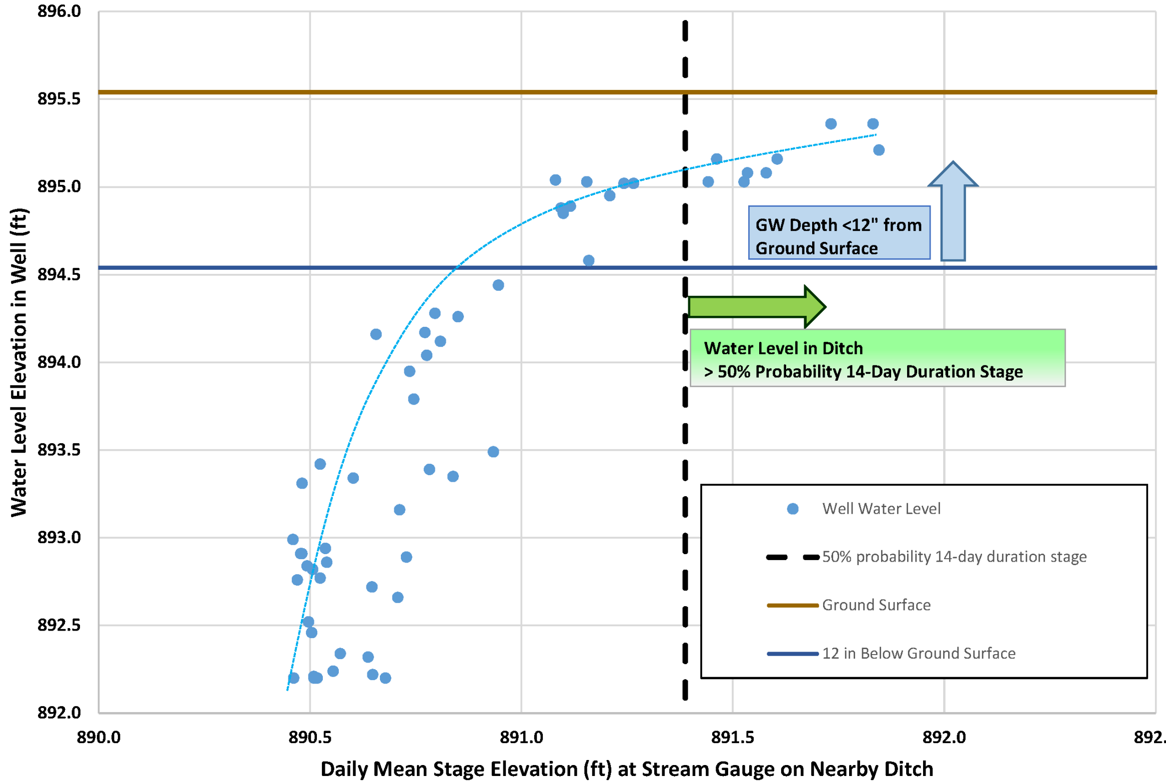


Figure 4

*Same graph as Figure 2 with the addition of the **50% probability 14-day-duration stage** of the ditch water level (vertical dashed line). All days when the ditch water level was greater than the 50% probability, 14-day-duration stage, the water level in the well was within 12 inches of the surface.*

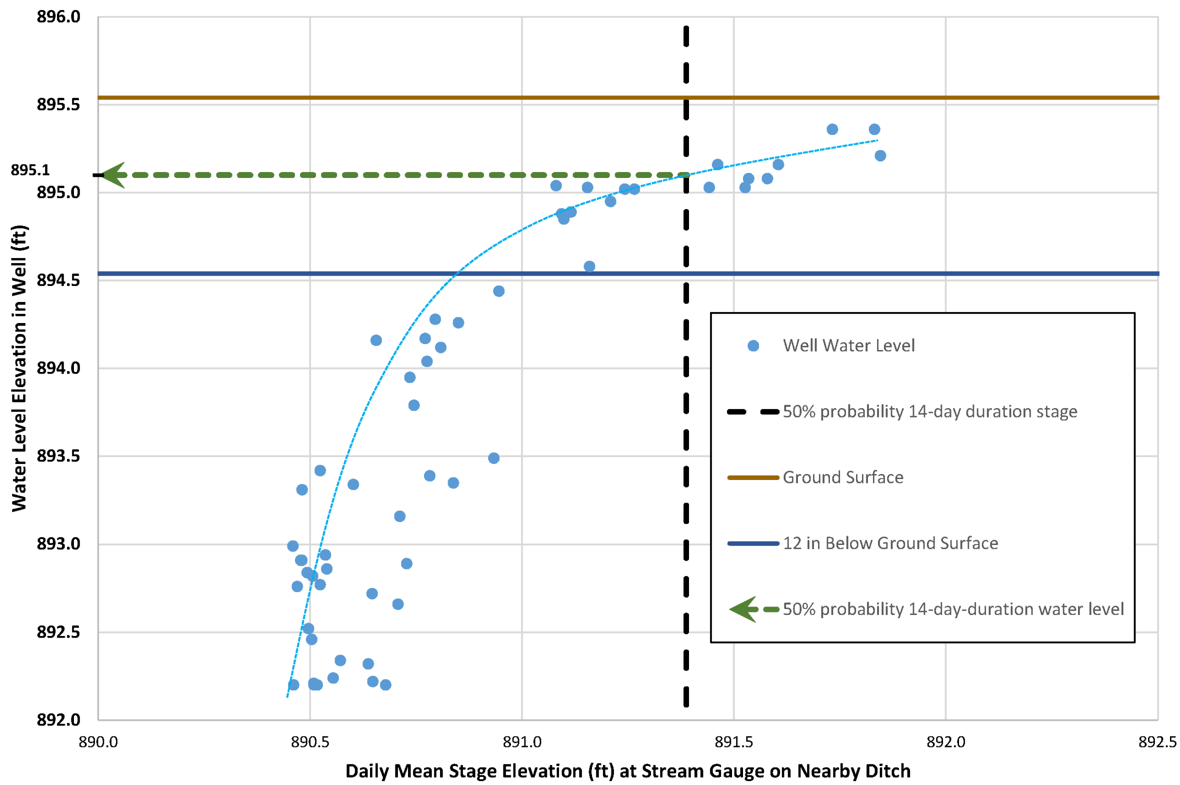


Figure 5
Using the correlation to estimate the 50% probability 14-day-duration water-level elevation ***at the well location****: The value where the estimated correlation curve crosses the 50% probability 14-day-duration stage line estimates the 50% probability 14-day-duration water level elevation in the well (y-axis) – in this case* ***895.1 feet****. Adding 1 foot to this value yields* ***896.1*** *feet – an elevation that can be mapped on the land surface as an estimate of the wetland hydrology boundary. In other words, in the vicinity of this well, elevations lower than* ***896.1*** *feet are estimated to meet the technical criterion for wetland hydrology.*

We can also use the correlation to estimate the 50% probability 14-day-duration water level **in the monitoring well**. It is the well water level elevation (y-axis) value of the point where the estimated correlation curve crosses the 50% probability 14-day duration stage line (**Figure 5**). This value is an estimate of the water level elevation that is reached for 14 consecutive days in most years at the well location. Adding 1 foot to this elevation yields an elevation that can be mapped on the land surface as an estimate of the wetland hydrologic boundary – at least in the vicinity of the well. In other words, in the vicinity of this well, elevations lower than **896.**1 feet (895.1 feet + 1 foot) are estimated to meet the technical criterion for wetland hydrology. Having such a line on the map can be a very useful starting point for wetland delineation.

Example 2: Probability-Duration Analysis

Probability-duration analysis is used to estimate probabilities, depth, duration, and frequency of inundation or saturation. Probability-duration refers to the probability that a condition exists for a specified duration. A detailed description of this technique can be found in the new (2015) version of Hydrology Tools for Wetland Identification and Analysis. The analysis is a straightforward exercise using spreadsheet software.

We want **daily** data from a longer-term station for doing probability-duration analysis because we typically want to determine its 50% annual probability (i.e. most years) **14-day-duration** condition. The data from a nearby monitoring site can be stage or discharge data from a stream gauge, lake or wetland water level data, or observation well water level data – any type of hydrologic data that could possibly correlate with shorter-term data from a site.

In the second example (**Figure 6**), we consider some short-term (one growing season's) data from two monitoring wells on a potential wetland restoration site. We are interested in knowing if the locations meet the technical criterion for wetland hydrology. The wells lie in a river floodplain, and a stream gauge downstream of the sites provides many years of discharge data. The well data (**Figure 7**) suggest that wetland hydrology may exist at the wells' locations, especially #6. However, during the monitored year, the early growing season was particularly wet, so there are disagreements about whether wetland hydrology exists at the locations. To settle these questions, we can use the long-term data from the stream gauge to better interpret and make use of the short-term monitoring well data.

Figure 6
Example 2: Monitoring wells on a potential wetland restoration site. We want to know if the locations meet the criterion for wetland hydrology. The wells have only one growing season's data, but a gauge downstream on the river has many years of discharge data.

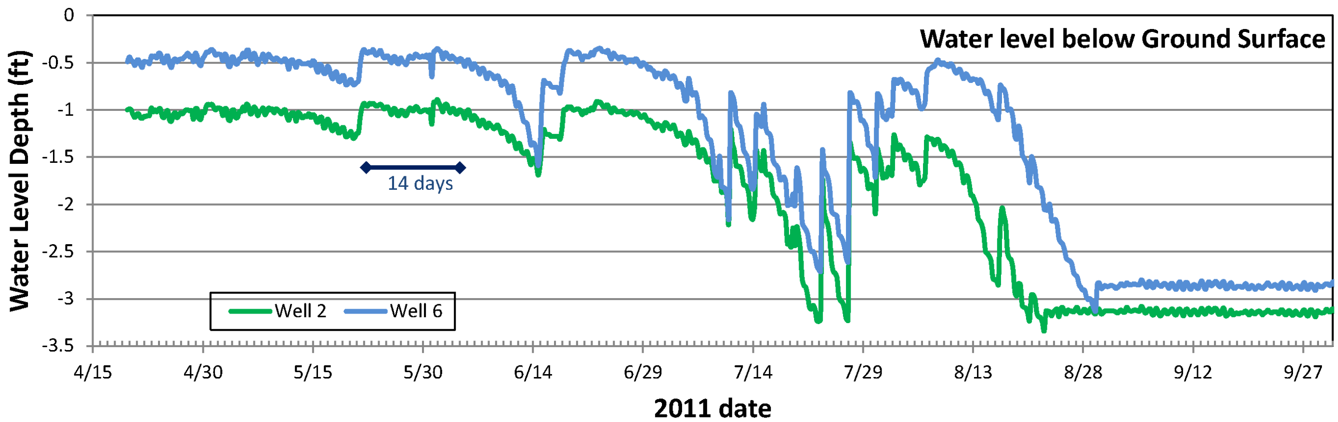


Figure 7: Water-level data from two monitoring wells in Example 2 for the 2011 growing season.

Dealing with Raw Data

We have discharge data from the stream gauge going back many decades – at least daily but in recent years collected on 15-minute intervals. We need an average (mean) daily discharge to correlate with an average daily well-water level. We could painstakingly go through and calculate daily averages for the stream gauge and well data, but far easier is using a spreadsheet tool known as a **Pivot Table**. A pivot table can summarize data records by one of the variables, such as the "Day" in our example (**Figure 8**). In this case, we use a pivot table in Microsoft Excel to produce a column that calculates average daily discharge. Excel has help screens to coach you through using pivot tables.

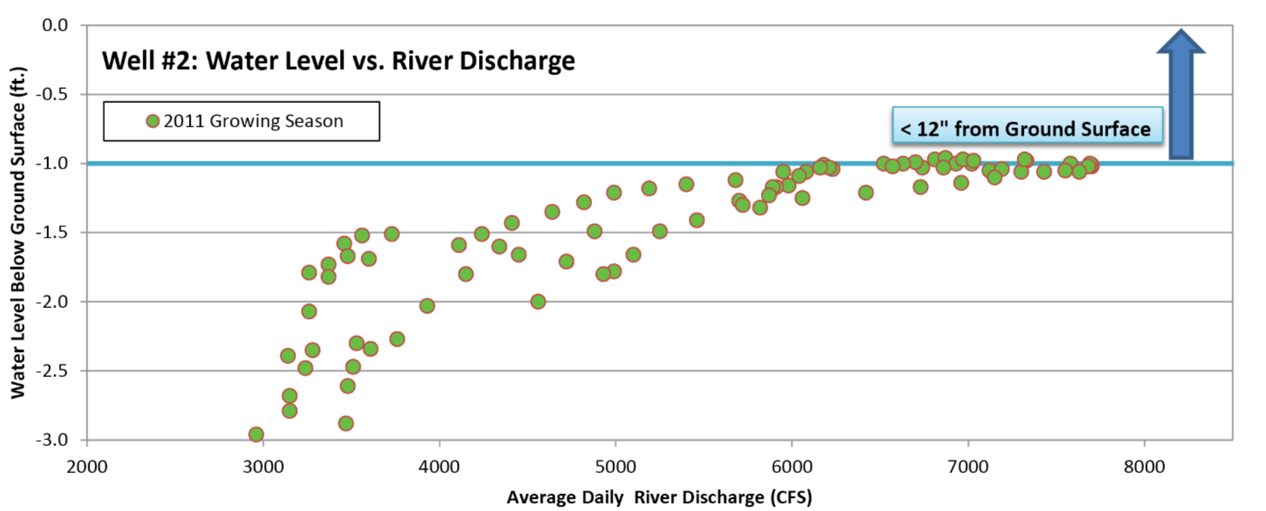
In a similar approach, a pivot table is also used to calculate average daily water levels in the monitoring wells. Once that calculation is complete, we can construct a table with rows being dates and columns being the daily averages for the stream gauge and the two wells (**Figure 9**). From these data, we then plot average daily river discharge against average daily well-water levels to see if there is a correlation (**Figure 10**).

	A	B	C	D	E
1				Pivot Table	
2	Timestamp	Day	Discharge (cfs)	Day	Average of Discharge (cfs)
11187	8/20/2014 5:15	8/20/2014	2850	8/11/2014	2690
11188	8/20/2014 5:30	8/20/2014	2860	8/12/2014	2690
11189	8/20/2014 5:45	8/20/2014	2850	8/13/2014	2710
11190	8/20/2014 6:00	8/20/2014	2860	8/14/2014	2800
11191	8/20/2014 6:15	8/20/2014	2860	8/15/2014	2910
11192	8/20/2014 6:30	8/20/2014	2850	8/16/2014	2930
11193	8/20/2014 6:45	8/20/2014	2860	8/17/2014	2850
11194	8/20/2014 7:00	8/20/2014	2860	8/18/2014	2770
11195	8/20/2014 7:15	8/20/2014	2870	8/19/2014	2808
11196	8/20/2014 7:30	8/20/2014	2860	8/20/2014	2852
11197	...	...	...	...	...

*Figure 8: Example of part of a **pivot table** used to calculate average daily discharge from discharge data collected at a variety of time intervals.*

	A	B	F	J
1		Daily Ave. River Discharge	Daily Ave Water Level Depth (ft)	
2	Date	(CFS)	Well 2	Well 6
30	5/2/11	7020	-1	-0.43
31	5/3/11	7120	-1.05	-0.48
32	5/4/11	7190	-1.04	-0.48
33	5/5/11	7330	-0.98	-0.41
34	5/6/11	7580	-1	-0.42
35	5/7/11	7700	-1.02	-0.47
36	5/8/11	7700	-1.01	-0.44
37	5/9/11	7690	-1	-0.45
38	...	...	...	...

Figure 9: Example of a table with daily averages for the river discharge and well water levels for each date.



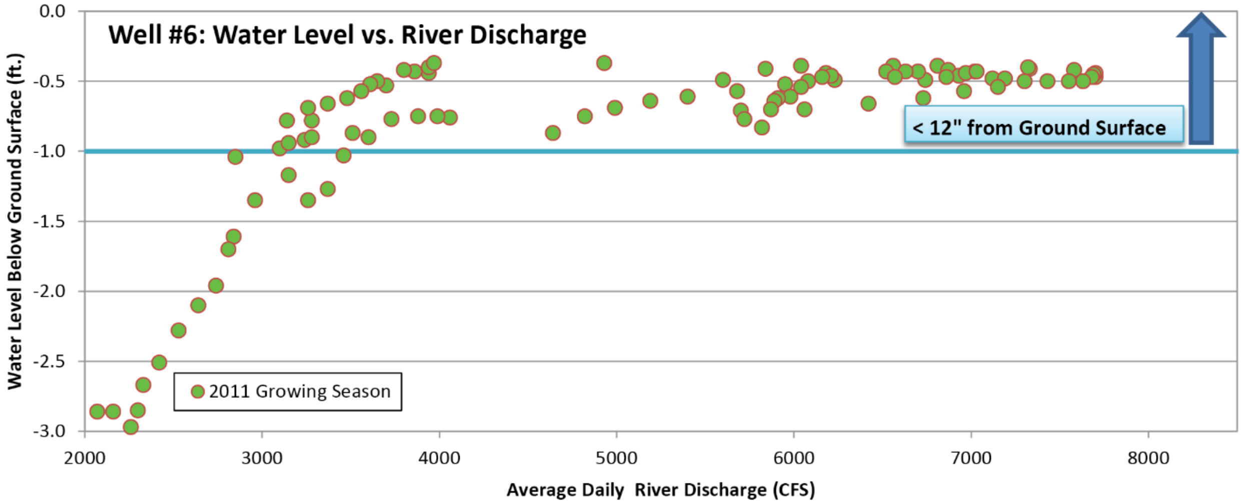
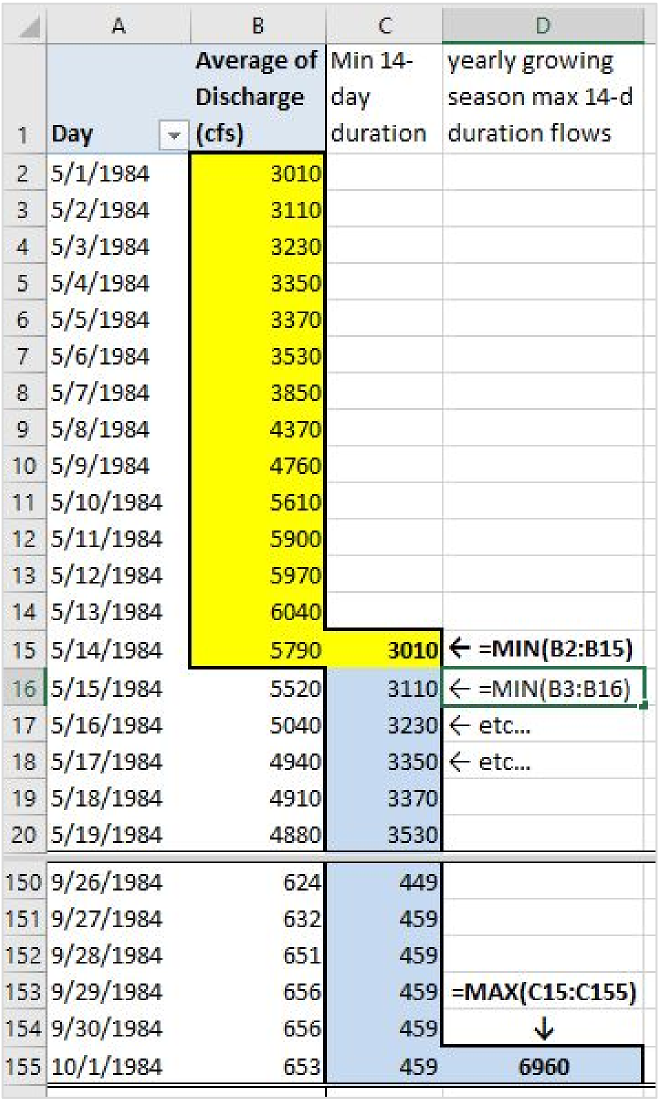


Figure 10: *Graphs of water levels in wells #2 and #6 vs. average daily river discharge in cubic feet per second (cfs).*

Probability-Duration Analysis Using Spreadsheet Methods

We see a respectable correlation between the average daily river discharge and the average daily water table depth as measured by the monitoring wells (**Figure 10**). Now we are ready to use probability-duration analysis with the average daily river discharge data to estimate the 50% annual probability 14-day-duration river discharge. We have obtained daily discharge data from the stream gauge from 1984 to 2014 and will assume a growing season of May 1 through October 1 for this location. **We are only interested in data during the growing season**.

We organize the data into spreadsheet columns; an example portion of an Excel spreadsheet is shown in **Figure 11**. Dates are in Column A, followed by the average daily river discharge value in Column B. Column C is used to calculate the lowest flow exceeded in the 14-day period preceding and including each successive date, so the first figure appears at the 14th day of record, and the cell formula is **=MIN(B2:B15)**. The formula for the next cell is **=MIN(B3:B16)**, and so on. The cell formula will automatically increment when copying the first cell's formula downward to the bottom of the spreadsheet. Column D is used to calculate the **maximum** 14-day-duration discharge for each growing season. The formulas in Column D are placed the end of the growing season data for each year of record – in our example October 1. The formula in Column D for the first year is **=MAX (C15:C155)**, which yields the maximum 14-day-duration flow for the 1984 growing season. This formula is then copied to the end of each growing season's data for all the years we are analyzing. The cell formula will automatically increment to calculate the **maximum 14-day-duration discharge for each growing season**.

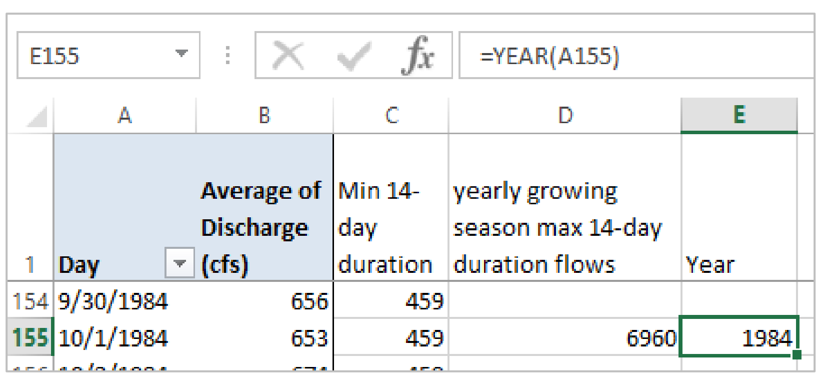


	A	B	C	D
1	Day	Average of Discharge (cfs)	Min 14-day duration	yearly growing season max 14-d duration flows
2	5/1/1984	3010		
3	5/2/1984	3110		
4	5/3/1984	3230		
5	5/4/1984	3350		
6	5/5/1984	3370		
7	5/6/1984	3530		
8	5/7/1984	3850		
9	5/8/1984	4370		
10	5/9/1984	4760		
11	5/10/1984	5610		
12	5/11/1984	5900		
13	5/12/1984	5970		
14	5/13/1984	6040		
15	5/14/1984	5790	**3010**	← **=MIN(B2:B15)**
16	5/15/1984	5520	3110	← =MIN(B3:B16)
17	5/16/1984	5040	3230	← etc…
18	5/17/1984	4940	3350	← etc…
19	5/18/1984	4910	3370	
20	5/19/1984	4880	3530	
150	9/26/1984	624	449	
151	9/27/1984	632	459	
152	9/28/1984	651	459	
153	9/29/1984	656	459	**=MAX(C15:C155)**
154	9/30/1984	656	459	↓
155	10/1/1984	653	459	**6960**

Figure 11: *Example portion of a spreadsheet for probability-duration analysis.*

Next, the goal is to transfer these annual maximum values to a new column with no intervening blank cells. The procedure below is an efficient way to accomplish this.

Here is a handy shortcut to avoid having to manually transfer each maximum value: Have the year of the maximum in the cell adjacent to the maximum value (Column E in **Figure 12**). This can be entered manually, or better yet, calculated from the date in Column A using the **=YEAR( )** formula (e.g. **=YEAR(A155)**). Then copy and paste Columns D and E to a new location, using "**Paste Special**" and selecting "**Values**", so only the values are copied and not the formulas. Finally, sort the new columns by the year variable. (In Excel, select the columns, then on the [Data] tab, select 'Sort', click 'My Data has headers', then select Sort by: Year.) The result looks like Columns J and K in **Figure 13**.



E155 =YEAR(A155)

	A	B	C	D	E
1	Day	Average of Discharge (cfs)	Min 14-day duration	yearly growing season max 14-day duration flows	Year
154	9/30/1984	656	459		
155	10/1/1984	653	459	6960	1984

Figure 12: *The formula =YEAR(A155) in Column E calculates the year from the date in Column A. This is useful for sorting the maximum values into a new column.*

J	K	L	M	N
Year	yearly growing season max 14 day duration flows	50% probability	Rank	% probability of exceedence (Weibull Plot) rank/(n+1)
2014	8770		6	18.8%
2013	7940		11	34.4%
2012	13300		1	3.1%
2011	7020		13	40.6%
2010	2070		31	96.9%
2009	6440		17	53.1%
2008	8050		10	31.3%
2007	3040		27.5	85.9%
2006	5320		18	56.3%
2005	6560		16	50.0%
2004	3040		27.5	85.9%
2003	2810		29	90.6%
2002	5050		21	65.6%
2001	12700		2	6.3%
2000	3260		26	81.3%
1999	8380		8	25.0%
1998	4690		22	68.8%
1997	9430		4	12.5%
1996	11000		3	9.4%
1995	4380		23	71.9%
1994	6980		14	43.8%
1993	8590		7	21.9%
1992	3630		25	78.1%
1991	5270		20	62.5%
1990	5300		19	59.4%
1989	8340		9	28.1%
1988	2160		30	93.8%
1987	4130		24	75.0%
1986	9380		5	15.6%
1985	7590		12	37.5%
1984	6960	6560	15	46.9%
=PERCENTILE.EXC(K2:K32,0.5) ↗				

***Figure 13:** The annual maximum 14-day duration discharges are in Column K. We use the spreadsheet formula to calculate the 50th percentile of this data set, yielding the value **6560 cfs in** Column L.*

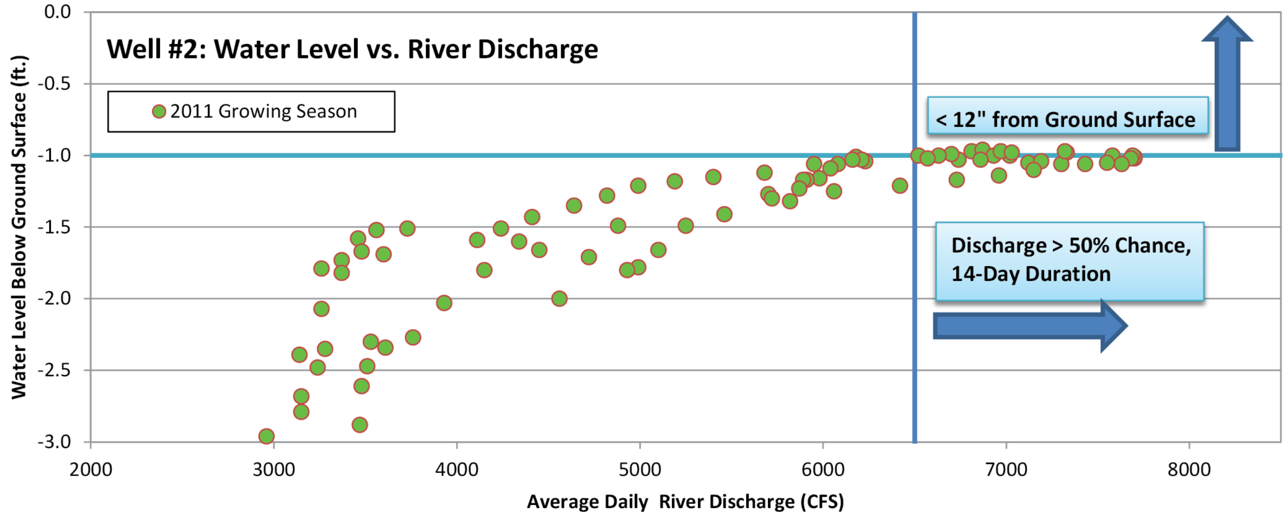
We now calculate the **50th percentile** of the annual 14-day-duration maxima from Column K (**Figure 13**). This is the annual 14-day-duration discharge that has a 50% probability of being reached or exceeded. In other words, this is the discharge that will occur or be exceeded for 14 consecutive days in most years. The Excel spreadsheet formula that does this is **=PERCENTILE.EXC(K2:K32,0.5)**. For our data set, this comes out to be **6560 cubic feet per second (cfs) in Column L**.

(Columns M and N are not necessary, but we will discuss these later as they provide a way to visualize the data.)

Finally, we plot the 50% probability 14-day-duration discharge (from Column L) as a vertical line on the graph of river discharge vs. well water levels (blue vertical line in **Figure 14**). We can now interpret the well data as we did in the first example.

For **Well #6**, we see that for all days when the river discharge was greater than the 50% probability 14-day-duration discharge (>6560 cfs), the water level in the well was within 12 inches from ground surface. Since the correlation is good, we can reasonably infer that the durations of high water table measured at that well were the same as the river-discharge durations, so we conclude that well experiences water table within 12 inches of the surface for 14 consecutive days with a greater than **50% annual probability** – that is in **most years**. We then conclude that the well location **does** meet the technical criterion for wetland hydrology.

For **Well #2**, we see that for all days when the river discharge was greater than the 50% probability 14-day-duration discharge, the water level in the well was at or slightly below 12 inches from ground surface. Because the correlation is also good here, we can also infer that the durations measured at the well were the same as the river discharge durations. In this case we conclude the well location, though right on the edge, **does not** quite meet the technical criterion for wetland hydrology.



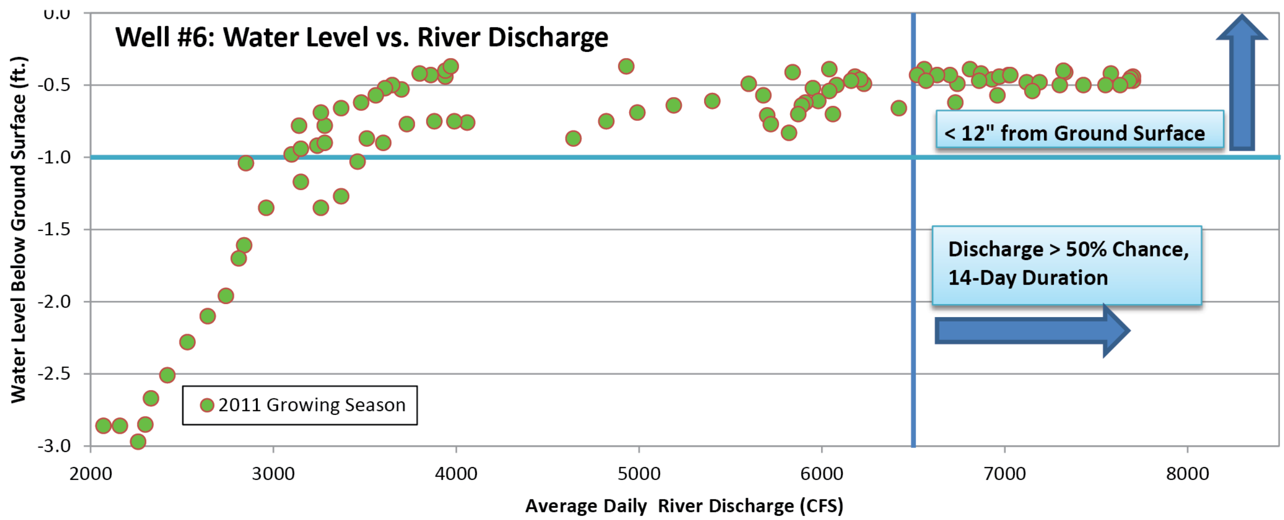


*Figure 14: Graphs of water levels in wells #2 and #6 vs. average daily river discharge, now with the added 50% probability 14-day duration river discharge of **6560 cfs** (vertical line).*

Visualizing the Analysis (Extra)

The next discussion is extra – it is not necessary for the analysis but is a useful way to visualize what the probability-duration analysis is doing.

Column M (**Figure 13**) assigns **ranks** to the annual maximum 14-day-duration discharges in order of highest to lowest (rank "1" being highest). Determining the rank in Column M uses the Excel formula: **=RANK.AVG(K2,K$2:K$32)** for the 2014 row. As with other formulas, it automatically increments when copied to the cells below. Next, we calculate the **percent probability of exceedance** of the discharge values in **Column N**. This percent probability is the chance that this discharge will occur or be exceeded for 14 consecutive days in a year. The formula that calculates the percent probability is $\boldsymbol{Rank}/(\boldsymbol{n}+\mathbf{1})$ where $\boldsymbol{n}$ is the number of years of data. The 50th percentile of this data set is the same as its median.

The exceedance probabilities for different 14-day-duration discharges can be visualized with a "Weibull Plot" (**Figure 15**). In this plot, the annual maximum 14-day-duration discharges in Column K (**Figure 13**) are plotted against their **percent probability of exceedance** (Column N). The plot is a way of visualizing the probabilities that a given 14-day-duration discharge will be reached or exceeded in a year.

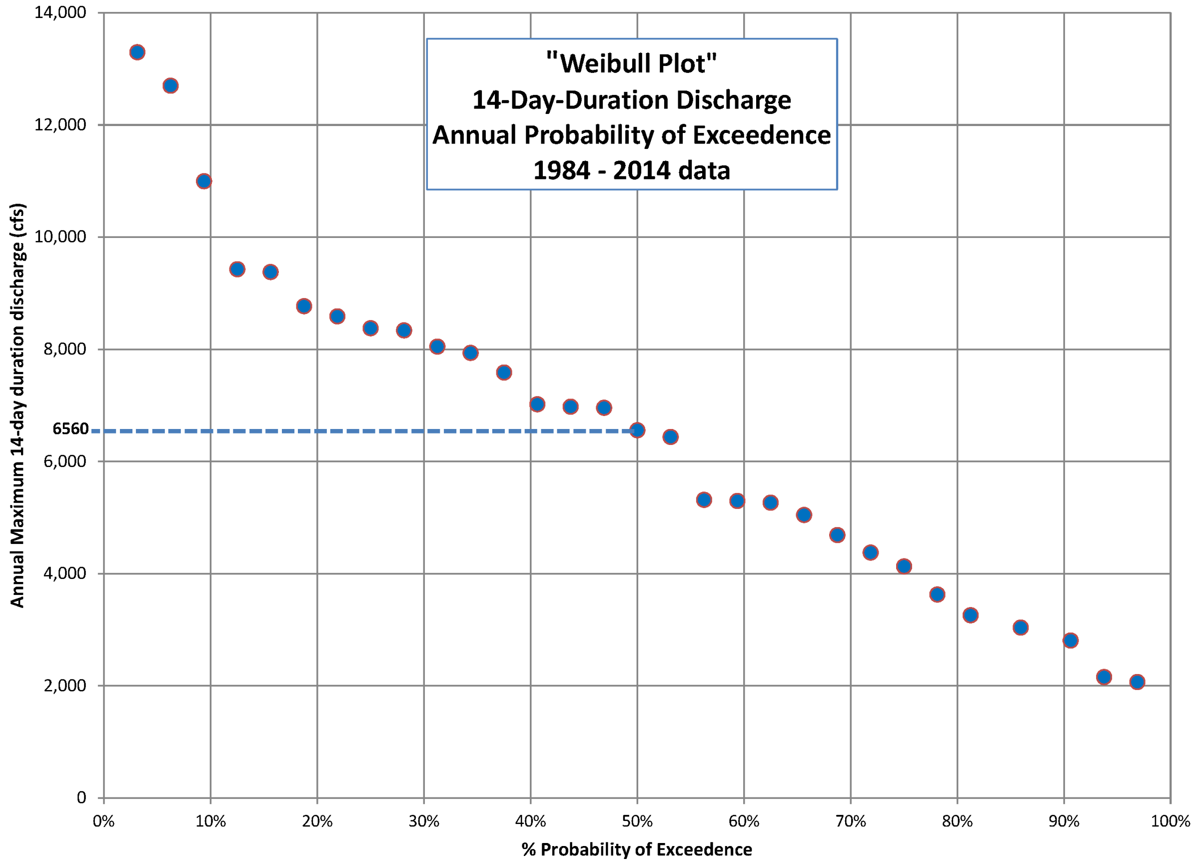


__Figure 15:__ Plot of the annual probability of exceedance of 14-day-duration discharges (Weibull Plot"), based on 1984-2014 data. It shows the probabilities that a discharge will be met or exceeded for 14 days in a year. The number we are interested in is the __50% annual probability__ – i.e. the discharge that will be reached or exceeded for 14 consecutive days in __most years__.

We are most interested in the **50% probability** – that is the discharge that will be met or exceeded for 14 consecutive days in **most years**. The plot illustrates that the **50% probability 14-day-duration discharge** is 6560 cfs, the same value that was calculated by the spreadsheet formula **=PERCENTILE.EXC(K2:K32,0.5)** in column L (**Figure 13**). The 50% probability value is the same as the **median** of the data set. Again, although plotting the data this way is optional, it is a useful way to visualize the probability relationship.

Summary

In summary, here are the steps to make better use of short-term hydrologic monitoring data from a site using their correlation with longer-term data and probability-duration analysis:

1. Find out if longer-term daily hydrologic monitoring data exist from one or more nearby locations. Ideally, there should be at least 10 years of data.
2. Calculate the average (mean) daily values of both the longer-term data and the short-term site data. Using a spreadsheet pivot table makes this calculation easy. Make a table with rows for each date and columns being the average daily values from the site and the longer-term data set (**Figure 9**).
3. Using this table, plot the average daily values from the site against those from the longer-term data (**Figure 2**, **Figure 10**). Spreadsheets also make this step easier to do.
4. Examine the plot to see if the data appear to correlate. If they are scattered all over the place, **STOP HERE** - it won't work. If they appear to be correlated, continue to probability-duration analysis.
5. Generate a correlation curve, either manually or using a spreadsheet tool.
6. For the longer-term daily data:
   a. For each date, using a spreadsheet formula, calculate the minimum value in the 14-day period preceding and including each successive date.
   b. Then use the spreadsheet to calculate the maximum of these minimum values for each growing season (**Figure 11**). These maximum values are the maximum 14-day-duration value for each growing season.
   c. Calculate the 50th percentile (median) of the annual 14-day-duration maximum values (**Figure 13**).
7. See where this median value (vertical line) falls on the plot you made in step 3 (**Figure 4**, **Figure 14**).
8. Look at the values from the site (short-term) corresponding to values from the longer-term data that are greater than the 50th percentile 14-day-duration value. If all or most of these values are greater than the threshold (usually 12 inches below ground surface), then one can conclude that the location meets the criterion for wetland hydrology. Repeating these steps using long-term data from more than one source (i.e. more than one stream gauge, lake, or observation well in the vicinity) will further test the conclusion and make the conclusion more robust.

References

USDA Natural Resources Conservation Service, 1975. Part 650 Engineering Field Handbook, National Engineering Handbook, Chapter 19: Hydrology Tools for Wetland Identification and Analysis (210–VI–NEH, Amend. 75, September 2015).

Evaluating Antecedent Precipitation Conditions Using Climate Data Available in Minnesota, Board of Water and Soil Resources Guidance Document, 2015