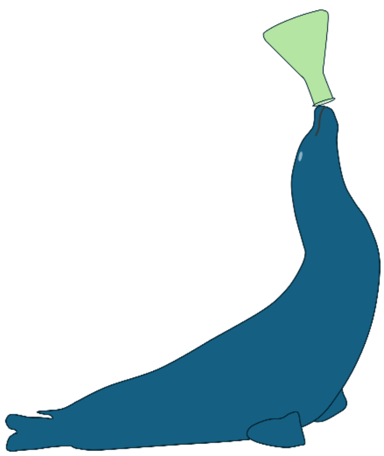
The “100-Year Flood”, a Skeptical Inquiry in ~~3~~ ~~4~~ 5 Parts.

Shawn D. Mahaney

December, 2024

scienceisjunk.com



Part One – Are the 100-Year Flood Maps We Use Any Good?

Anyone who has tried to buy a house has probably seen a form somewhere in the closing packet certifying that the structure does not lie in a "flood plain". Or if any part of the house is in the flood plain there will be extra paperwork (and a big check) for flood insurance. Someone at the insurance company has to work out a competitive premium to charge for the insurance. Someone else already worked out the map that defined the flood plain. How did they do it, and how sure are they? If they're wrong many more houses than expected get destroyed and insurance companies go bankrupt, or insurance rates are sky high and whole towns get abandoned.

History suggests that Mother Nature takes some delight in proving people wrong in these projections.

Challenges

Official flood plain maps usually are attempts to plot the limits of a "100-year flood", or an event with a 1% chance of happening next year. For other purposes one might want to know about a less- or more-rare event, e.g. the 1000-year event for "worst case" planning around something of critical importance. To estimate these probabilities, one might lean on historical weather data.

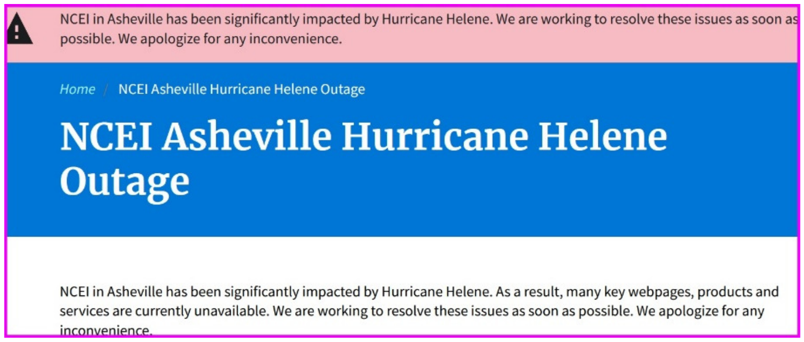
This has a number of problems,
- Rare events don't show up in data very often, by definition,
- Extrapolations beyond history are hyper-sensitive to sparsity and noise in the data,
- The available data is (much) worse than you probably expect,
- The assumptions made in statistically modeling weather are wrong,
- Translating weather data into real effects, i.e. local flood level, is messy,
- People easily forget all the above when citing statistical predictions.

Rare and extreme events by nature are difficult to handle statistically. They sit largely outside limited historical data. Functions fitted to that data are burdened by a large number of mild events and can be highly sensitive to the values of the few semi-extreme observations. Attempts to build math- and/or physics-based models of weather at the extremes will similarly run into hyper-sensitivities to small changes in debatable parameters.

The only thing that history shows us for sure is one example of what was "not impossible" at a previous point in time. To get anything else from that bit of information requires some assumptions, which in the context of other information and knowledge may be reasonable. To make claims based on a series of events over a span of time usually means making additional assumptions, most of which are convenient but untrue (more on this in later parts).

We're going to try some simple modeling techniques to demonstrate the challenges. We're going to get some real data, then we'll make up some big batches of fake data to mine.

The first job is to get the real data. This project is partly spurred on by hurricane Helene charging up into the Appalachians, wiping out some towns with raging flood waters and putting large parts of Asheville, North Carolina, under ten feet of water. At the start of the project naturally it was desired to get a bunch of historical weather data. But the NCEI (National Centers for Environmental Information) and its GHCNd (Global Historical Climatology Network daily) database are housed in Asheville…


NCEI in Asheville has been significantly impacted by Hurricane Helene. We are working to resolve these issues as soon as possible. We apologize for any inconvenience.

Home / NCEI Asheville Hurricane Helene Outage

NCEI Asheville Hurricane Helene Outage

NCEI in Asheville has been significantly impacted by Hurricane Helene. As a result, many key webpages, products and services are currently unavailable. We are working to resolve these issues as soon as possible. We apologize for any inconvenience.

This was not a good start to a project which took this author much more time than expected (another lesson in prediction error!). The first analyses here used data from near Clemson, SC, which was already in hand.

Rain Amount Is Not Flood Level

Weather data, like precipitation, is somewhat easy to get, but it's not exactly what we want for flood prediction. The effects of total rain fall are greatly dependent on local geography. When a river crests, it's getting flow from everywhere up stream, so we really want to know how much water fell into the entire upstream basin.

Stream level history may be better for direct estimation of flood events, but rainfall is directly related, and such data is more widely and deeply available. We're going to focus on precipitation data, using daily totals. Snow totals will be converted at 0.1 inch-per-inch, and we'll stick to temperate locations where snow does not often accumulate for long.

Flooding is very time-and-order sensitive. A long all-day soak followed by a sudden heavy burst is a great recipe for flash floods. The reverse order isn't as bad. And what if there is rain two days in a row? If many inches of water came spread out or in separate showers over the two days that's a low-impact event compared to three inches just before midnight then another four before one am.

For this analysis were going to sum pairs of adjacent daily rain totals. When there are many days in a row of rain, considered one long weather event, the days are paired, even or odd as counted in the event, such as to give the largest max for the event.

Clemson	Rain	count	Paired
2014-08-07	0	0	0
2014-08-08	0.9	1	0.9
2014-08-09	1.94	2	**3.26**
2014-08-10	1.32	3	*0*
2014-08-11	0	0	0

These max pairs are used here as a surrogate for true flood level data. Total 48-hour rain is presumed to be a useful marker for flood potential.

Toward the end of this article we will work on connecting flood levels and rainfall.

Lousy Data, But Lots of It

That 3.26-inch rain total in Clemson? That's more than either the 1.94 and 1.32 that make it up, but it's still badly wrong. I personally recall rain falling for 15 days in a row in that month. A contemporary news account relays that, "According to the National Weather Service, Seneca received 6.91 inches of rain between midday Saturday and midday Sunday.[1]" (Seneca is the actual locality of the Clemson-Oconee County airport.)

That mismatch is on a very modern bit of data. Going back farther through other data sets I've not compared archive data to contemporary reports too many times. A deep survey would be required to cross-validate the available mass of weather data.

The complete GHCNd database (as of the end of October, 2024) extracts to a 32-gigabyte archive file. The list of stations alone is almost 11 megabytes. This is a lot of data in total, but few of the streams are long and continuous and of high quality. In the United States alone 48,415 stations are listed. But about 7,000 of them have less than a year of data. The average range of dates covered is almost 22 years, and over 4000 of them span 75 years or more, but most of the longer sets are discontinuous. The longest data set with modern reports goes back to 1840, from Carlisle, Pennsylvania. But that station has no reports from the 1880s through the 2000s, and one wonders about the compatibility of data collected almost two centuries ago with the modern block.

The problem remains, how do we make projections from a limited pool of recent high-quality data?

First Predictions

We will start with a very simple way to extrapolate from a set of rainfall data directly into those 10/100/1000-year rain event numbers[2]. Data was in hand for the Clemson-Oconee County Regional Airport. This location is near where the author lives, and the data was used in a mathematics course at Clemson University. It's a relatively new airport, and even newer official weather station. 23 continuous calendar years of daily data are available for this station.

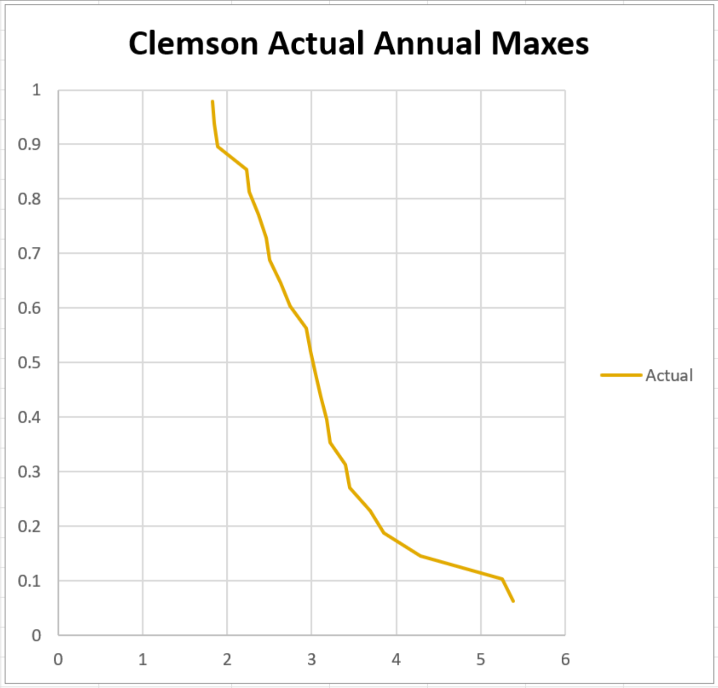
For this station the annual maximum (2-day sum) precipitation totals are found. Those maximums are put into a list and sorted by size. Then we begin making wild-but-useful assumptions!

If we assume the data are a perfectly representative sample of all possible maximums (very optimistic), and that the values are independent of each other (false), and that the probabilities don't change over time (false), then the chance of any new year maximum falling between any two of the 23 known values is about 1-in-24. Another half of a 24th of probability lives above and below the highest and lowest values. The probability values can be plotted against the rainfall amounts and we get a graph as follows. The form is a type of "waterfall chart".

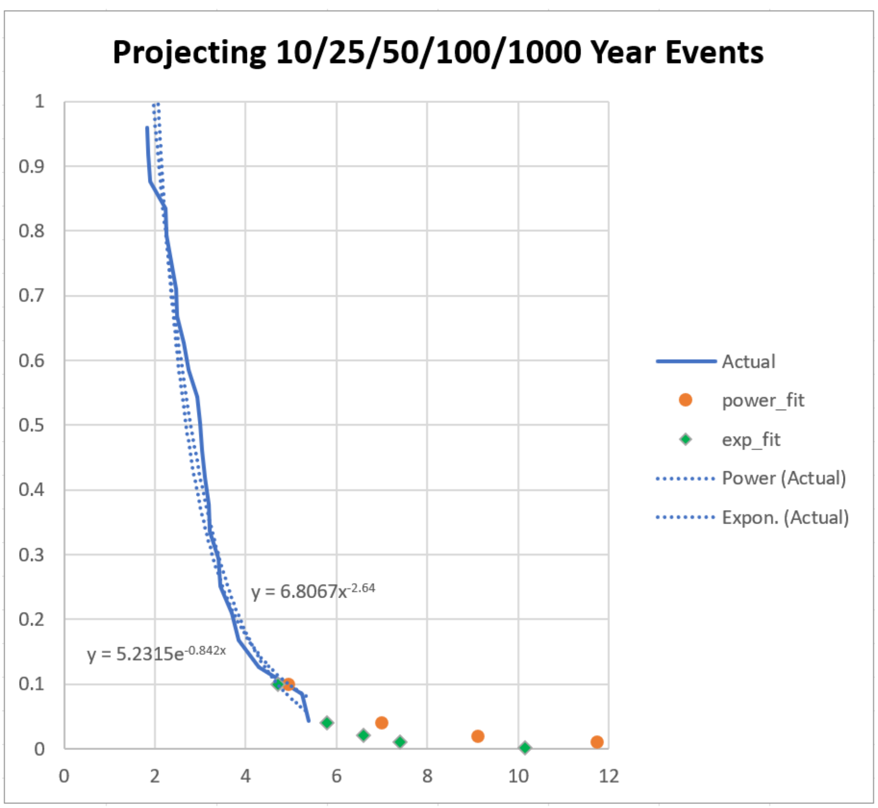
---

[1] Hardesty, Abe, (August 13, 2014), *Clemson Campus Slowly Recovering from Deluge*, Anderson Independent Mail
[2] Soltys, Mike, (September 14, 2013), "How to calculate the 100-year flood", https://www.mikesoltys.com/2013/09/14/how-to-calculate-the-100-year-flood/

annmax	prob exc
1.83	0.979
1.85	0.938
1.89	0.896
2.23	0.854
2.26	0.813
2.37	0.771
2.46	0.729
2.5	0.688
2.63	0.646
2.74	0.604
2.94	0.563
2.99	0.521
3.05	0.479
3.11	0.438
3.18	0.396
3.22	0.354
3.4	0.313
3.45	0.271
3.69	0.229
3.85	0.188
4.29	0.146
5.25	0.104
5.38	0.063



While based on a list of assumptions we know aren't true, these plots do usually, so far as I've seen, make a pretty reasonable curve. We can fit the right sort of simple function to them and get a visibly acceptable fit. Then that function can be used to extrapolate to more distant probabilities – those one-in-a-hundred and one-in-a-thousand rare events.



Here we have two different functions fit to the real data. One is a power-law format, the other is an exponential[3]. Both fits “look” good, and neither varies much from the other in the range of the actual data. But these sorts of ‘scalable’ functions are very dynamic at extremes, very sensitive to those last few semi-extreme data points.

Both give a 10-year (probability 0.1) maximum likely rain amount of just under five inches. But the 100-year values are almost twelve inches from the power law versus 7.4 inches out of the exponential. Trying to get a 1000-year event (from 23 years of data) the projections spread from 28 to 10 inches of rain.

This example demonstrates the difficulty in getting long-range projection from short-range data.

But what if we had a significantly longer data set? In the next part we will try to use the longest available data set in the US archive.

---

[3] What we specifically want to fit is a monotonic decreasing function with a zero asymptote, preferably which can be scaled to an integral area of one, but such is not necessary at this stage.

Part Two – Sampling the Longest Available Data Set

Looking for information about rare events, we naturally gravitate toward records with a long history. In this section we try to generate some statistics and projections about extreme rain events from a large sample of data. In the Global Historical Climate Network daily database, the largest record for the United States is that for Fort Bidwell, California

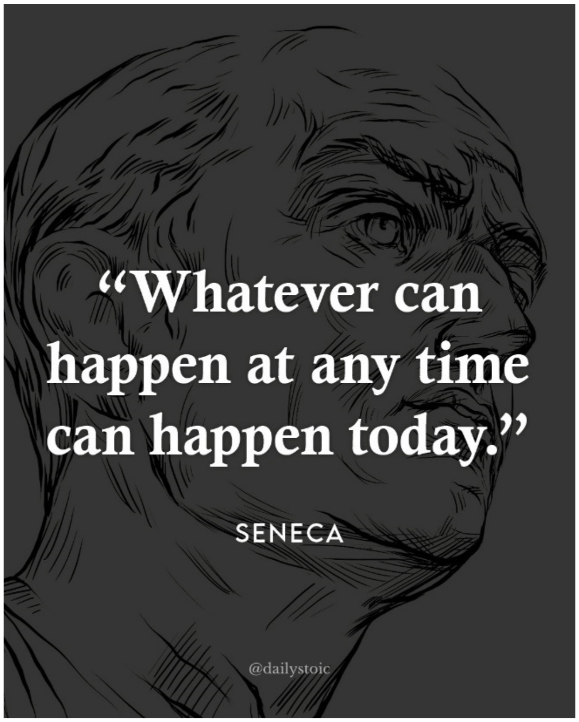
Holey Data

The data set on offer for this location promises data all the way back to June of 1867. Sticking to whole years we take 1868 through 2023. Great! For this 156-year data set we should get 56,979 days of observations. But there are 46768 observations available in the file. Where are the other ten thousand days? We mentioned before about long gaps in data sets. The range from October, 1893, through mid-May, 1911, simply isn't there at all (6438 days). But small gaps are scattered throughout as well, many apparently in the middle of long weather events.

Year	Month	Day	Rain		GAP
1874	10	31	0.00		**4**
1874	11	04	0.27		**1**
1874	11	05	0.12		**2**
1874	11	07	0.05		**2**
1874	11	09	0.25		**3**
1874	11	12	0.22		**3**
1874	11	15	0.05		**6**
1874	11	21	0.12		**4**
1874	11	25	0.12		**6**
1874	12	01	0.00		**1**

There are 451 other gaps in the set, mostly single days but up to a six-month gap, making up the other 3770 missing data points.

Gaps in the data are not a huge impediment to finding the peak values; we still have a pretty large sample of days. But we are going to look at 48-hour totals, and many of these gaps were in the middle of stormy periods, so those two-day sums are gone. This means we lose some chances to catch extreme values, and the missing pairs will affect the fitted curves we're going to generate to simulate super-long data sets.



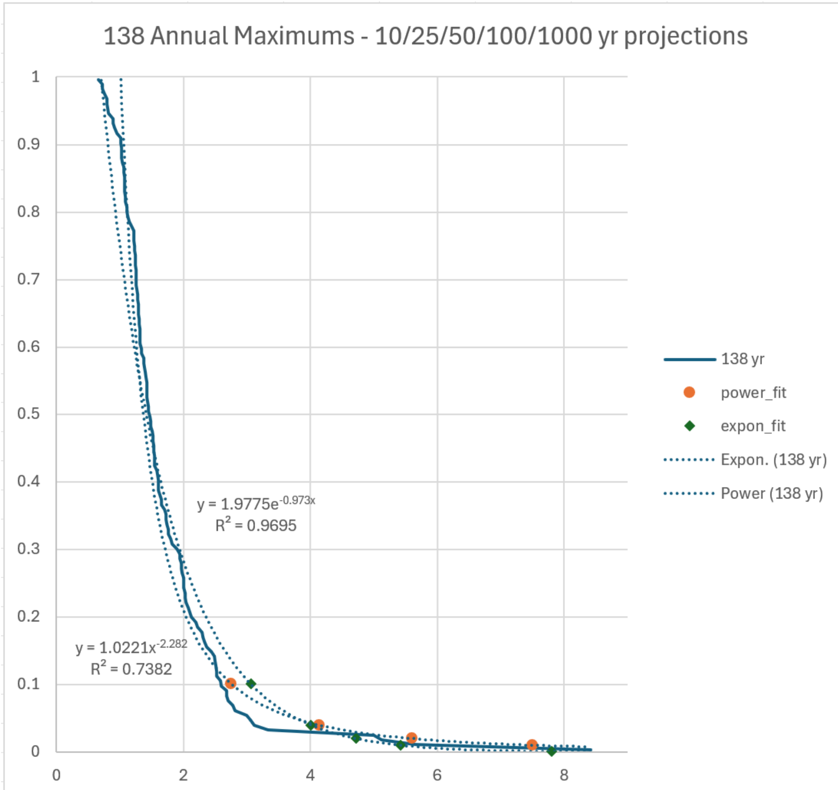
In the past I have had to deal with many unexplained holes in weather data. Sometimes a station is unmanned or out of commission *because of severe weather*, an obvious drawback in extreme-weather event analysis. In October of 1945 Typhoon Louise swept across Okinawa, ripping through the camps and piers of what was to have been the invasion force of mainland Japan. Official wind speed on the first day was recorded at 130 mph – only because that was the limit of the local instrument. The next many days were not recorded because the instrument tower was destroyed by even faster winds. [The storm was also completely unpredicted. "They knew a typhoon was running through to the south of us. But for no reason, perhaps the whim of a bored Greek god, it stopped and turned north, growing stronger by the hour as it was nudged along by that neglected ancient immortal." – X-Day Japan [4]]

Older data, while valuable, has its own problems. Differences in methodology and instrumentation may not be documented. At a glance one might be skeptical of the precision of rain amounts from the 1860s recorded at Fort Bidwell as exactly 2, 3, or 5 inches (which is only seen after converting back from the archive's tenths-of-millimeter). In the same year we see records of 2.10, 3.05, and 5.12, so something was possibly amiss on those other days.

The GHCNd database is full of flags noting where data has been estimated or assumed from other sources. Those methods may be "reasonable", but they still introduce further uncertainty and noise in the data which may properly erode confidence in resulting statements

But taking what we have, we do curve fits like before on annual maxima for the entire long data set.

[4] Keylor, Howard, "Surviving Typhoon Louise", http://danielborgstrom.blogspot.com/2012/08/surviving-typhoon-louise.html

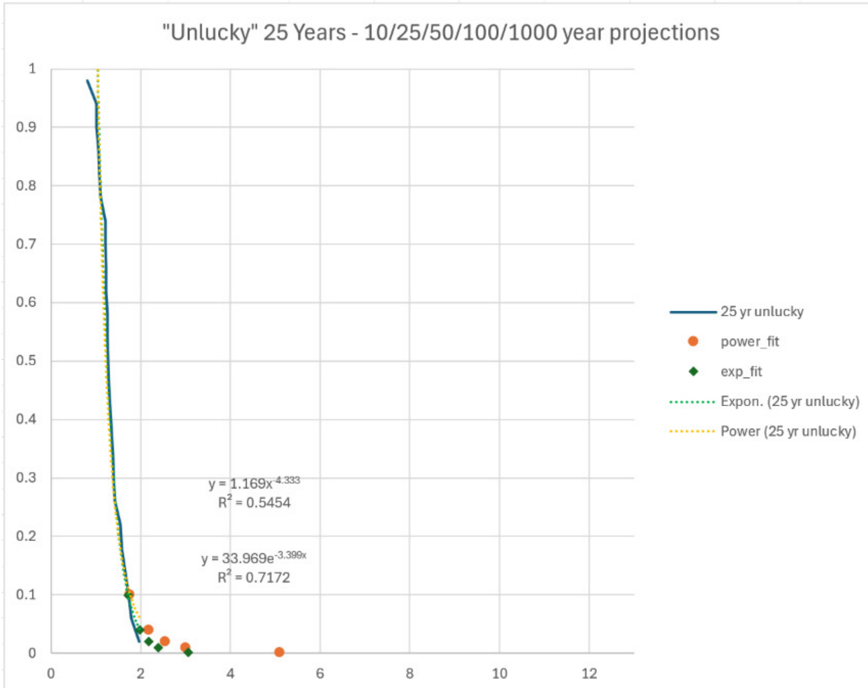


It's no surprise that the 100-year projections, about 5.5 and 7.5 inches, lie inside the real data set, which is a good bit longer than 100 years.

Different Previews of the Big Show

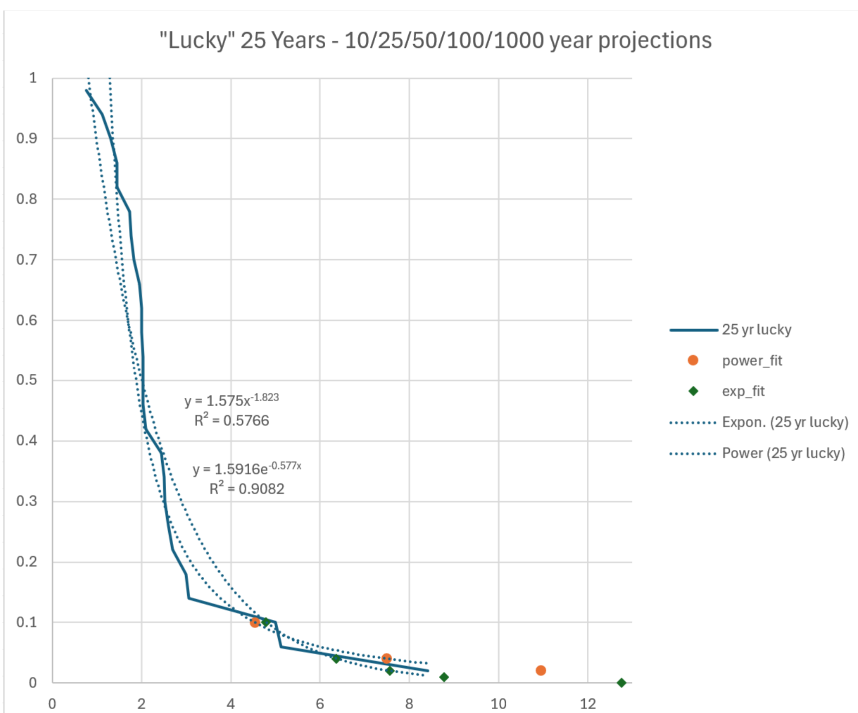
But this gives us a chance to experiment. What if this was a location for which we had less data, like at most other stations? We could be trying to do our estimations at any time in history. So, we will look at a couple different 25-year-long samples.

If we start in 1911, for writing an insurance policy for someone lucky enough to afford a home during the Great Depression, there are very few large rain events to use.



Extrapolating from this short history we get hundred-year events under three inches and even the thousand-year event predictions are in the low single digits.

If we start back at the beginning in 1868, because we want to build near a river in 1893, we get a very different graph.



The 100-year event in the former case is under 3 inches of rain. In the latter "lucky" case, the 100-year event is *16 inches of rain* (both from the power laws, which had better fit).

Using a weather history with 128 years of observations was certainly better than using the 23-year set. But as noted very few weather stations have data going back that far. Can we say something about the certainty of our calculations when there is a limited number of years of data to use? We will try in the next section.

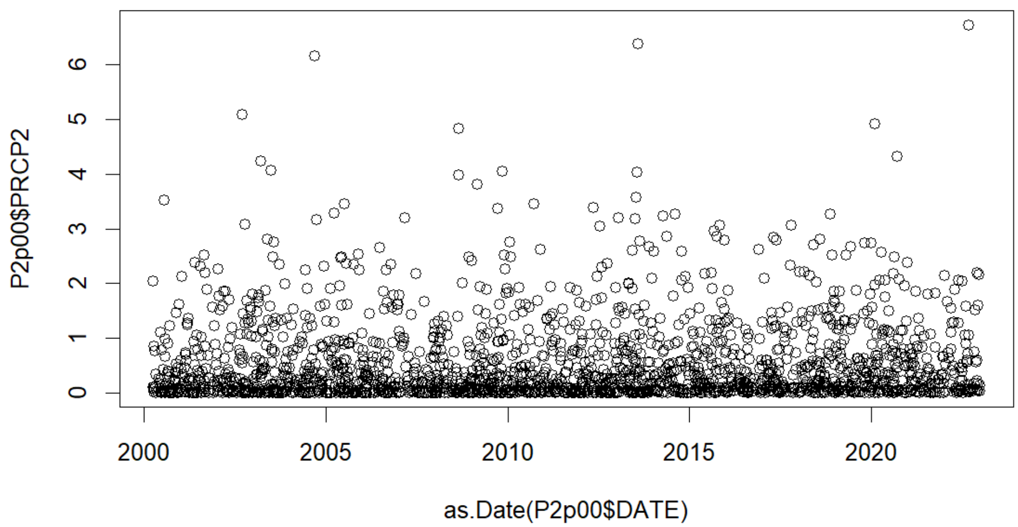
Part Three – An Enormous Data Set (and small bites of it)

We've come to the question of how well or badly we can make estimations of future severe weather events from limited data. One fun way to find out is a Monte Carlo simulation (this is only fun when one has a fast computer). We will try to make a very long, practically infinite, string of fake weather history then sample it many times (randomly or exhaustively) and see how likely we would be to observe the most severe events in different situations.

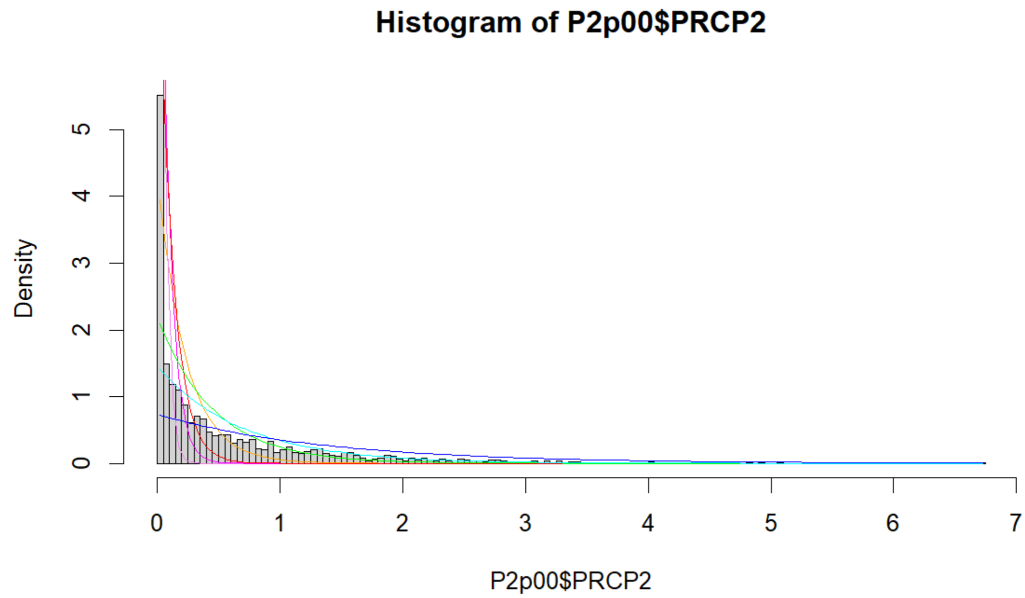
Starting From Short-Term Real Data (again)

The first attempt at a Monte Carlo study will use the Clemson-Oconee data. Its 23 years of good data is about average in length for stations in the U.S, as we saw earlier.

First all the data for total precipitation (two-day events) is loaded into our program.



There are an awful lot of zero values, a majority in fact. Right away the positive values are split off and we drop the dry days. The days which had positive precipitation can be put into a histogram that looks like this. (This histogram is normalized to have a total area of 1.0, making it usable as a piecewise probability distribution.)

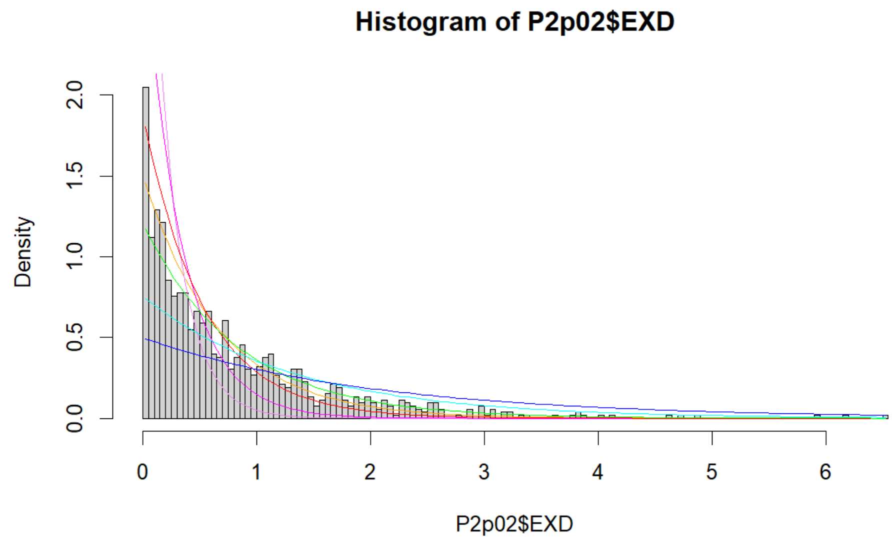


There are still an awful lot of days with just trace amounts of rain. The colored lines are attempts to fit “exponential distributions” to the data. The exponential is appreciated for being simple mathematically and tending to look a lot like plots of real events that have diminishing frequency related to the magnitude of the event. Using the exponential is the only good assumption we’re going to make today.

But in this case trying to fit a single exponential to the whole set doesn’t look very good. In the real world we expect that weather events are related to specific conditions, like a cold front moving through from Canada, or a warm damp air mass coming from the Gulf of Mexico, or a collision of the two. Each of those conditions might have a distribution of resulting weather effects like precipitation. The combination would have another, probably more severe, distribution. Every other notable atmospheric condition, and all their combinations, might have different distributions, and each of those conditions has a likelihood of being present in the first place. A single smooth curve may not fit the sum of all those situations.

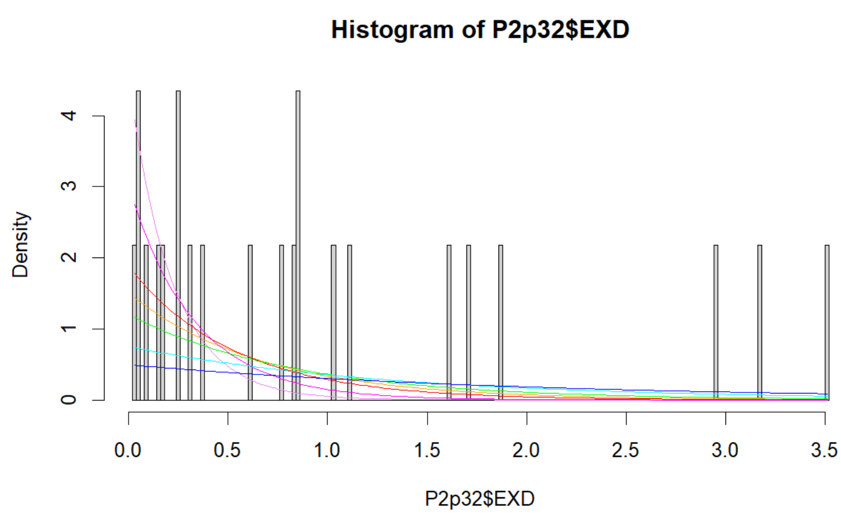
Data Fitting by Block

For purposes of this study, we’re going to break up the data by amount of rainfall, assuming that ranges of rainfall correspond broadly to the different conditions (I did say our last good assumption was behind us).

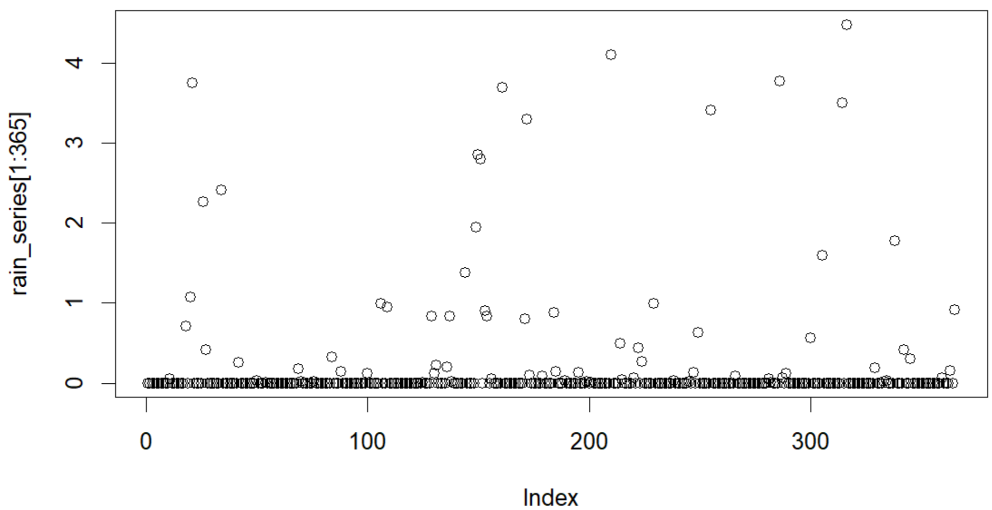


This graph shows the frequency of all rain events that exceeded 0.2 inches, by the amount of exceedance. Now a single exponential curve fits much better. The best fit, the orange line, has an error 91% lower than in the first graph, according to one standard measure.

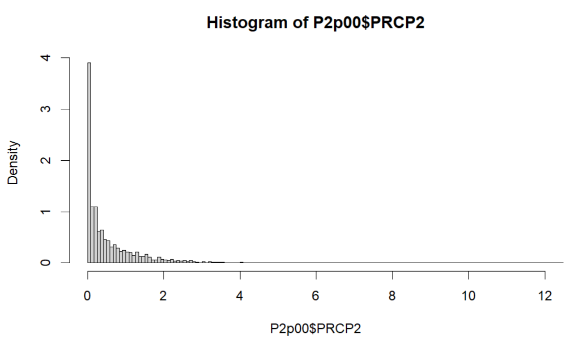
The same thing is done for exceedances over 0.2, 0.4, 0.8, 1.6, and 3.2 inches of precipitation (in two days). Data does get a bit sparse by 3.2.

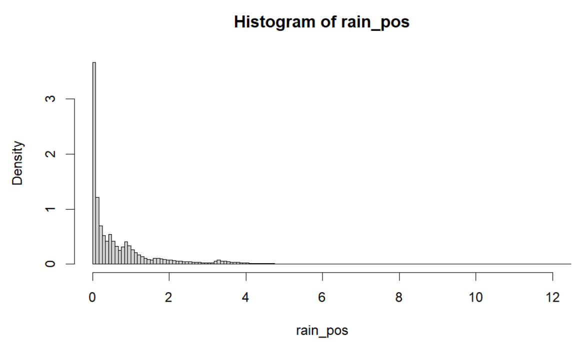


The trick then is to add the six distributions together to get one "reasonable" bundle that we can use to make a very long realistic synthetic history. We know how many days in the real data had no rain, over 0.2 inches of rain, over 0.4", etc. Weighted samples of each distribution are taken roughly by those frequencies. Some tweaking is done to get the annual total rainfall to match history and for the total distributions to be comparable. One year of fake data looks like this,



The real and made-up distributions are compared.





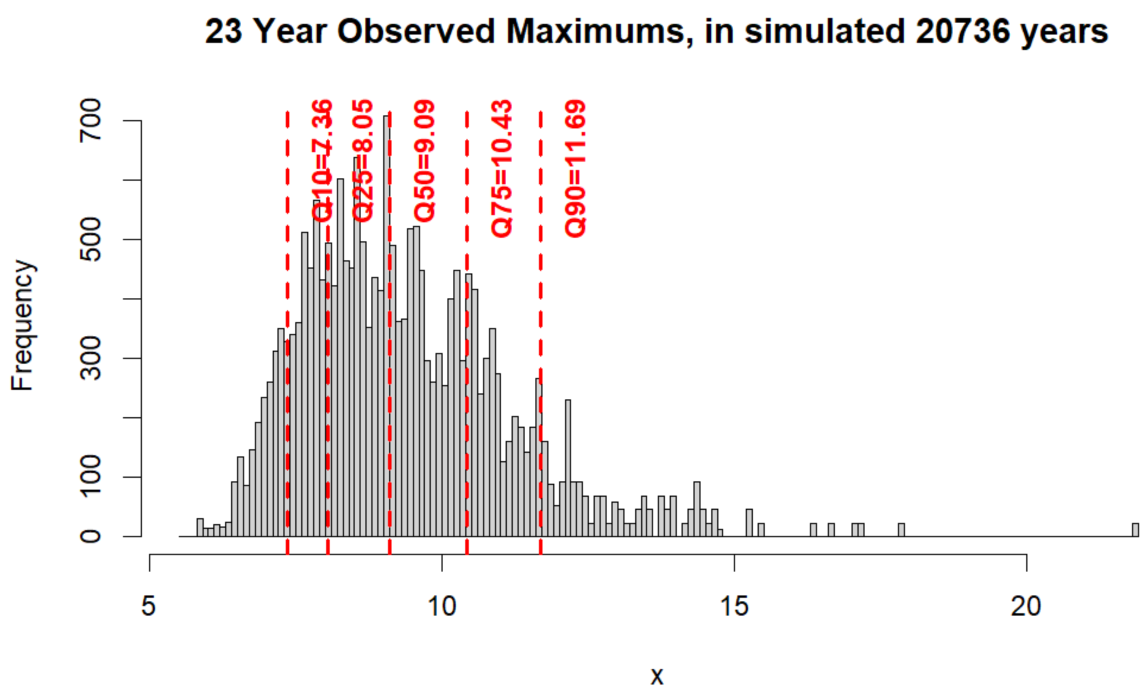
Each gives an average annual rainfall of about 47 inches. The fake distribution is a bit lumpy, which is no surprise since the contributing functions overlap. No attempt was made to fit these together smoothly. Other defects in the fake data include lack of seasonality and effective independence of all the data. But here our interest is in the extreme events, not the mundane.

The simulation distribution may be juiced a wee bit to give a better chance at big values. This allows for better illumination of how good or bad a short sample can be. Remember that we only have about 23 years of data, out of thousands of possible years. The point of this Monte Carlo study is to test the likelihood of a relatively small sample revealing the true distribution of severe events.

The new distribution is used to randomly generate daily data for 20736 years (12^4), over 7.5 million days. The daily draws are independent, but the distribution is made from two-day data already, so the simple one-step influence (rain today is correlated with rain tomorrow) is baked in.

Checking Every Possible Subsample

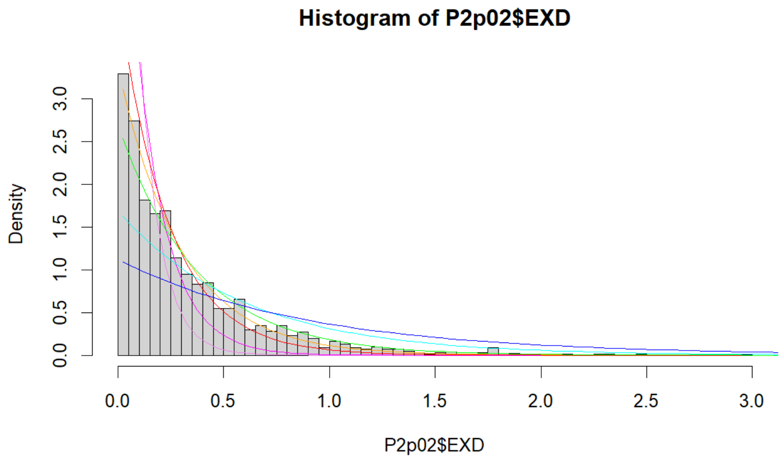
Now we're going to pretend that we only have a 23-year long window of observations into that long history. There's no telling where in the timeline our particular sample starts. With all the rain data stored, the maximum of each 365-day block is saved into a list. Every possible 23-year set (year 1-22, year 2-23, year 3-24, etc.) is checked for its observed maximum and that list of 20714 maximums is charted in a histogram below.



In all those possible short histories, we have even odds of observing nine inches of rain and 75% of the time we would observe annual maximum rainfall of under 10.43”. Rainfall in excess of that was found 252 times in our simulated long history, with a maximum of 21.8”. Some of the possible histories could hint at severe possibilities, but many will fall well short.

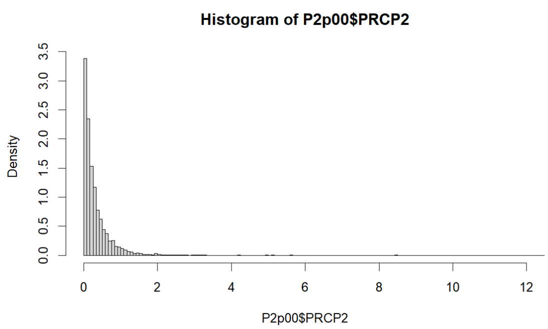
This particular Monte Carlo simulation gave a single outlier of 21.8 inches of rain. The series could be generated again many times (as it was) and not make another data point like that. Still, it reinforces the point that the best we can ever do is come up with an odds-of-exceedance beyond some semi-extreme value. The true extreme of a scalable event knows no bounds short of the limits of the entire system. In this case that would be every molecule of the atmosphere conspiring at once to reinforce a storm in one spot!

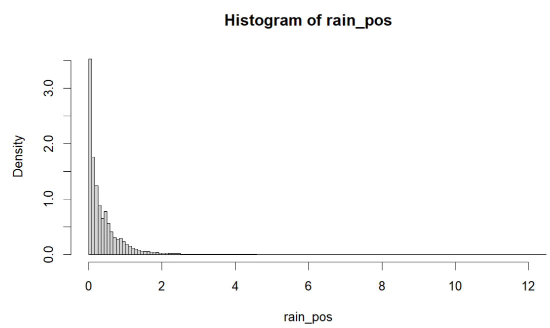
So, what if we have a lot more data? Let’s try that long Ft. Bidwell set. It’s a much drier location, with about 17.6 inches of rain per year. We fit data to each exceedance range as before.



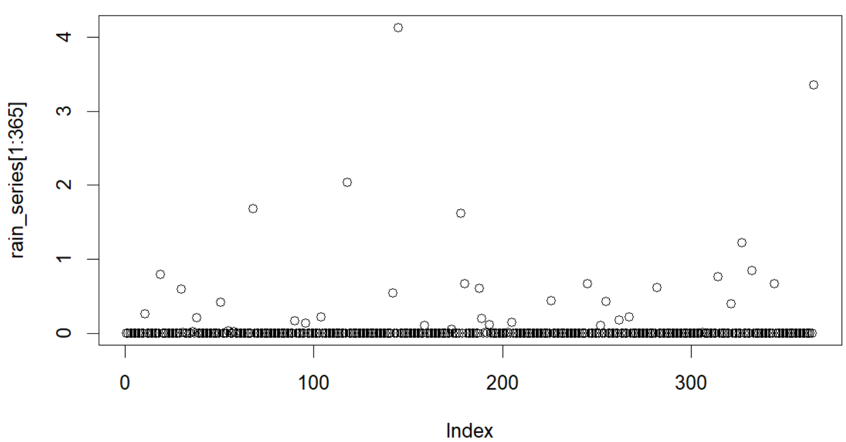
It's still difficult to fit exponentials which catch the large bunches near zero but also allow enough events in the mid-range or higher values.

With more fitting and fiddling we get a distribution for simulation that gives a similar looking graph and about the same annual rain fall (17.3" simulated vs 17.6" historical).

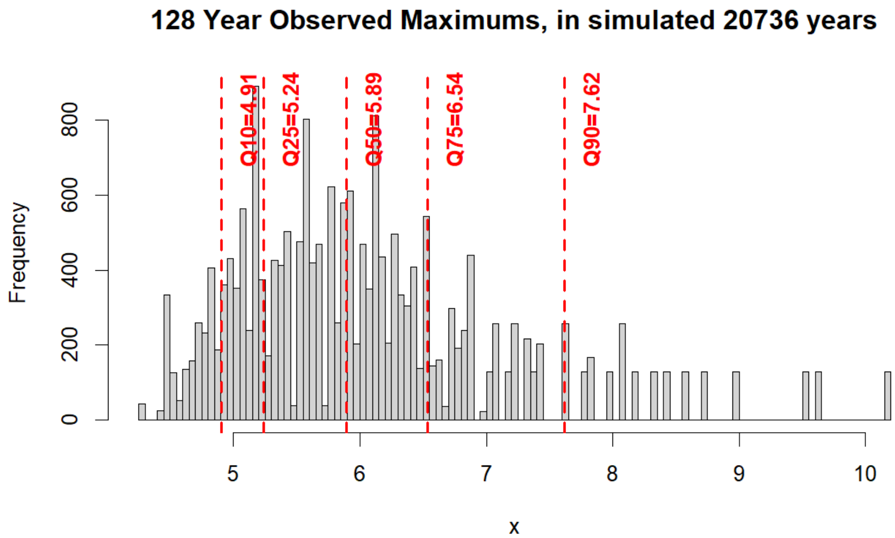




One sample year of precipitation looks like this,

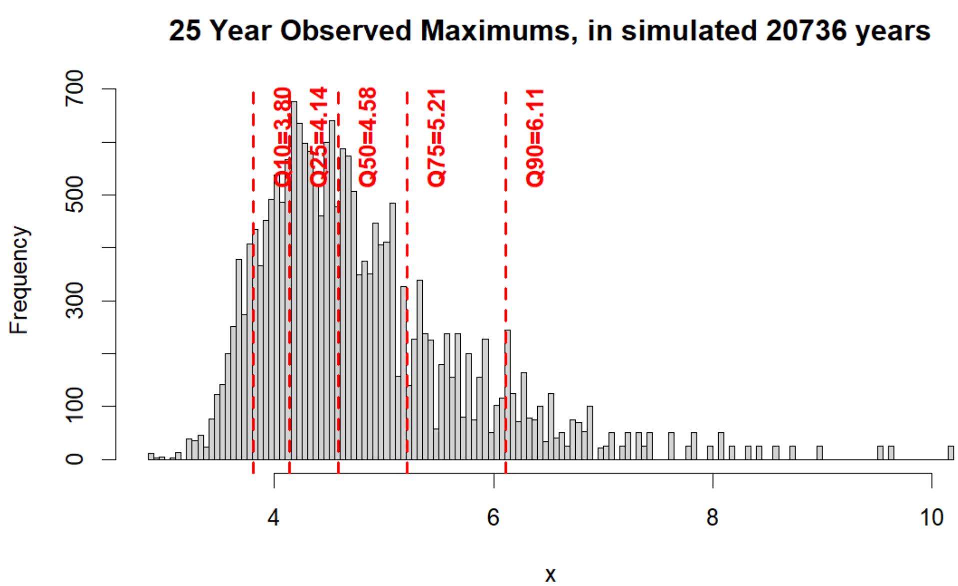


Again, a very long series is simulated, 20736 years. If we had the same 128 years to sample this long history at every possible starting year, our maximum observations would plot like this.



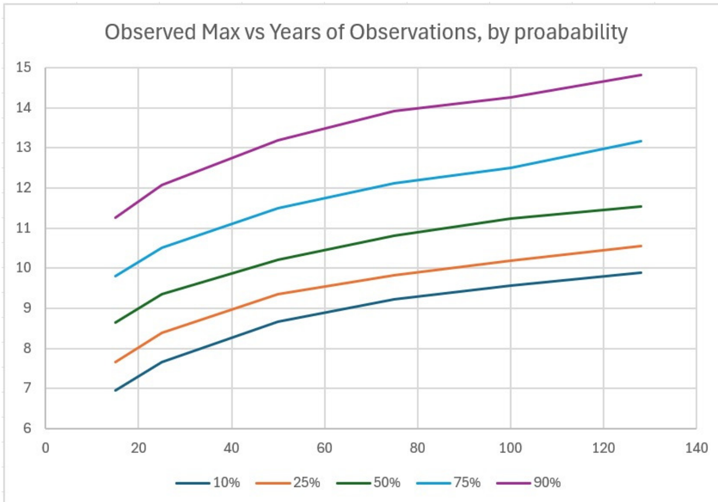
Sampling 128 years at a time there's a 75% chance that we will never see a maximum rainfall event over 6.54 inches of water. Such an event happened 49 times in the long history, with a maximum of 10.15". That's a 55% difference, and a big absolute jump for a dry place.

How much worse would it be if we only had, say, 25 years of data?



Now the 75% line is at 5.21 inches, which is less than 243 rain events in the long history. Compared to the same overall maximum, it's now a 95% jump!

If we do this for 15, 25, 50, 75, 100 and the full 128-year sample sizes, we can tabulate the quantiles and graph them together.



As expected, the likelihood of finding an extreme value increases when we take a longer sample, and it's not quite linear. There's less payoff to adding more years the more years we have, and at the other end having fewer years things get bad faster. We see now how the selected "lucky" and "unlucky" periods in the previous section had such different 100-year event predictions. The prediction is sensitive to the last few maximum values, and they don't come by very often.

It should be noted that this synthetic sample probably gives milder results than a true long-term weather series would give, for a few reasons.
- The data are truly independent (as well as a computer random number generator can make them), limiting clustering as might be seen in periods of drought or prolonged wetness.
- There is no seasonality, which would lead to smaller scale clustering.
- The data is "stationary", with no long-term trend. Long term climate shifts are a matter of record, and at this epoch in history increasing weather volatility is expected pretty much everywhere.

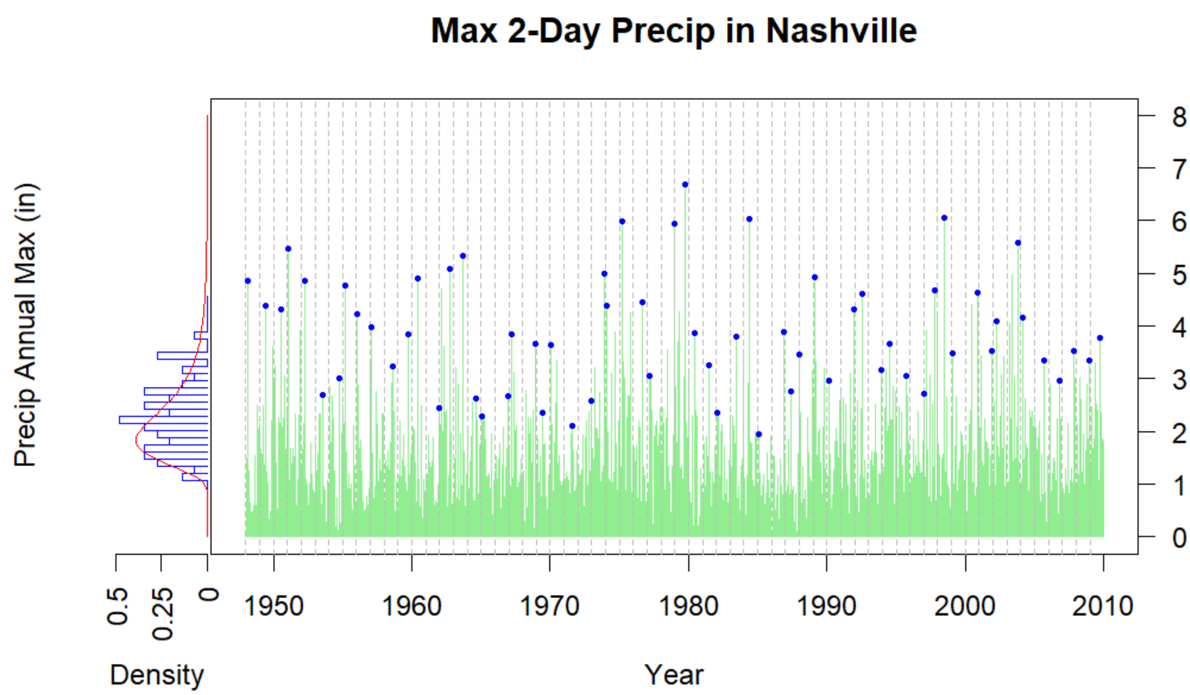
So far, we've tried generating 100-year event numbers in a simple way on short data sets and then experimented with a very large but very made-up data set. As a last step we'll look at results from a more sophisticated analysis method (and then try to break it).

Part Four – Pretending to Be Super Smart

Probability theory is full of theorems that work “for very large n”, the number of samples in a sequence. Taken to the limit, the observed maximum from a large data set should fall in line to one of a few simple-ish distributions, no matter the original distribution. The theory is air-tight, at the limit. But it leaves a person living in a real particular location and time, with explicitly limited sets of data, asking, “How many ‘n’ is large enough? How small of an ‘n’ is too dangerous to use?”

The Fancy-Pants Way to 100 Year Floods

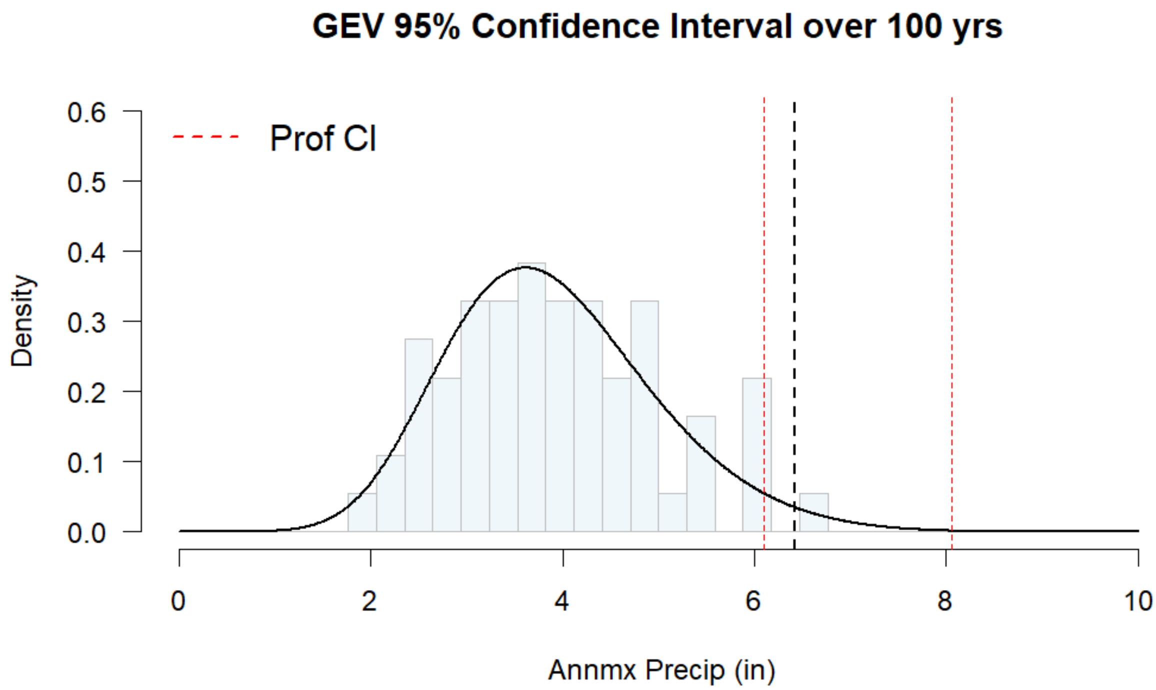
We will apply one analysis method to data from near Nashville, Tennessee. This is a relatively long and high-quality data set. (There are “only” 213 missing values in the raw precipitation data from 1948 onward.) To start, I insist we use data only through 2009, for a reason that will become clear.



In the busy plot we have combined daily rain data for Nashville, select annual peaks marked by blue dots, and a histogram of all the annual maximums plotted up the left side[5]. A red curve is fitted to the histogram. This curve is from what is called the “generalized extreme value distribution”. It gives us probabilities of extreme values, higher than the maximums already observed, just like the exponential or power law plots of the first section. But the GEV has formal theoretical rigor behind it: for enough samples, the solution is exact, if certain requirements are met and we have all eternity to sit and watch.

[5] Code for generating these plots was adapted from materials provided by Professor Whitney Huang, Clemson University.

Since we don't have eternity to wait or an infinite amount of data from the past, we need some idea how the size of our actual data set affects our answer. It's a straightforward thing to calculate "confidence" intervals based on the assumed distribution and how many samples we got from it.



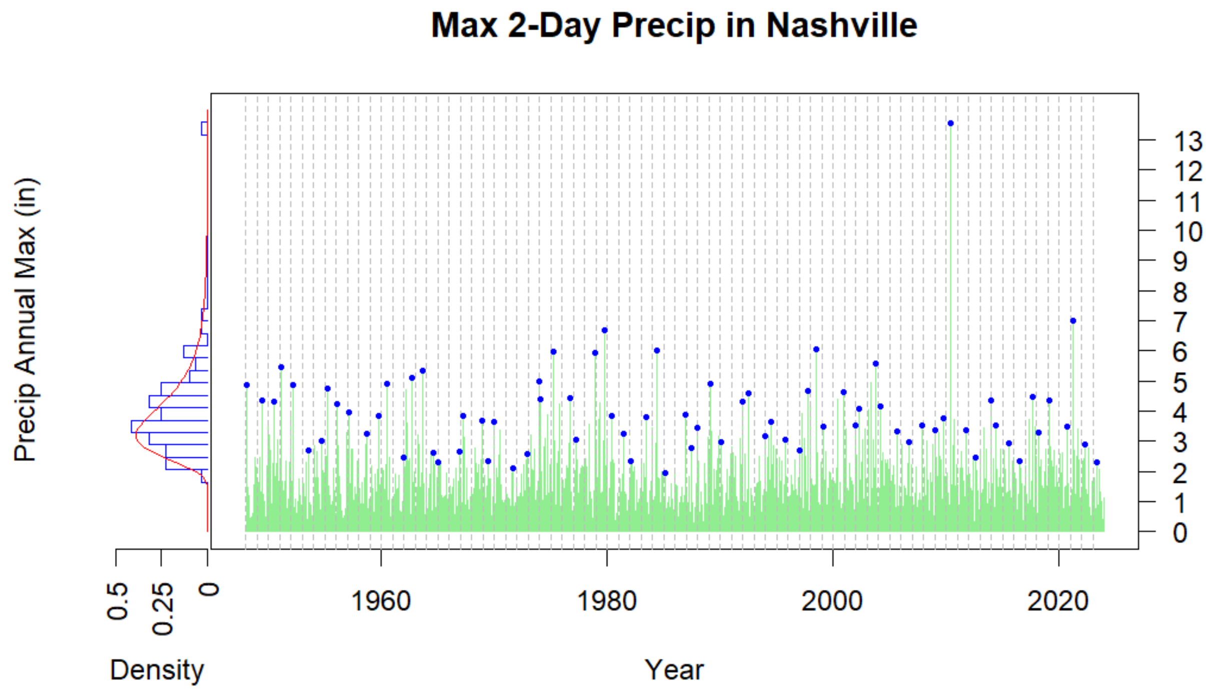
This is the same histogram as the sidebar in the previous plot (note that the bin spacing is not the same) laid out normally, with the predicted 100-year extreme and a 95% confidence band around it. We have a 100-year rain event of 6.4 inches, with a 95% confidence band between 6.1 and 8.2 inches.

Breaking the Model with new Data

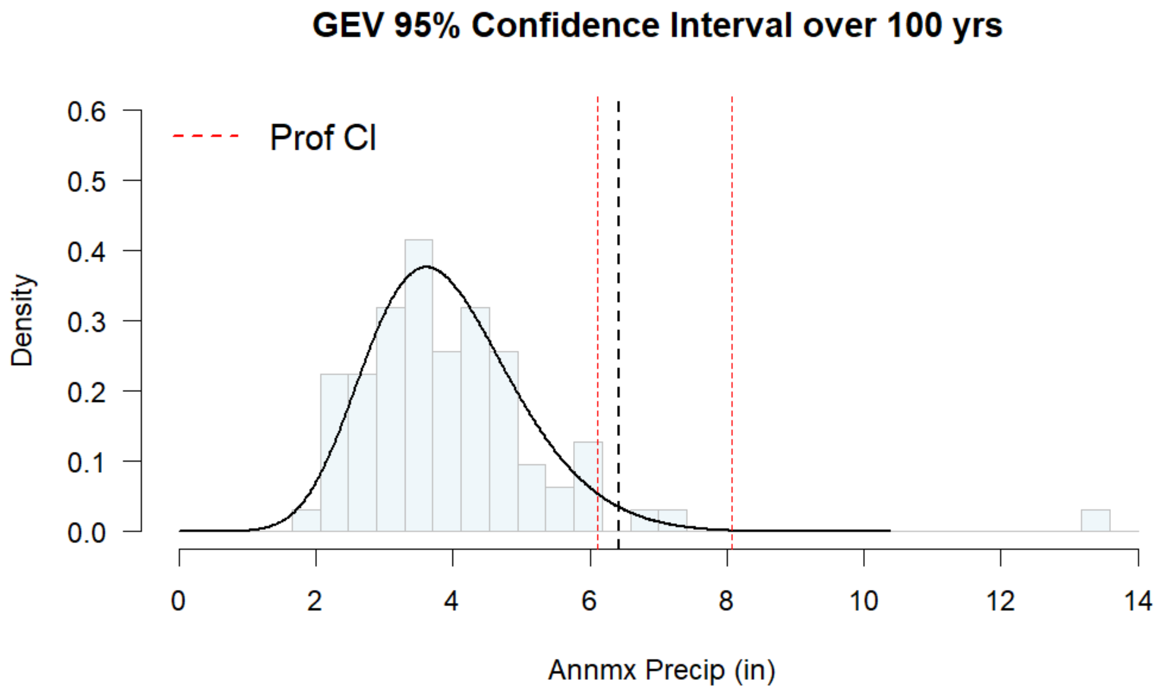
OK, now I'm going to get mean. Nashville was chosen because of a known single super-extreme event. A great real-world example of two-day pairing is the May, 2010, flooding of downtown Nashville. The National Weather Service reports that 7.25" of rain fell on May 2nd, besting the previous record of 6.6 inches. But the two-day total record, adding in 6.32" from May 1st, leapt to 13.57 inches of rain. The previous two-day record was 6.68", only 0.08" greater than the old one-day record[6].

The previous maximum 2-day rain event was less than seven inches. So now we have a serious outlier to deal with. How will that affect the analysis? The whole set of available rain data is brought in, from 1948 through 2023.

[6] National Weather Service, (2020), "10th Anniversary of the May 2010 Flood", https://www.weather.gov/ohx/10thAnniversaryMay2010Flood



The event from 2010 stands out dramatically. Only one event since has even equaled the old record, about seven inches. The GEV model fit didn't allow much chance of this happening (it's beyond the 10,000-year event). But we can at least update the model with the new information and talk about how much smarter we are now.



Our expected 100-year event is now at 6.4 inches of rain, all the way up from 6.4! We might have expected the value to go up, so that's curious. The good news is we have an even tighter confidence band, ranging from 6.1 to under 8.1 inches (previously 6.1" and 8.2"). How could that super-extreme event not have a greater impact on our analysis?

The new data came with one extreme but a bunch of other new data that looked much like the old. The key assumption of this GEV process is that all the data comes from the same "generator", the real physical process that makes it happen, imagined as a probability distribution. The weather is not like that! Different storms are made in distinct contexts. The generalized extreme value theorem is not general at all – it applies only to unlikely situations where nothing ever changes and every day is (probabilistically) like the last.

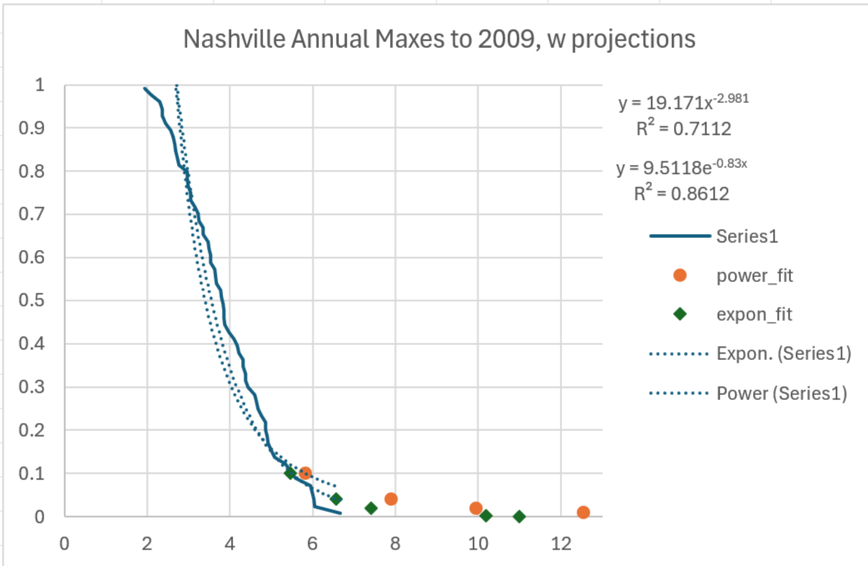
Mathematical rigor cannot keep your house from floating away if the axioms are materially false. But it can fool journalists. In one piece we see a promising headline, "Why the May 2010 flood won't happen again in our lifetime". The report goes on to completely fail at both skepticism and probability[7]. The headline is not backed up and is in fact countered by the canned explanation of "1000-year floods."

Back to Basics

What about that "crude, simple" analysis method we tried at the start?

Using just the data through 2009, the sort, plot, and fit routine, all done on a spreadsheet, gives some usable conservative curves.

[7] Melanie Layden, (May 3, 2024), "Why the May 2010 flood won't happen again in our lifetime", https://www.wsmv.com/2024/05/03/why-may-2010-flood-wont-happen-again-our-lifetime/



annmax	(prob target)	power
5.83	0.1	0.10004
7.9	0.04	0.04044
9.95	0.02	0.02033
12.54	0.01	0.0102
27	0.001	0.00104
		exponential
5.47	0.1	0.10151
6.57	0.04	0.04074
7.42	0.02	0.02012
10.2	0.01	0.002
11	0.001	0.00103

As usual, the power-law rule fits better and gives more conservative estimates. We have a 100-year event of 12.54 inches (compared to the real 13.57”), and a 1000-year event of a dangerous 27 inches of rain. Numbers like that might have prompted people to build and provide dams and drainage in a very different way.

Which brings us back to the beginning. This all started when the south side of Asheville, North Carolina came to be under ten feet of water. We look at the rain history and how it translates into flooding in the next section.

Part Five – Back to Where It All Started

Some weeks ago, this author began looking into how weather history relates to “100-year flood” numbers that are used in many contexts. It was supposed to be a short study with a quick bit of math to show how history may not give good extrapolations very easily. It has been a lot more work than that – because this is hard!

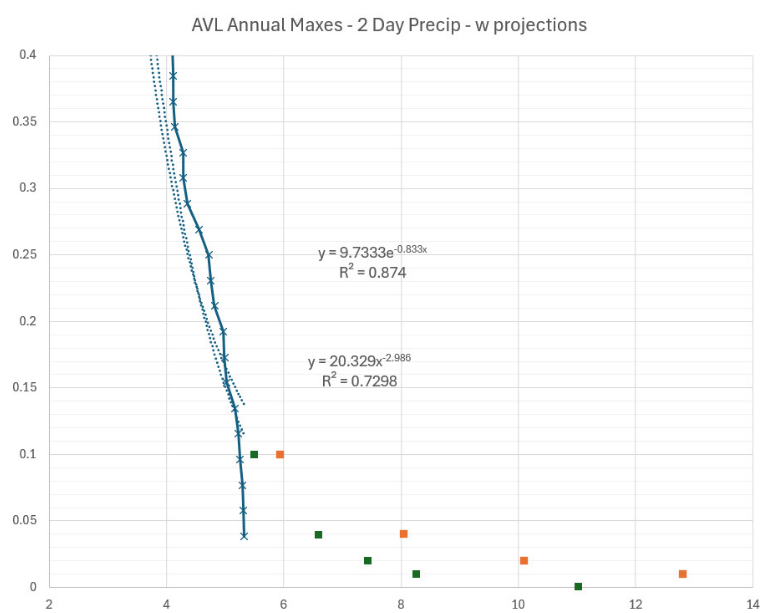
Shopping for Numbers in Asheville, NC

At the end of September, 2024, hurricane Helene pushed up from the Gulf of Mexico right into the southern Appalachian Mountains. Many roads and parts of towns were washed out. The southern part of Asheville, NC was put under ten feet of water. What caused this was a multi-day rain event which started with an all-day soak and then two days of heavier rain, as seen in this data from the Asheville airport, a bit south of town.

2024-09-23	0.1
2024-09-24	0.21
2024-09-25	4.09
2024-09-26	5.78
2024-09-27	4.11

By our two-day max pair method, the high value out of this storm is 9.89. We could reasonably add in the 4.09 for a total of 13.98 inches of rain.

Taking data from a nearby airport, back to 1973, we go through the same old procedure. At the high end of the annual maxes there is a cluster of years near the same value, about 5.25 inches. In 52 years, there are no events within 60% of the 9.89” observed in September, 2024.

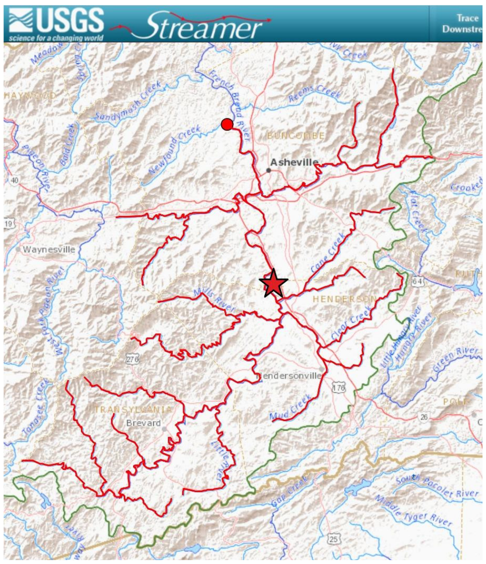


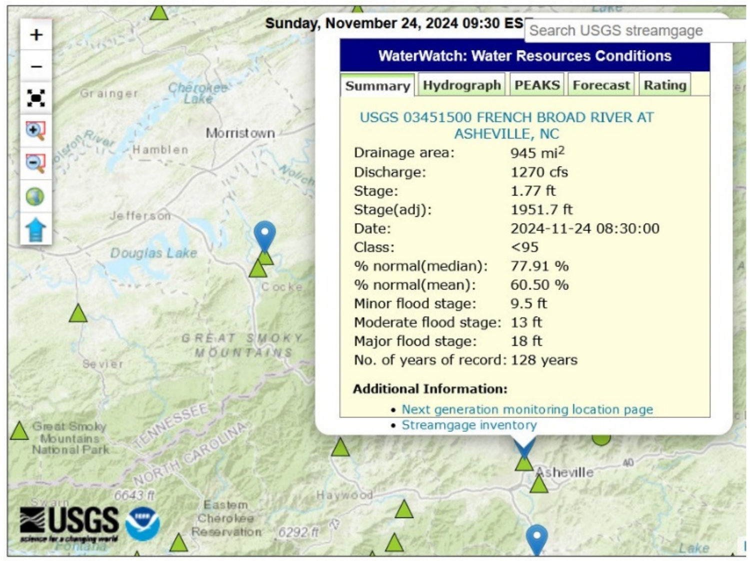
For this graph we’ve zoomed in some to see the details. The exponential fit gives a 100-year event of 8.25 inches, the power-law fit 12.8 inches of rain. The exact numbers are debatable (as we have well shown), but probably the basin around Ashville should expect rain as heavy as was recently seen more often than once in a lifetime.

Transmogrifying to Flood Levels

This extends to actual flooding, of course. The Swannanoa River hit a new record of 26.1 feet above nominal level. It's time we investigated the leap from rain amount to flood level. We're not going to do this deterministically, but some of the problems with that are noted next.

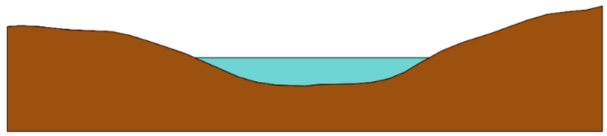
The first matter is the total basin area collecting water. At this point most of the USA is well mapped and encoded in databases like behind the USGS "Streamer" web site. All rivers upstream of the gage point are highlighted in red. The location of the AVL airport is marked with a star.

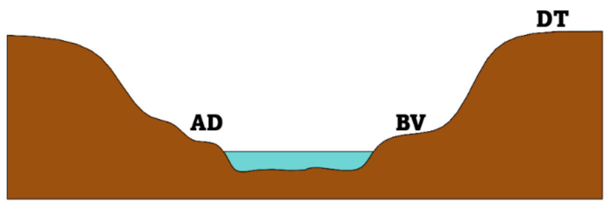


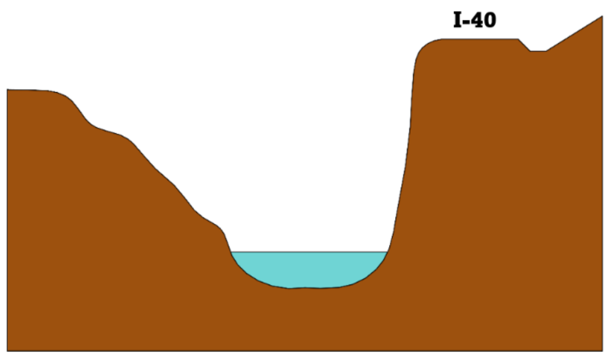


We noted before that flooding results from rain are time dependent. On top of simple upstream area, we would need some idea of the permeability of all the ground to estimate if it may be already saturated.

The point marked on the preceding map is the location of a stream gage on the French Broad River on its way out of Asheville. A simple gage can give flow speed at one point in the water. Another might give height of the water over some reference. With those two values in hand, we still need to know the shape of the river bed to calculate anything. Those shapes can vary a lot.



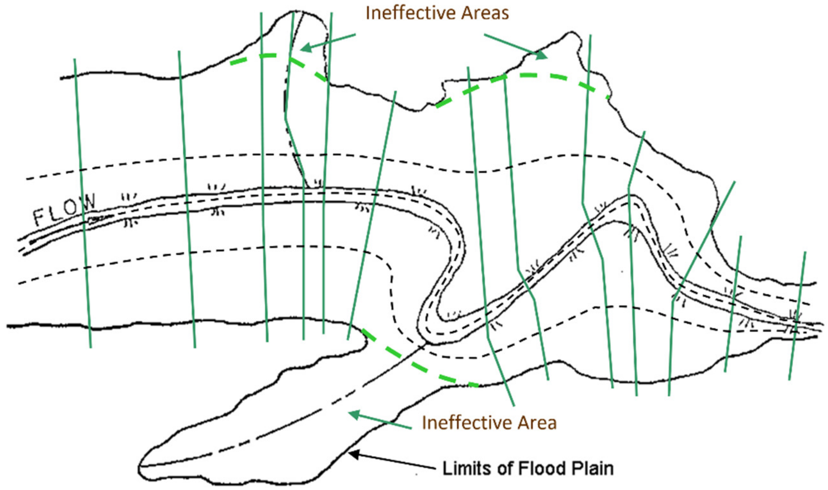




These cross-section diagrams show different stream beds. The top might be through Mississippi farmland in Iowa. As the water level rises the flow area increases quickly, but the water volume is moving more slowly downstream. The middle diagram is an (exaggerated) Asheville, showing Biltmore Village and the arts district in the flood plain, but downtown up on a plateau. Here the flow-versus-depth is quite variable. In another spot we might have an artificial bluff, like the supporting fill of a major highway like Interstate-40 near the Tennessee border. As the water level increases here the flow speed stays elevated, which is how we get the severe scouring that undercut the freeway in several places.

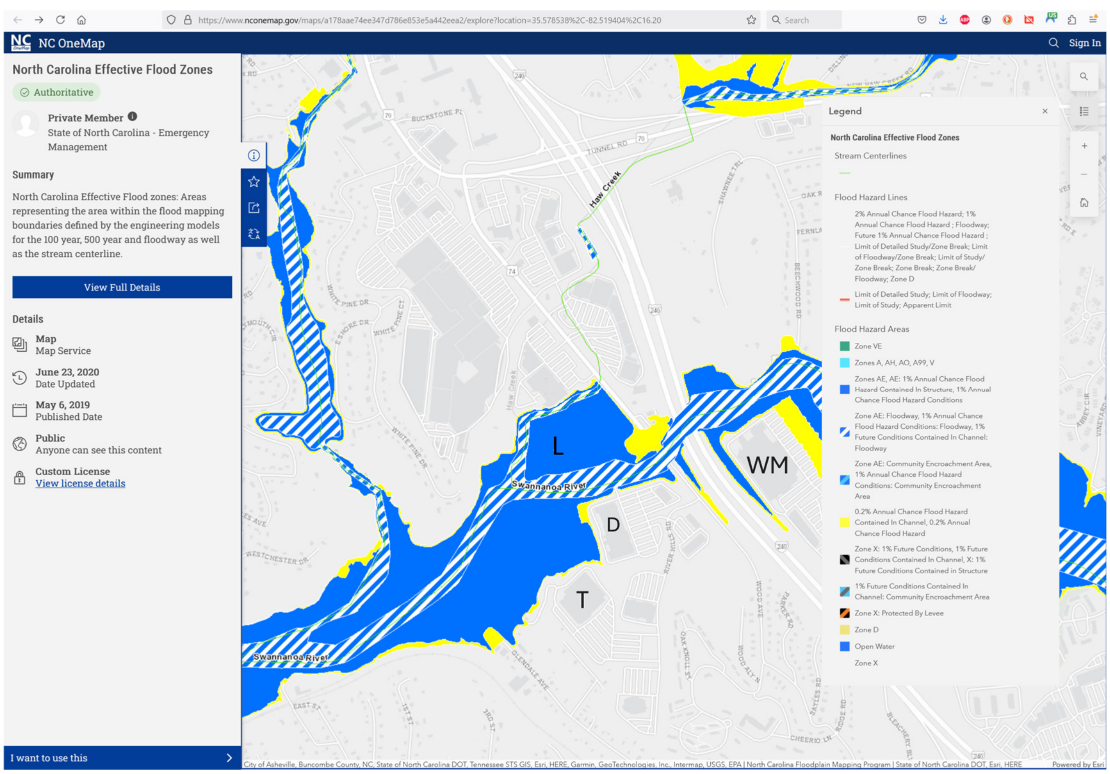
The appropriate cross-sections of a river at a point of interest may not be flat or perpendicular to the mid-stream [8].

[8] US Army Corps of Engineers Hydrologic Engineering Center, HEC-RAS Hydraulic Reference Manual, https://www.hec.usace.army.mil/confluence/rasdocs/ras1dtechref/6.3/basic-data-requirements/geometric-data/cross-section-geometry

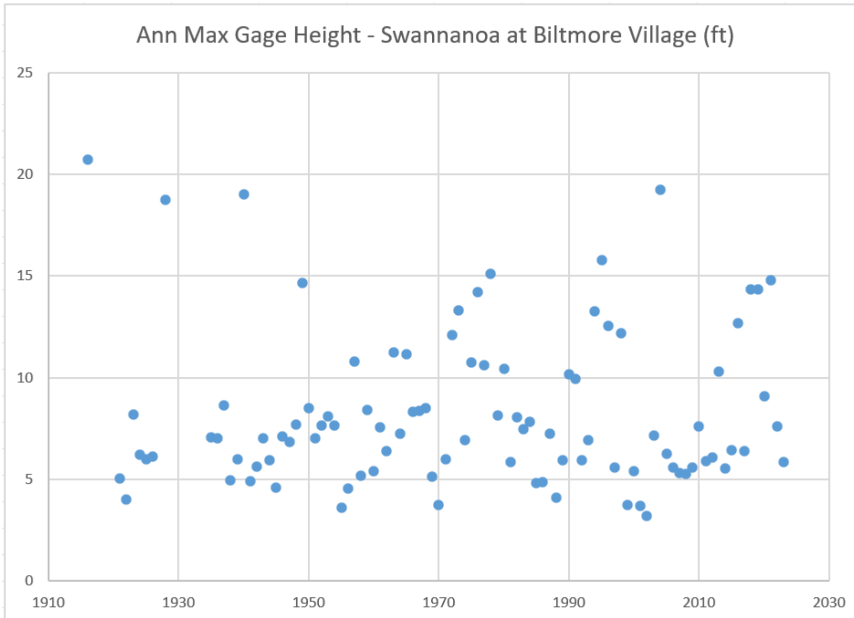


Some diligent mapping (and a measure of educated guesswork) goes into making a usable flow map. All that is to say it's a lot of work. With all that work done one can begin to record data on water flow versus water depth for a gage location. From there a flood map can begin to be put together along a river.

Here we see such a map along part of the Swannanoa River, through the primary big-box retail district for Asheville.

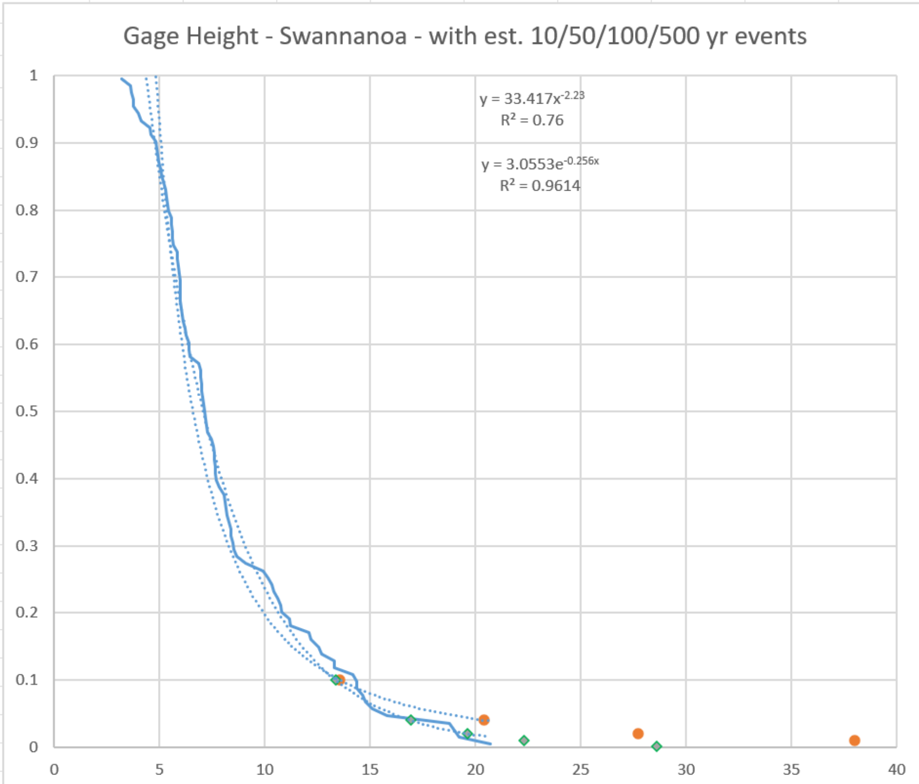


In this official map solid blue is the "100-year" flood zone, yellow the "500-year" zone. Walmart, Lowe's, Dick's, and Target, labeled by letter, were all under water up over their front doors in the September, 2024, flood (One may wonder who Lowe's has for an insurance company…). These retail outlets could have been a big help during storm recovery and clean up, but they were wiped out. Was this really a beyond-500-year event? Let's look at the data in our simple way.



Plotted here are data for maximum level of the river near this district. Consistent records from a stream gage go back to 1921, and there looks to be reliable information from a famous flood in 1916 so we include that. The new max for 2024 (as of this writing) is 26.1 feet. But the river has been near twenty feet twice in recent years and over 15 another few times[9], so we might suspect that gives us some semi-extreme events to direct a projection of the more extreme.

[9] Victoria A. Ifatusin, (October 4, 2024), "It's official: Helene's rainfall, flooding broke all-time records", https://avlwatchdog.org/its-official-helenes-rainfall-flooding-broke-all-time-records/



annmax	(prob target)	power
13.53	0.1	0.100271
20.4	0.04	0.040132
27.7	0.02	0.020288
38	0.01	0.010024
78	0.002	0.002016
		exponential
13.36	0.1	0.09993
16.91	0.04	0.040273
19.63	0.02	0.020073
22.3	0.01	0.010133
28.6	0.002	0.00202

1995 rain event:	5.25 in
2004 max gage ht:	15.79 ft
2004 rain event:	4.4 in
2004 max gage ht:	19.22 ft
2024 rain event:	9.89
2024 max gage ht:	26.1 ft

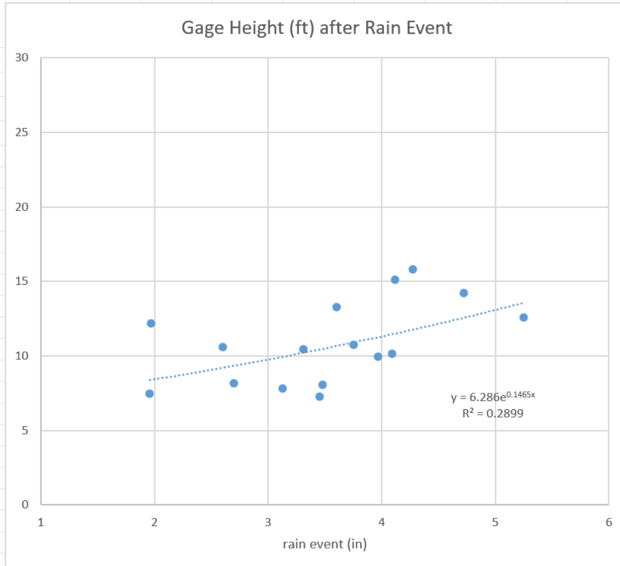
We put all the annual maximum gage heights in a row again and plot them by assumed equally spaced probability (At the beginning of this project I thought this was a horrible assumption; I see now it's the most minimal assumption we can make and much better than the assumptions that lead us into more 'sophisticated' analysis.) As usual the power law fits better but gives scary big numbers. We have 100-year flood levels of 22.3 or 38 feet, and the 500-year numbers are 28.6 or 78 feet. Again the truth is probably in between the pairs.

Plugging the new 26.1 feet observation into the fitted curve formulas, it would have been predicted as a 43-year or 261-year event. Those numbers are well apart, but both are far under the 500-year mark, which the flood zone map puts at least twelve feet under the actual water level we saw in recent photos.

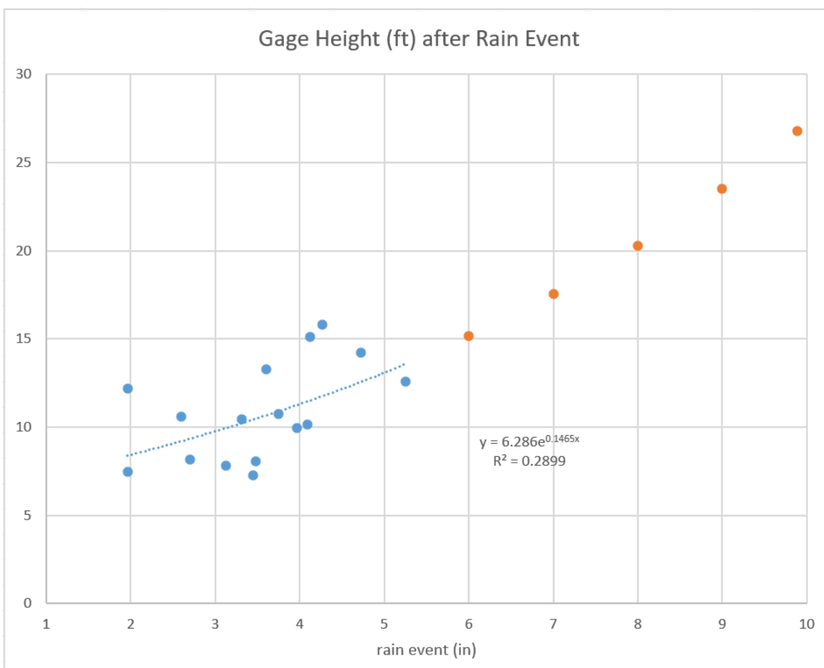
Correlating back to Rain Fall

We also listed a few comparisons of gage heights and the actual rain event that drove them (the preceding two days of precipitation). If four and a half inches of rain in 2004 got us to over 19 feet at the gage, it shouldn't be a surprise that almost ten inches of rain could put the gage over 26 feet. We can look at a lot of them and check the correlation.

	rain event	gage ht, ft
1975-03-14	3.75	10.75
1976-05-29	4.72	14.19
1977-04-04	2.6	10.61
1977-11-06	4.12	15.12
1979-03-05	2.7	8.14
1979-11-02	3.31	10.44
1982-02-03	3.48	8.07
1983-03-18	1.96	7.46
1984-02-14	3.13	7.83
1987-03-01	3.45	7.25
1990-03-17	4.09	10.15
1991-03-29	3.97	9.93
1994-08-17	3.6	13.28
1995-01-15	4.27	15.79
1995-10-05	5.25	12.56
1998-01-07	1.97	12.17
2003-05-06	3.33	7.16
2004-09-08	4.4	19.22
2010-01-24	3.36	7.6
2013-05-06	3.97	10.32
2015-12-29	1.84	12.7
2018-05-30	2.93	14.34
2019-04-19	5.31	14.35
2021-08-18	5.33	14.79



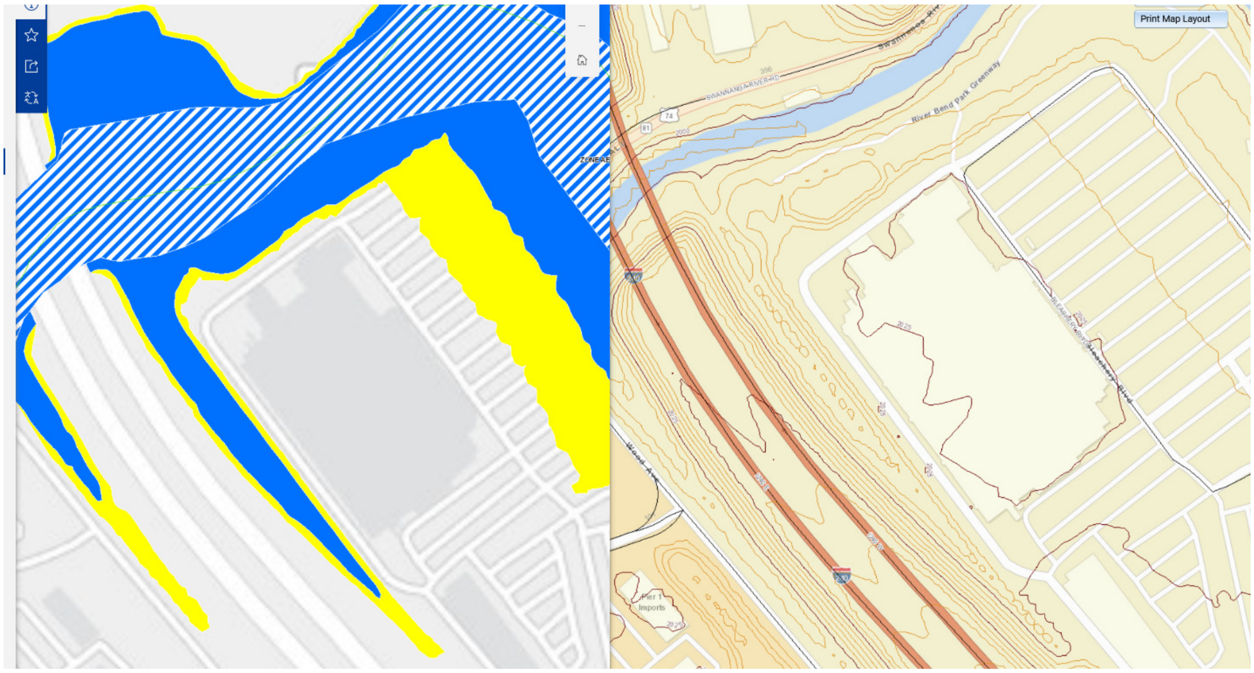
All the events where the gage height got over seven feet, and for which we have precipitation data from AVL, are plotted against the associated two-day rainfall. (Some expected snow melt was added to one of the smaller rain events.) It's a bit noisy, for all the reasons we explored earlier. A trend line only vaguely fits. But since we have a trendline, what nominal expectations does it give us?



With a humble level of confidence in our loose fit, we expect seven inches of rain should best the previous high-water mark from this set. Also, we project a flood level of 26.8 feet from a 9.89-inch rain event, whisker-close (for a coarse beard) to the true observed value of 26.1.

So how likely was that amount of rain? Going back to our fit of the rain data, we put that 9.89 value into the fitted curves and get a likely average frequency of 46 or 389 years, again somewhat divergent but both well under 500 years.

Comparing the official flood map to the Buncombe County GIS contour map of the area, it appears that the official 100-year flood line is 2015 feet above sea level and the 500-year line is at 2020 feet above sea level [the lines follow so closely one suspects the flood map uses the same 5-foot contours]. The front door of the Walmart is right at 2025 feet elevation, and the front doors were at least five feet into the flood water, for a peak water line over 2030 feet above sea level.



From this we estimate that the official 100- and 500-year flood levels correspond to about 11 and 16 feet on the nearby stream gage. The stream gage has been over that 100-year level 18 times in the last hundred years, and over that 500-year level three times. From our simple fitted function, we would expect these levels to be reached with average frequency of 31 and 66 years.

The analysis here is stubbornly simple, but it is as honest take direct from the data, and it is back-tested in this case by one dead-accurate prediction. We conclude that the official flood maps for the river basin upstream of Asheville, North Carolina, are, to the use technical term, “garbage”.

We should never forget; You Can’t Fool Mother Nature.

There's no hiding from her
She sees everything
Rules with the iron fist
You know she is the matriarch
https://www.youtube.com/watch?v=fs8RUFS18g0
[Unleash the Archers – The Matriarch]

Part Six - The Light at the End of the Tunnel is a Thunderstorm

Decisions need to be made. A building must be declared inside, or outside, of a flood plain. A dam must be built to specific dimensions. Decision makers want clear-cut answers to their questions. Diligent scientists can work long and hard to provide those answers. But science is hard on a good day – and there aren't very many good days. People providing answers need to understand and to be clear about the remaining uncertainty. People making decisions need to learn how to make decisions under uncertainty.

Conclusion and Recommended Further Work

In this article we mostly discussed ways to get hints about the future from past data, and limits of those methods. We also mentioned deterministic ways to work out how a drainage responds to rain events, and how difficult that can be in the details.

On the use of applying conventional statistical tools on historical precipitation data we saw that:
 - The available data is usually pretty bad, short in duration and full of holes or bad information,
 - Rare events do not show up often enough to give a good representation in small data sets,
 - The standard assumptions of statistics (independence, stationarity, identical distributions, thin tails) are false for real-world data, especially weather,
 - Sophisticated analysis techniques (e.g. GEV) may under-predict extremes for those reasons,
 - Simple statistical techniques (e.g. waterfall plots) may be a better choice, eschewing the precision of other methods when precision was never there to begin with.

This article is the result of two months part-time work from a diligent amateur (with a few graduate college math credits). It only superficially covers the problem of local weather modeling. And the issues of water basin mapping and functional modeling are barely introduced. But this article has enough firm results that it should inspire further related work.

We might suggest from among the following.

Data Cleanup and Expansion

This might be one good job for an AI tool. We saw one instance where archive rain data was clearly wrong, from a contemporary news report. Comparing billions of archive records to other sources is just the kind of laborious job that computers are built for. In addition to correcting data, more detail might be found, finer than daily data, where it's not already on hand. We don't need hourly observations for every day going back centuries, just more detail during storms, as might be found from written word accounts.

Checking Limits of the GEV/GPD Methods

This author is not well versed in the literature on the derivations and known limits of the formal extreme value theorems, but it has not been hard to find data sets which stretch credibility of their results. In particular, I suspect that time-series rain data is never going to fit a simple distribution well, and I think it might be best modeled as having a compound probability distribution. In such a distribution there are multiple steps. First there is a discrete (stepwise) distribution giving the odds of a few different distributions corresponding to different broad weather patterns, and their possible combinations.

Does this compound probability resolve to a stable distribution? Does that distribution obey a form of the central limit theorem (which we need to start a GEV derivation)? Someone more motivated than me might attempt an exact solution using a small model (perhaps three different exponentials). What I might be tempted to do is a large Monte Carlo attack on the problem, which can practically be much more complicated.

Fix the Flood Maps around Asheville!

The most direct and certain result from the work presented here is that the official flood maps for southern Asheville, NC, are on the dangerous side of wrong. The frequency of property-damaging floods is underestimated by a factor of at least two, but possibly ten. To narrow down that range, to get precise but properly conservative flood maps, will require diligent, honest, earnest effort. That effort is demanded.