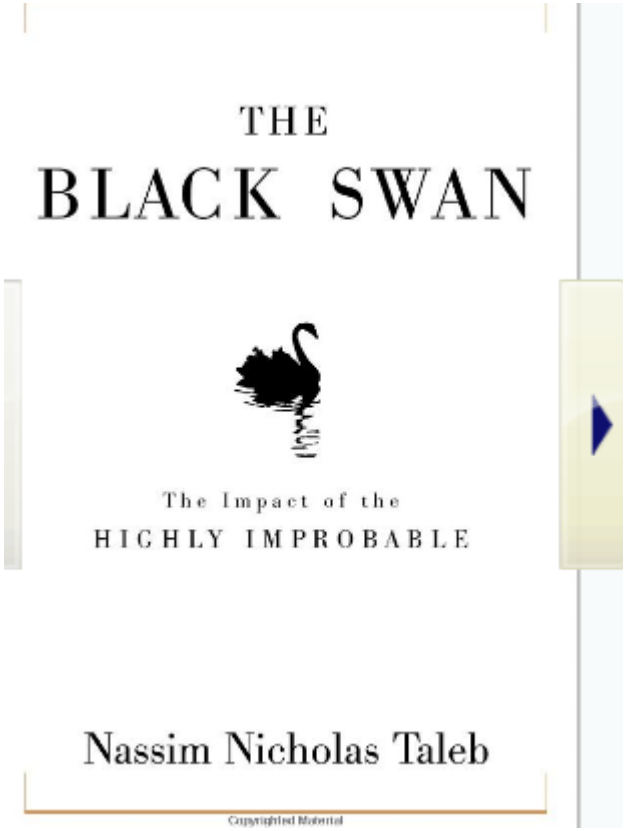
Odds of Dying

- http://www.huffingtonpost.com/tag/odds-of-dying

- NY Times review

The human mind suffers from three ailments as it comes into contact with history, what I call the *triplet of opacity*. They are:

a. the illusion of understanding, or how everyone thinks he knows what is going on in a world that is more complicated (or random) than they realize;
b. the retrospective distortion, or how we can assess matters only after the fact, as if they were in a rearview mirror (history seems clearer and more organized in history books than in empirical reality); and
c. the overvaluation of factual information and the handicap of authoritative and learned people, particularly when they create categories—when they "Platonify."



Assignments/Dates

- Chapter 3 notes due Feb 15
- http://journals.ametsoc.org/doi/pdf/10.1175/BAMS-88-6-853 Cerveny Article
- Amazon Drought paper review Due Feb 17
- Odds Are It's Wrong. Due Feb 22
  - http://www.sciencenews.org/view/feature/id/57091/title/Odds_Are,_Its_Wrong
  - Read and summarize issues about significance testing in a few paragraphs
- Chapter 4 notes due Feb 24
- Exam March 1.

Daily diet soda tied to higher heart attack risk - Health - Diet and nutrition - msnbc.com - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.msnbc.msn.com/id/41479869/ns/health-diet_and_nutrition/

diet coke stroke

Daily diet soda tied to higher heart at...

Daily diet soda tied to higher risk for stroke, heart attack

61 percent higher risk of vascular events for those who drank diet soda each day, study finds

Jump to text
Just as you were starting to feel virtuous for having switched from sug...

Below: Text Discussion Related

Tweet 380

Recommend 8K

Open in new wind

TODAY

Feb. 10

A new study indicated that diet soda drinkers are more likely to suffer from vascular complications. NBC's chief medical editor Dr. Nancy Snyderman discusses the dangers these drinks

Related video

Diet sodas may be hazardous to your health

Just as you were starting to feel virtuous for having switched from sugary sodas to low- or no-calorie substitutes, a new study comes along suggesting that diet sodas might be bad for your head and your heart.

The study, which followed more than 2,500 New Yorkers for nine or more years, found that people who drank diet soda every day had a 61 percent higher risk of vascular events, including stroke and heart attack, than those who completely eschewed the diet drinks, according to researchers who presented their results today at the American Stroke Association's International Stroke Conference in Los Angeles.

Still, the researchers aren't ready to tell consumers to skip diet sodas. More studies need to be done before that happens, said the report's lead author Hannah Gardener, an epidemiologist at the University of Miami Miller School of Medicine.

"I think diet soda drinkers need to stay tuned," Gardener said. "I don't think that anyone should be changing their behaviors based on one study. Hopefully this will motivate other researchers to do more studies."

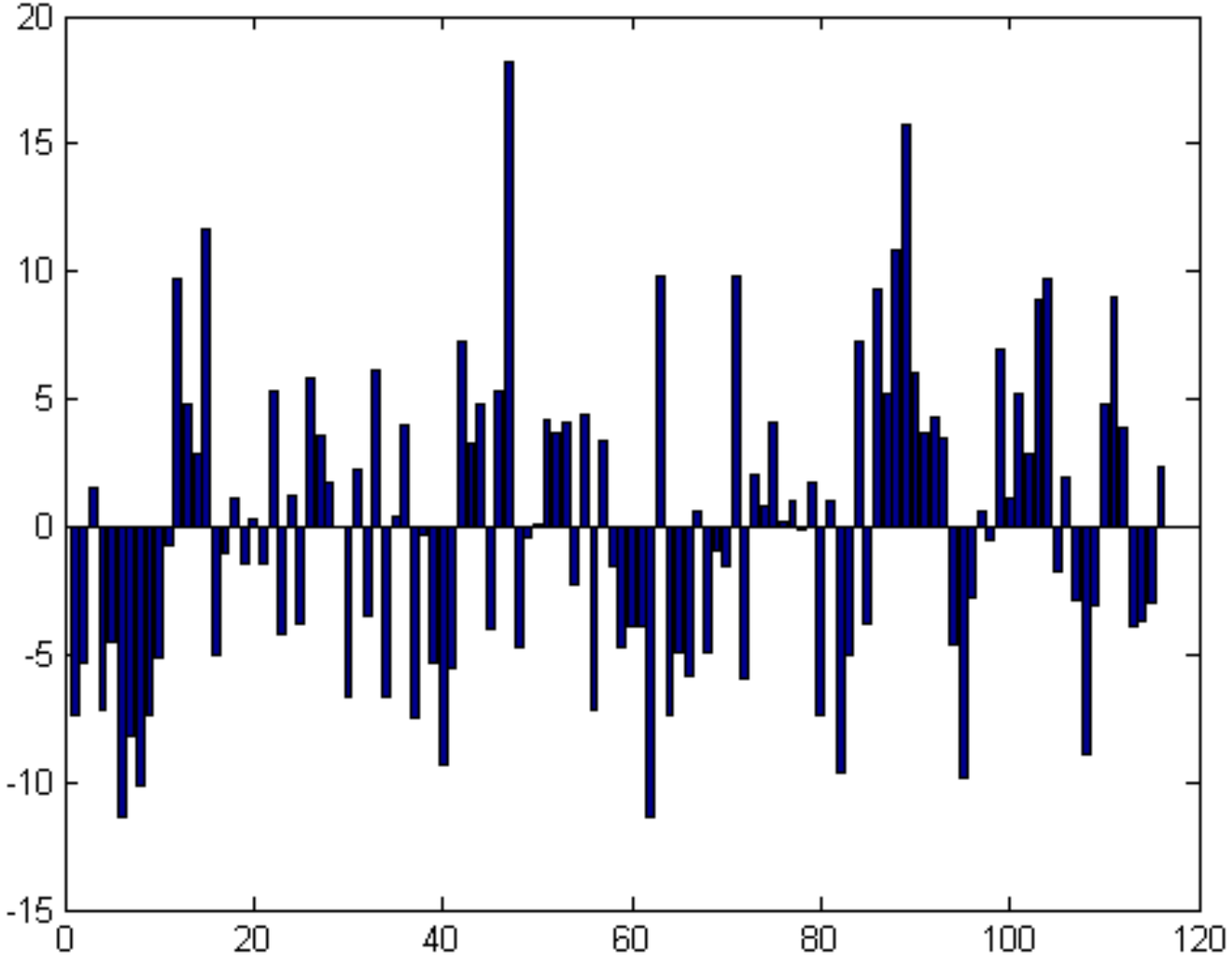
For the new study, researchers surveyed 2,564 north Manhattan residents about their eating behaviors, exercise habits, as well as cigarette and alcohol consumption. The study volunteers were also given physical check-ups that included blood pressure measurements and blood tests for cholesterol and other factors that might affect the risk for heart attack and stroke.

The increased likelihood of vascular events remained even after Gardener and her colleagues accounted for risk factors, such as smoking, high blood pressure and high cholesterol levels. Pointing the finger more squarely at diet drinks, the researchers found no increased risk among people who drank regular soda.

Does this mean there's something in diet sodas that hurts our blood vessels? Nobody knows the answer to that question, yet, Gardener said. There could be

Annual Precipitation of Utah

- Have we been in a drought the past 5 years?
- Define drought as average precipitation over 5 year period substantively below 0
- Values: 3.9, -3.9,-3.8,-2.9,2.4
- 5 yr mean -0.9, sd = 3.9
- 116 year sample mean 0, sd = 5.8


20
15
10
5
0
-5
-10
-15
0
20
40
60
80
100
120

Steps of Hypothesis Testing

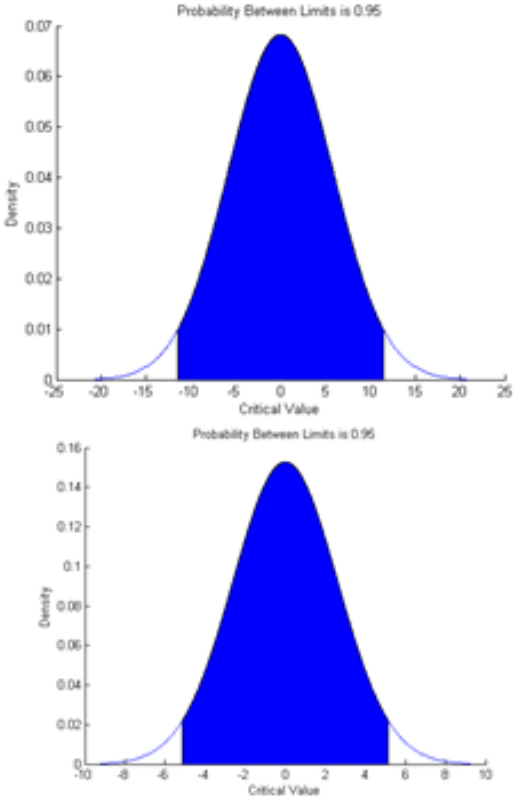
- Identify a test statistic that is appropriate to the data and question at hand
  - Computed from sample data values. 5 yr sample mean -0.9
- Define a null hypothesis, H_0 to be rejected
  - 5 yr sample mean 0
- Define an alternative hypothesis, H_A
  - 5 yr sample mean < 0
- Estimate the null distribution
  - Sampling distribution of the test statistic IF the null hypothesis were true
  - Making assumptions about which parametric distribution to use (Gaussian, Weibull, etc.)
  - Use sample mean of 0 and 116 yr sd of 5.8
- Compare the observed test statistic (-0.9) to the null distribution. Either
  - Null hypothesis is rejected as too unlikely to have been true IF the test statistic fall in an improbable region of the null distribution
    - Possibility that the test statistics has that particular value in the null distribution is small
    - normspec([-1.96*5.8,1.96*5.8],0,5.8)
    - OR
    - The null hypothesis is not rejected since the test statistic falls within the values that are relatively common to the null distribution

Caution!

- NOT rejecting the null hypothesis is not the same as saying the null hypothesis is true
  - There is insufficient evidence to reject H_0
- H_0 is rejected if the probability p of the observed test statistic is ≤ α significance or rejection level
- If odds of test statistic occurring in the null distribution less than 1 or 5%, then we may choose to reject the null hypothesis
- Rejecting the null hypothesis MAY be same as accepting alternative hypothesis BUT there may be many other possible alternative hypotheses
- You must define ahead of time the α significance or rejection level
  - 1% or 5%, 1 in 100 or 5 in 100 chance that you accept the risk of rejecting the null hypothesis incorrectly
  - Type 1 category error of a false rejection of the null hypothesis

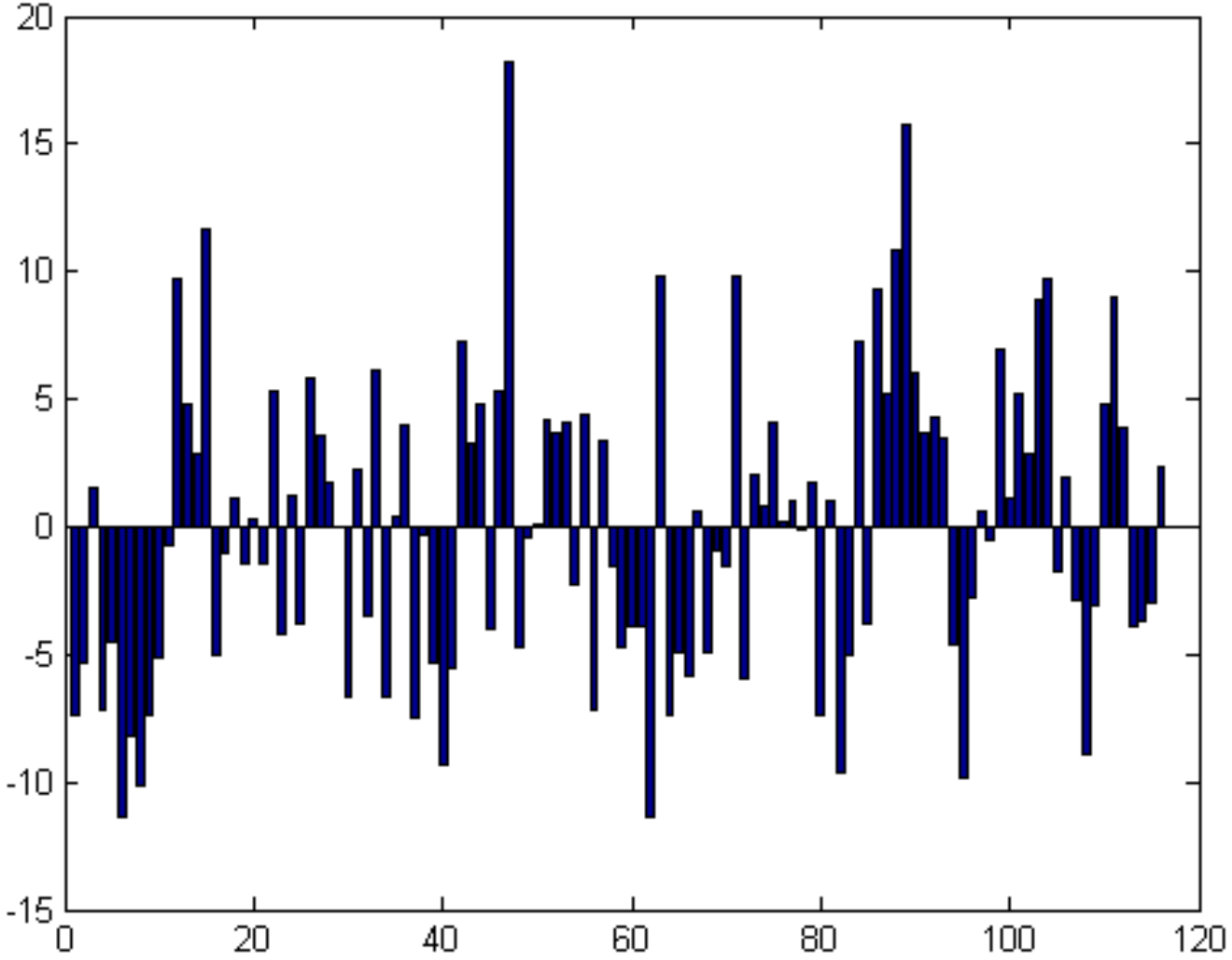
Central Limit Theorem

- Sum (or mean) of a sample (5 dice) will have a Gaussian distribution even if the original distribution (1 die) does not have a Gaussian distribution, especially as the sample size increases
- $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$
- Individually 116 yr sd = 5.8 cm
- 5 yr samples sd = 5.8/sqrt(5)

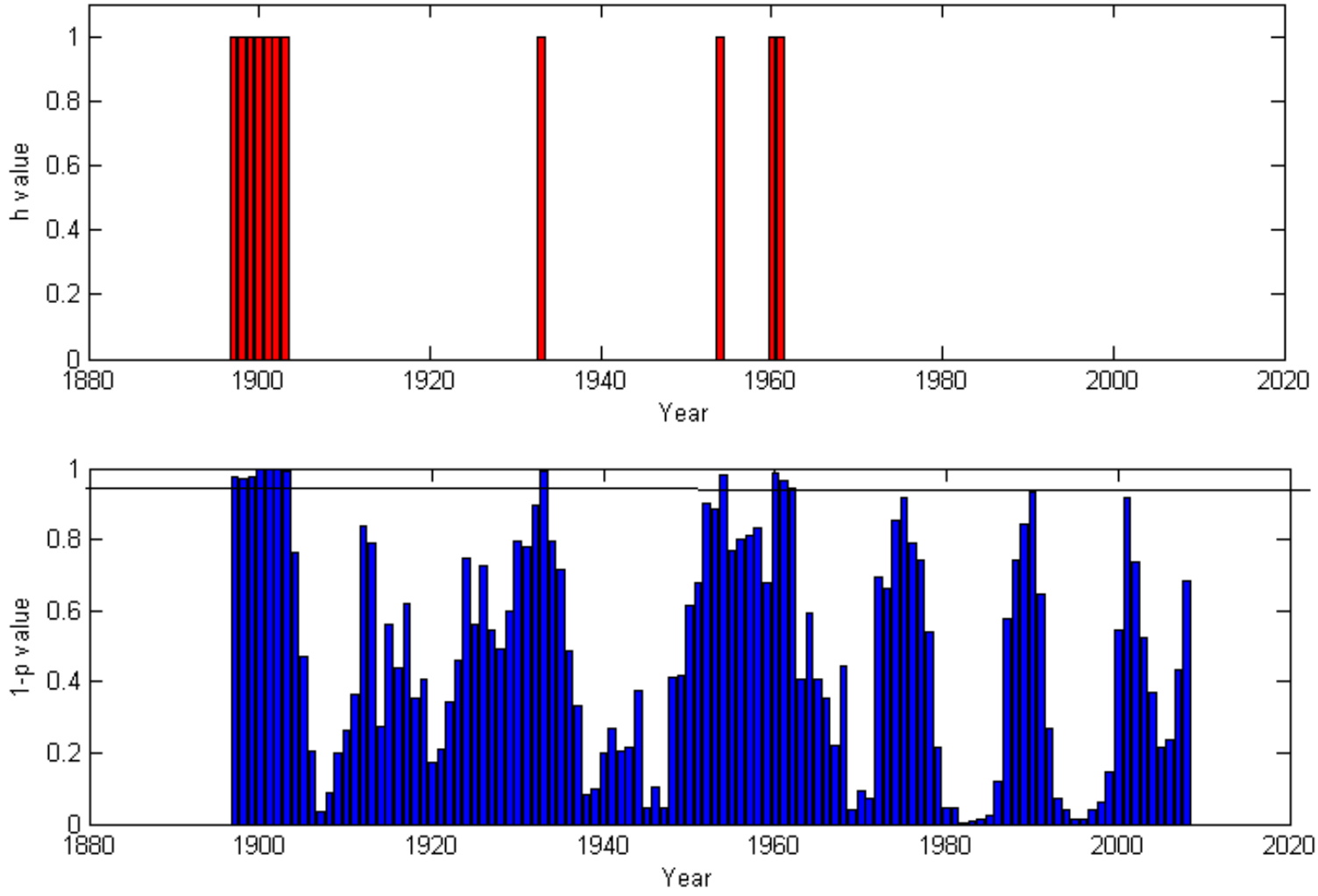


Students' t test

- $\sigma_{\overline{X}} = \frac{s_x}{\sqrt{n-1}}$
- Estimate of population variance from sample
- T value:
- Numerator: signal $t = (\bar{x} - \mu)\sqrt{n-1} / s_x$
- Denominator: noise
- At t gets larger, confidence in rejecting the null hypothesis (sample mean differs from population mean) gets higher
- T large IF:
  - Spread between sample and population means large
  - Degrees of freedom is large
  - Variability in sample is small


20
15
10
5
0
-5
-10
-15
0
20
40
60
80
100
120

Which 5-yr samples would be considered a drought?



Summary

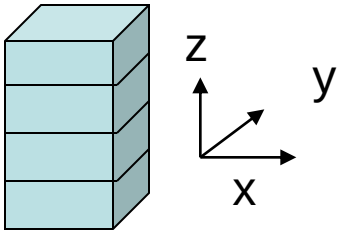
- Research involves defining a testable hypothesis and demonstrating that any statistical test of that hypothesis meets basic standards
- Typical failings of many studies include:
  - (1) ignoring serial correlation in environmental time series that reduces the estimates of the number of degrees of freedom and
  - (2) ignoring spatial correlation in environmental fields that increases the number of trials that are being determined simultaneously.
    - Inflates the opportunities for the null hypothesis to be rejected falsely.
- Use common sense
- Be very conservative in estimating the degrees of freedom temporally and spatially
- Avoid attributing confidence to a desired result when similar relationships are showing up far removed from your area of interest for no obvious reason
- The best methods for testing a hypothesis rely heavily on independent evaluation using additional data not used in the original statitiscal analysis

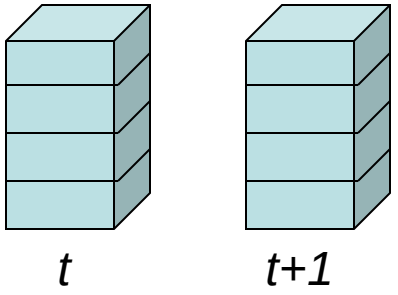
4. Exploratory Multivariate Data Analysis

- want to relate how one or more phenomena are related to others
- Think of it as a multidimensional problem and the goal is to reduce the dimensionality through exploration of relationships
- Dealing with the dimensionality of environmental data sets in statistical analyses is of general concern (Murphy 1991; Mon. Wea. Rev., 1590-1601).
- Exploratory multivariate data analysis encompasses an array of tools to assess relationships between two or more samples

Large Dimensionality of Geophysical Data Sets

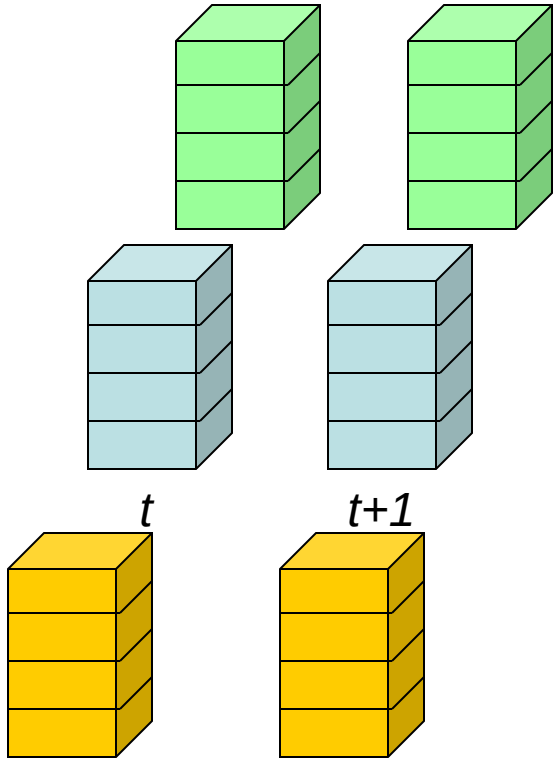
•Space: x, y, z



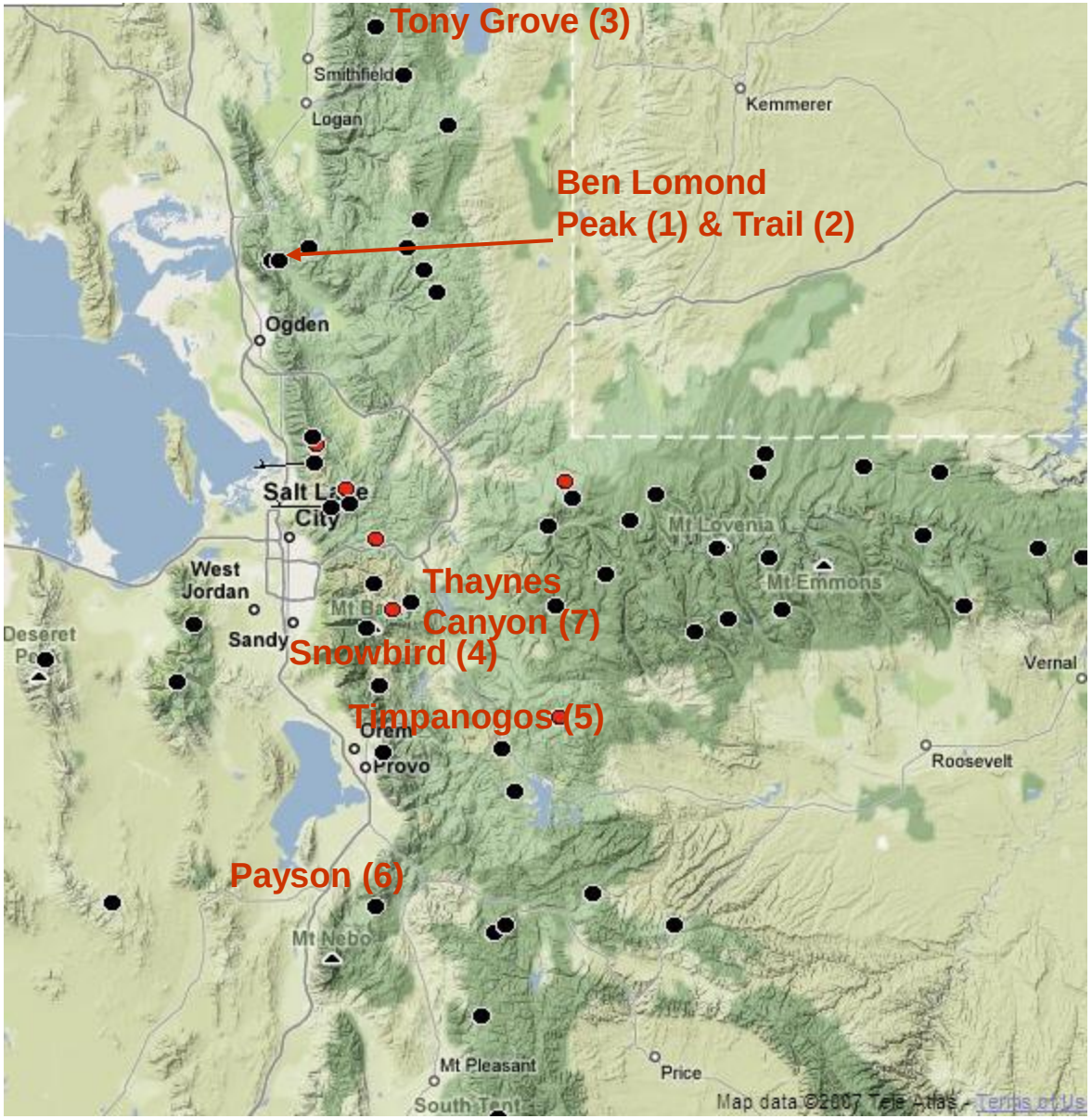


•Time: time (t) and forecast time (t_f)

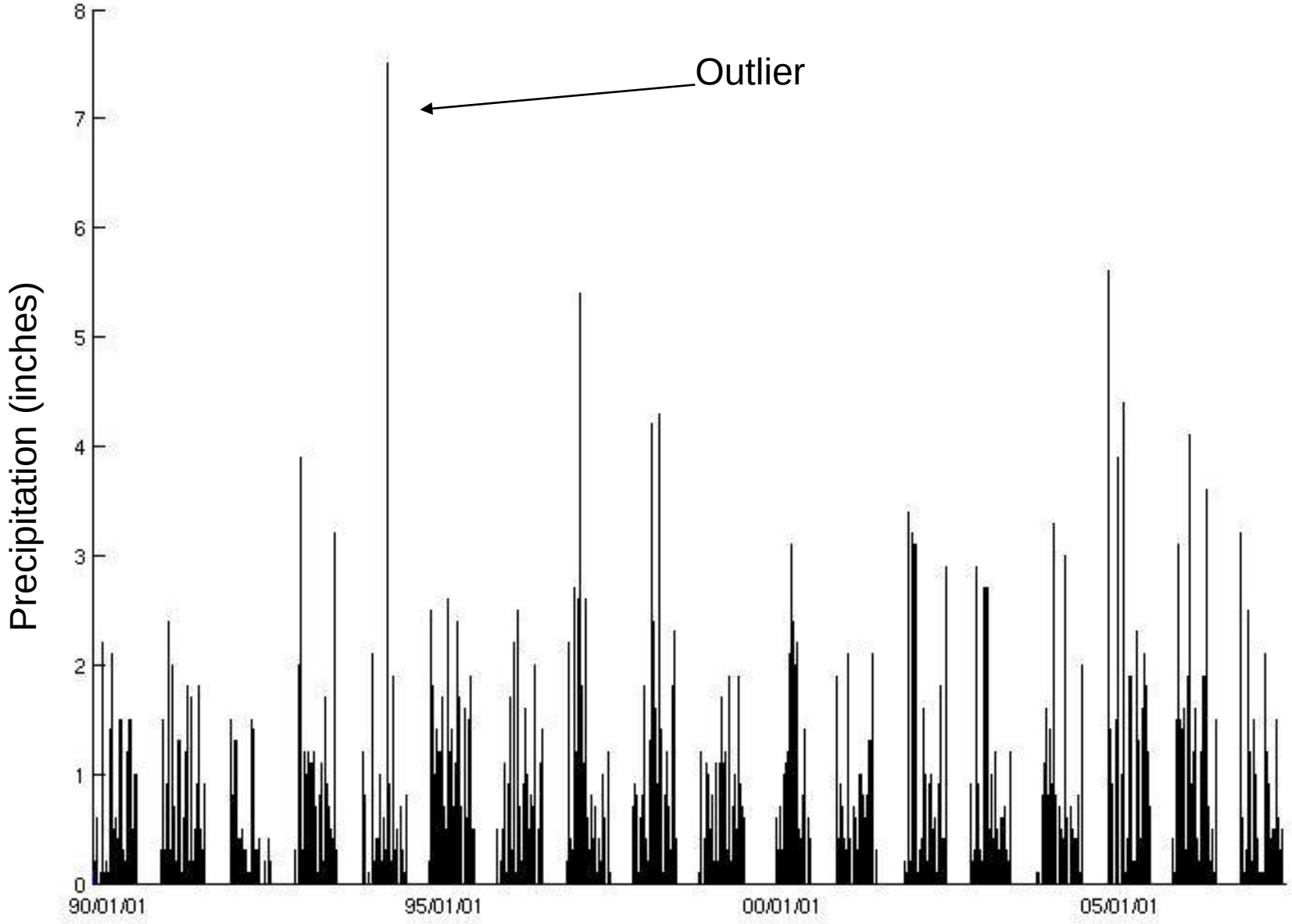
•Parameter & Source (temperature, winds, different models, measuring systems, perturbations)

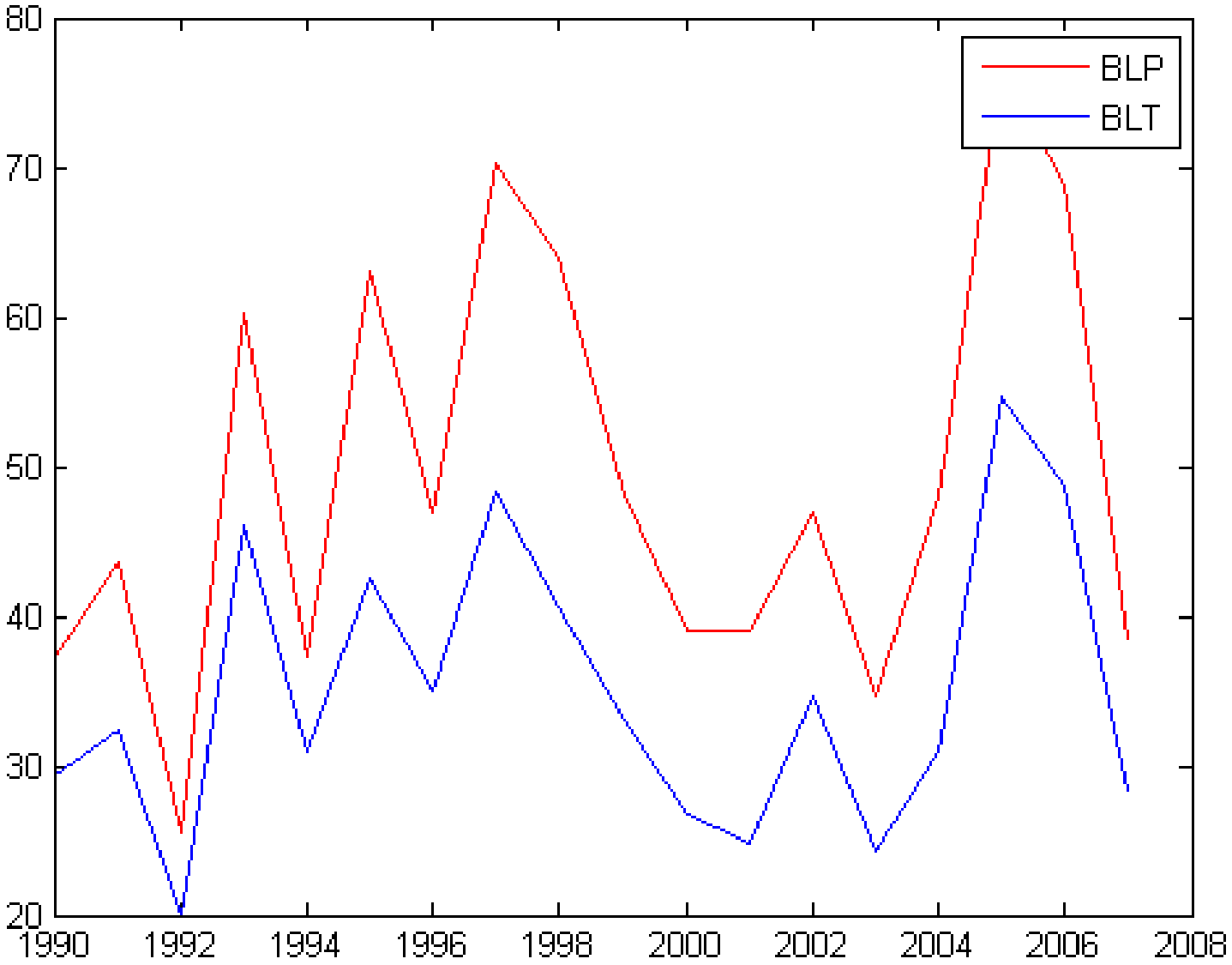


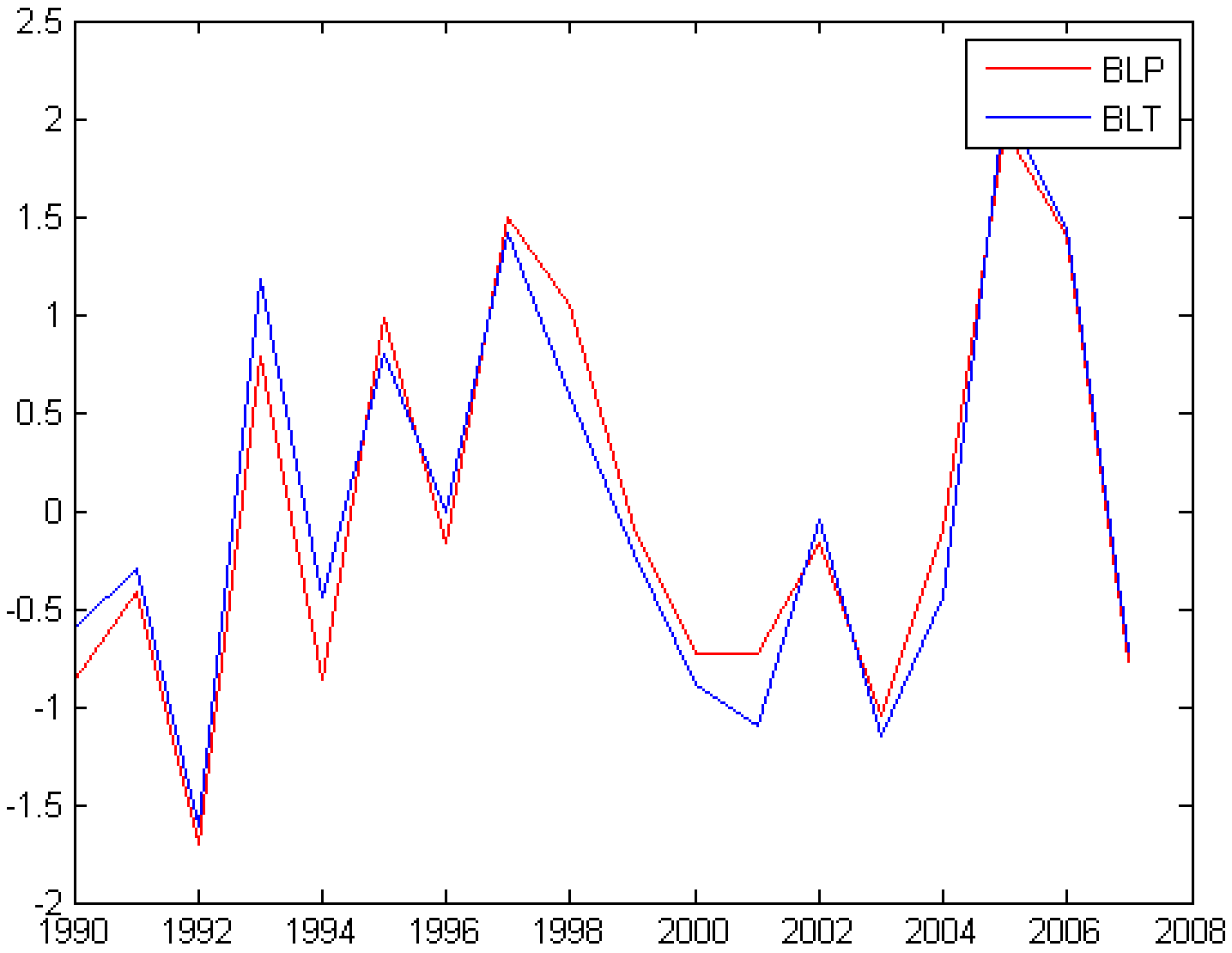
SNOTEL Sites

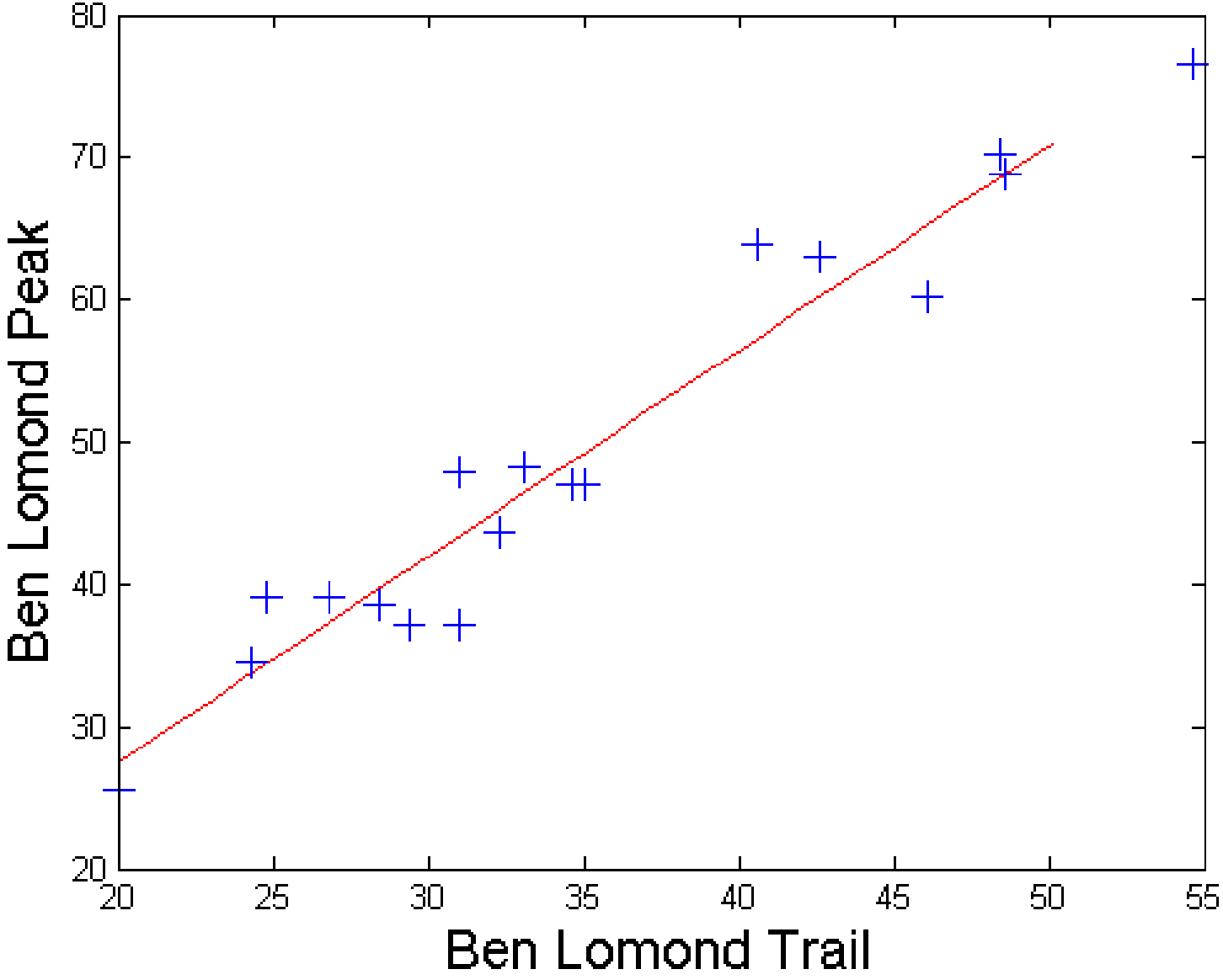


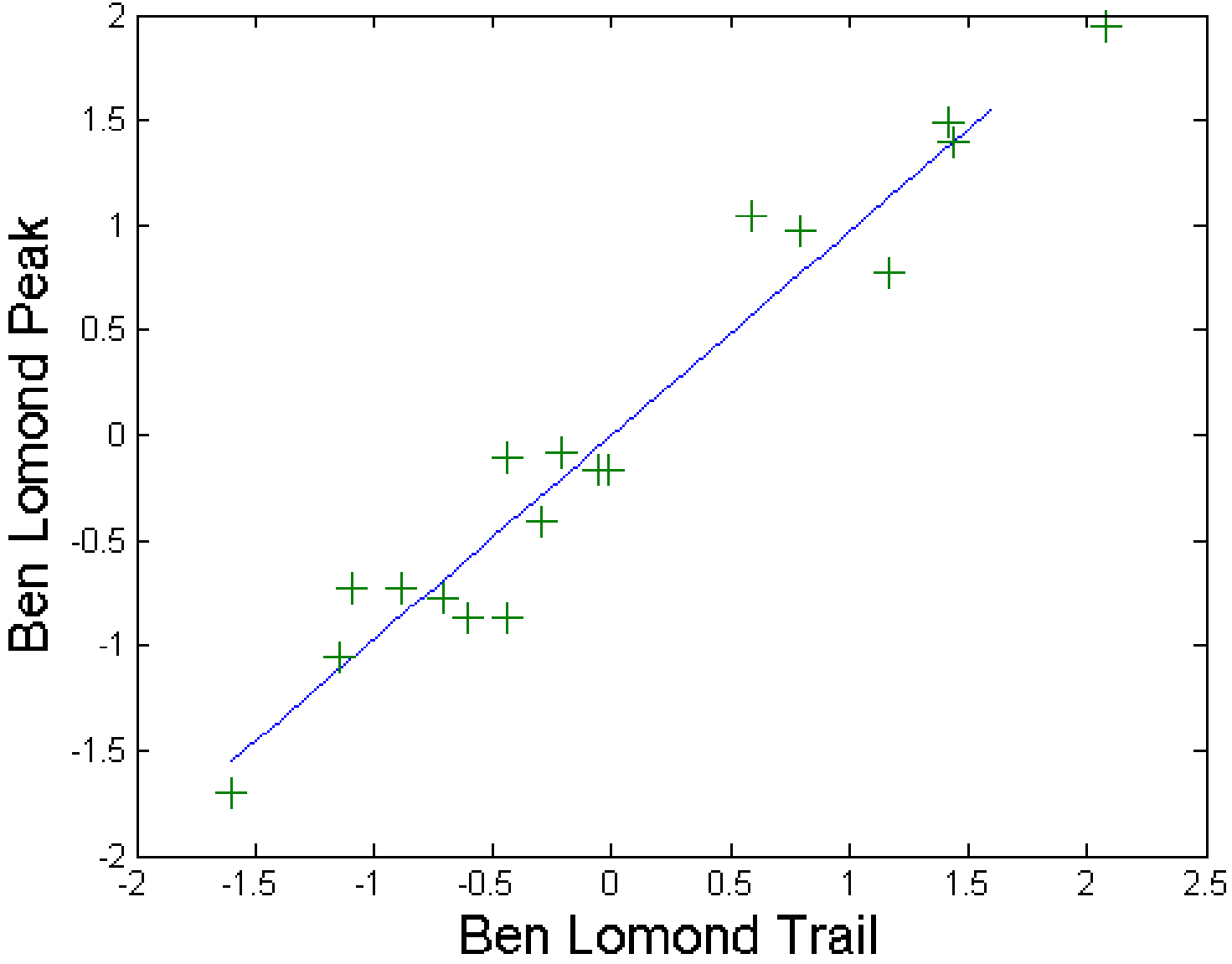
Ben Lomond Daily Precipitation




BLP
BLT
80
70
60
50
40
30
20
1990
1992
1994
1996
1998
2000
2002
2004
2006
2008


BLP
BLT
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
1990
1992
1994
1996
1998
2000
2002
2004
2006
2008


Ben Lomond Peak
Ben Lomond Trail
80
70
60
50
40
30
20
20
25
30
35
40
45
50
55


Ben Lomond Peak
Ben Lomond Trail
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5

Estimating Values of One Variable From Another

- X- Ben Lomond Trail
- Y- Ben Lomond Peak
- Want to estimate Peak from Trail
- Use pairs of observations from sample
- Need to determine coefficient b or r
- b- slope of linear estimate
- r- linear correlation

$$\hat{y}_i = \bar{y} + b(\hat{x}_i - \bar{x})$$

$$\hat{y}^*_i = r\hat{x}^*_i$$

Be Careful…

- Avoid considering X as “predictor” and Y as “predictand”
- Estimate Y from X
- X is independent variable
- Y is dependent variable