

EXTREME EVENTS

Terry Peace
Hazard Impact Assessment, Ministry of Health
Victoria, BC
tbpeace@bcsc02.gov.bc.ca

ABSTRACT

Extreme events (such as floods) can be costly and dangerous. This paper examines the theory and practice of predicting n-year events from previous observations. SAS INSIGHT proves to be a useful tool.

INTRODUCTION

When looking at extreme events, the problem is usually posed more or less as follows: determine how high the water will come every twenty years, or every two hundred years, or some other return period. Unfortunately, the problem is fraught with uncertainty.

There are three major sources of uncertainty:

first uncertainty:	physical process constant?
second uncertainty:	previous data extreme in one direction?
third uncertainty:	will next n years be extreme in the other direction?

THE THEORY

Let's look at the first uncertainty:

underlying physical process	generates observable phenomena which we observe and record observations.
-----------------------------	--

What would we need to know to confidently assert that this physical process will remain unchanged? What would we need to know to make a reasonably guess? Too much. In practice, we rarely have the resources to examine this topic at all. We therefore transform the uncertainty into an assumption.

The second uncertainty is just what statistical analysis was designed for. For example, if the

data arises from a one-parameter distribution, we can simply generate a confidence interval for the parameter, and take the more conservative extreme of the interval as the estimate. This defines the distribution.

The third uncertainty is also, in theory, amenable to statistical analysis. All we need to do is to derive the appropriate order statistic from the above distribution. Then, noting that this order statistic (say P95) is also a random variable, and has a distribution, we can derive a confidence interval for P95. Choosing the most conservative end of the confidence interval, will then, in principal, result in a conservative estimate of an event with a 20 year return interval.

LIMITATIONS OF THE THEORY

That's what theory tells us. Anyone done this lately? Anyone seen this done in the literature (if, indeed, you can find it; the literature, I mean)?

No. The reasons are as follows.

1. Observations usually have poor fit to one-parameter distributions. Even two parameter distributions. Often need three parameter distributions. Even though you can, in principal, compute confidence intervals for each parameter, these do not interact in a simple way, so the only practical way to get a confidence interval over the entire distribution is by using Monte Carlo analyses (one for the possibility that the observations are atypical, one for the possibility that the future may also be atypical).
2. Providing for the possibility that BOTH the past and the future are atypical is usually unacceptable (as few of the accessible statistical procedures accommodate this; the t distribution of $SX_i/n - SY_i/m$ is a notable exception). Not only is providing for this possibility technically difficult, it renders the conclusions so extreme as to be politically indefensible, hence useless.

So what should we do?

COMMON SOLUTIONS

The first common solution is to fit a curve, compute an order statistic (like P95), and move on to the next problem.

The second common solution is to partition the data, fit curves, and take the most conservative P95.

However, neither of these solutions acknowledges the fact that we are dealing with random processes.

A SUGGESTION

Perhaps what we should be doing is explicitly acknowledging that we have random processes at work, and using a Monte Carlo analysis on the best fitting distribution. This is tantamount to assuming that either the data is typical, or the future will be typical.

Conceptually, the technique is as follows:

1. decide on return period n
2. define appropriate order statistic $P(1-1/n)$
3. fit the data
4. compute n random predictions, retaining the largest value only
5. repeat 1,000 or 10,000 times, order these retained values
6. choose the appropriately conservative figure from the list created in step 5 as the estimate of the event with an n -year return period.

In statistical practice, the technique can be refined by computing $100n$ (or $1000n$) predictions, and choosing the appropriately conservative figure ($P95$ in this case corresponds to the median of the list in step 5; $P995$ corresponds to $P90$ of the list in step 5); however, in practice, the small theoretical gain may result in political misunderstandings not easily resolved. The consultant should choose his procedure based on his client's willingness to endure statistical minutiae.

THE SAS, USING PROC INSIGHT

Start with the usual:

```
libname c 'c:\work\ad\nauseam\';  
proc insight data=c.data;  
run;
```

You then have a series of menus: click on

```
Analyze  
Distribution  
Riverlev (extreme event variable)  
Y  
Output  
Cumulative Distribution Function (so you can use K-S)  
Empirical  
OK, OK, OK
```

Then you get to see your data. Next you want to examine how well the data fits various well

understood distributions, such as Weibull and Lognormal. To do this, click on

Curves

- Parametric CDF

Weibull

OK

Play with parameter values and record them.

Then you are ready to test for a specific distribution. Click on

Curves

Test for a Specific Distribution

Fill in the parameters

OK

Record the parameter values, K-S statistic, P value.

Play with parameter values some more, look for the smallest K-S statistic, corresponding to the biggest P value. Record.

Don't get discouraged if your K-S statistics are high and your P-values low. INSIGHT has lousy resolution when you try and change parameters. I've had two parameter values (sigma), each giving me K-S statistics of about .2, $P \gg .1$. The sigma midway between gave me a K-S of .06, $P > .99$. So play around and estimate, and insert preferred parameter values from the

Curves

Test for a Specific Distribution

screens.

Try another distribution.