# EGRET: Exploration and Graphics for RivEr Trends

**An R-package for the analysis of long-term changes in river-water quality and streamflow, including the water quality method "Weighted Regressions on Time, Discharge, and Season" (WRTDS)**

**Developed by Bob Hirsch & Laura De Cicco:**

**In Beta testing**

USGS
science for a changing world

# The driving philosophy for EGRET:

**“The only way to figure out what is happening to our planet is to measure it,**

**and this means tracking changes decade after decade**

**and poring over the records.”**

*(Ralph Keeling, Science Magazine, 2008)*

# Outline

- Motivation
- Basics of the EGRET package
- Working some examples
- Thoughts about climate change and streamflow
- Final examples of other change drivers

# Motivations

- Interest in hydrologic change
- Make it easy to get at our daily values & organize them for viewing
- People look to the USGS for flow statistics, we need tools to see changes (or no changes)

# Potential change drivers:

- Urbanization
- Agricultural land drainage systems
- Irrigation
- Groundwater depletion
- Exports and consumptive use
- Reservoir regulation
- Quasi-periodic climatic oscillations
- Enhanced greenhouse forcing

# Quasi-periodic oscillations:

- ENSO
- PDO
- AMO
- ??O
- Ice Ages

# The climate exhibits long-term persistence at many time scales

- What looks like a trend at one time scale may be a part of a much longer oscillation.
- Hypothesis tests won’t protect you from this confusion.
- Using long records can help.
- “Cycle hunting” is a fool’s errand, in my opinion.

**EGRET is a package, written in R. Totally portable (Windows, Macintosh, Linux).**

**Currently available in Beta-test, with a manual.**
**https://github.com/USGS-CIDA/WRTDS/wiki**
**This presentation is located there as well.**

**Not a formal R package yet, but it is "built" and provides on-line help pages for every function.**

**You will need a basic knowledge of R (which is very similar to S+).**

# EGRET overview

- Ingests water quality sample data, daily streamflow data, and meta data from **USGS Web services** or from **user-supplied files**
- Sub-systems: **WRTDS** for river water quality data analysis and **flowHistory** for analysis of streamflow alone
- Goal: Exploration of the data to describe the evolving hydrologic system. Produce: graphs, summary statistics, understanding, and hypotheses.

# EGRET flexibility

- Data from NWIS Web services or spreadsheet
- Choice of units: cfs, cms, or mm/day
- Stores and uses meta-data
- Can look at water year, calendar year, user-defined seasons, or individual months
- Graphs & tables are self-labeled and self-scaled
- Data structure allows user to run their own analysis using standard R functions
- Graphics ready for PowerPoint or publications
- Stores data in a standard format for future use

# https://github.com/USGS-CIDA/WRTDS/wiki

- At the downloads page you will see something like this:

dataRetrieval_1.0.3.tar.gz — Latest dataRetrieval package instal
136KB · Uploaded 8 days ago

EGRET manual_3.doc — EGRET Manual. Updated May 24 with
7.0MB · Uploaded 14 days ago

WRTDS NWQMC course.pptx — Slightly edited version of 1May
23.4MB · Uploaded a month ago

EGRET_1.1.3.tar.gz — Improved formatting of help descriptions
512KB · Uploaded a month ago

Download the tar.gz files, start R, and then follow the instructions for installation given on the wiki

```
> sta<-"05474500"
> param<-"00060"

> Daily<-getDVData(sta,param,StartDate="1879-10-01",EndDate="2011-09-30")
```

There are  48212 data points, and  48212 days.
There are  0 zero flow days
If there are any zero discharge days, all days had 0 cubic meters per second added to the discharge value.

```
> sta<-"05474500"
> param<-"00060"
> Daily<-getDVData(sta,param,StartDate="1879-10-01",EndDate="2011-09-30")
```

There are  48212 data points, and  48212 days.

There are  0 zero flow days

If there are any zero discharge days, all days had 0 cubic meters per second added to the discharge value.

```
> summary(Daily)
      Date                  Q                 Julian           Month             Day            DecYear    
 Min.   :1879-10-01   Min.   :   141.6   Min.   :10865   Min.   : 1.000   Min.   :  1.0   Min.   :1880  
 1st Qu.:1912-09-30   1st Qu.:   926.0   1st Qu.:22918   1st Qu.: 4.000   1st Qu.: 92.0   1st Qu.:1913  
 Median :1945-09-30   Median : 1475.3    Median :34970   Median : 7.000   Median :183.0   Median :1946  
 Mean   :1945-09-30   Mean   : 1919.8    Mean   :34970   Mean   : 6.523   Mean   :183.1   Mean   :1946  
 3rd Qu.:1978-09-30   3rd Qu.: 2520.2    3rd Qu.:47023   3rd Qu.:10.000   3rd Qu.:274.0   3rd Qu.:1979  
 Max.   :2011-09-30   Max.   :12402.8    Max.   :59076   Max.   :12.000   Max.   :366.0   Max.   :2012  

    MonthSeg        Qualifier               i               LogQ             Q7                Q30        
 Min.   : 358.0   Length:48212       Min.   :    1   Min.   :4.953   Min.   :  234.2   Min.   :  345.7  
 1st Qu.: 753.8   Class :character   1st Qu.:12054   1st Qu.:6.831   1st Qu.:  935.3   1st Qu.:  975.6  
 Median :1149.5   Mode  :character   Median :24106   Median :7.297   Median : 1482.2   Median : 1536.7  
 Mean   :1149.5                      Mean   :24106   Mean   :7.338   Mean   : 1920.0   Mean   : 1920.4  
 3rd Qu.:1545.2                      3rd Qu.:36159   3rd Qu.:7.832   3rd Qu.: 2521.4   3rd Qu.: 2519.1  
 Max.   :1941.0                      Max.   :48212   Max.   :9.426   Max.   :11856.7   Max.   :10986.9  
                                                                     NA's   :6         NA's   :29       
```

# Reasons for the meta-data:

- Knowing what data you have and where they came from

- Putting labels on figures and tables

- Putting a name on your saved workspaces, using abbreviations

>INFO <- getMetaData(sta,param)

Your site for streamflow data is 05474500 .

Your site name is Mississippi River at Keokuk, IA ,but you can modify this to a short name in a style you prefer.

This name will be used to label graphs and tables.
If you want the program to use the name given above, just do a carriage return, otherwise enter the preferred short name(no quotes):

**The latitude and longitude of the site are: 40.39366 , -91.37432 (degrees north and west).**

**The drainage area at this site is 119000 square miles which is being stored as 308208.6 square kilometers.**

**It is helpful to set up a station abbreviation when doing multi-site studies, enter a unique id (three or four characters should work).**

**It is case sensitive. Even if you don't feel you need an abbreviation for your site you need to enter something (no quotes):**

**Keok**

Your water quality data are for parameter number **00060** which has the name:'

**Discharge, cubic feet per second** '.

Typically you will want a shorter name to be used in graphs and tables. The suggested short name is:' Stream flow, mean. daily '.

If you would like to change the short name, enter it here, otherwise just hit enter (no quotes):

Discharge

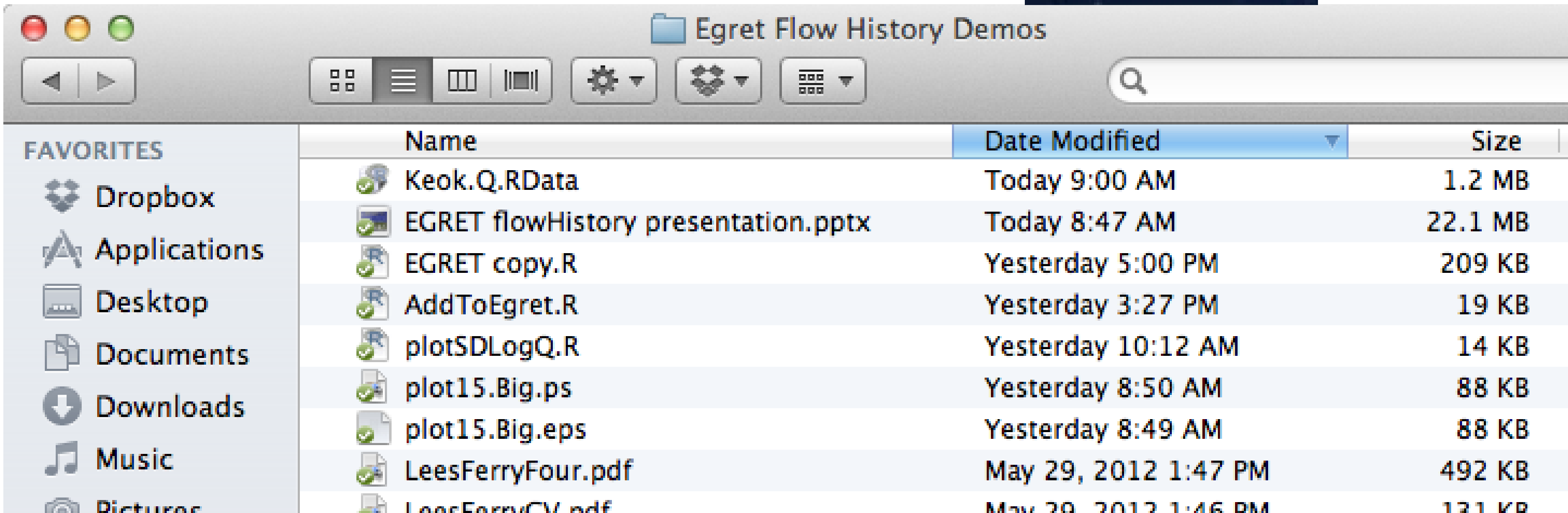
The units for the water quality data are:  cfs .

It is helpful to set up a constituent abbreviation when doing multi-constituent studies, enter a unique id (three or four characters should work something like tn or tp or NO3). It is case sensitive.  Even if you don't feel you need an abbreviation you need to enter something (no quotes):

**Q**

# Storing the workspace

```
> savePath <- "/Users/rmhirsch49/Dropbox/Egret Flow History Demos/”

> saveResults(savePath)
```

# What’s in the workspace?

R Workspace Browser

| Object | Type | Structure |
|---|---|---|
| Daily | data.frame | dim: 48212 12 |
| INFO | data.frame | dim: 1 42 |
| param | character | length: 1 |
| savePath | character | length: 1 |
| sta | character | length: 1 |

# What’s in the workspace?

R Workspace Browser

| Object | Type | Structure |
|---|---|---|
| Daily | data.frame | dim: 48212 12 |
| Date | Date | length: 48212 |
| Q | numeric | length: 48212 |
| Julian | numeric | length: 48212 |
| Month | numeric | length: 48212 |
| Day | numeric | length: 48212 |
| DecYear | numeric | length: 48212 |
| MonthSeq | numeric | length: 48212 |
| Qualifier | character | length: 48212 |
| i | numeric | length: 48212 |
| LogQ | numeric | length: 48212 |
| Q7 | numeric | length: 48212 |
| Q30 | numeric | length: 48212 |
| INFO | data.frame | dim: 1 42 |
| param | character | length: 1 |
| savePath | character | length: 1 |
| sta | character | length: 1 |

# Setting up the “Period of Analysis”

- Could be water year
- Could be calendar year
- Could be April-May-June
- Could be Dec-Jan-Feb-Mar
- Could be only May

paStart = calendar month that starts the Period
paLong = length of Period, in months

# Setting up the "Period of Analysis"

| | paStart | paLong |
|---|---|---|
| • Could be water year | 10 | 12 |
| • Could be calendar year | 1 | 12 |
| • Could be April-May-June | 4 | 3 |
| • Could be Dec-Jan-Feb-Mar | 12 | 4 |
| • Could be only May | 5 | 1 |

paStart = calendar month that starts the Period

paLong = length of Period, in months

```
> INFO <- setPA()
> annualSeries <- makeAnnualSeries()
```

If I wanted to look only at April, May, June then I would give the commands:

```
> INFO <- setPA(paStart = 4, paLong = 3)
> annualSeries <- makeAnnualSeries()
```

But for now, lets stick with the default:
paStart = 10, paLong = 12 which is the Water Year

I now have a data frame called "annualSeries" what is in it?

It is a matrix of dimensions [3,8,134]

The 3 values of the first dimension are:

1 = Decimal Year for the midpoint of the period
2 = The annual value of the particular flow statistic
3 = The smoothed value of the particular flow statistic

The 8 values of the second dimension are flow statistics:

* 1 = Annual 1-day minimum flow
* 2 = Annual 7-day minimum flow
* 3 = Annual 30-day minimum flow
  4 = Annual median daily flow
  5 = Annual mean daily flow
  6 = Annual 30-day maximum flow
  7 = Annual 7-day maximum flow
  8 = Annual 1-day maximum flow

If we use water year then these low-flow statistics are computed on a climatic year basis (April-March)

**The third dimension is the index of years (set by the size of the data set), in this case 134 years:**

**Note that it can contain “NA” values (“not available”)**

# What is the smoothing method?

**It is my own variation on LOWESS**

**I do a locally weighted scatterplot smooth on the relationship of log(Q) vs. year using a Tukey tri-cubed weight function with a half-window width of 30 years.**

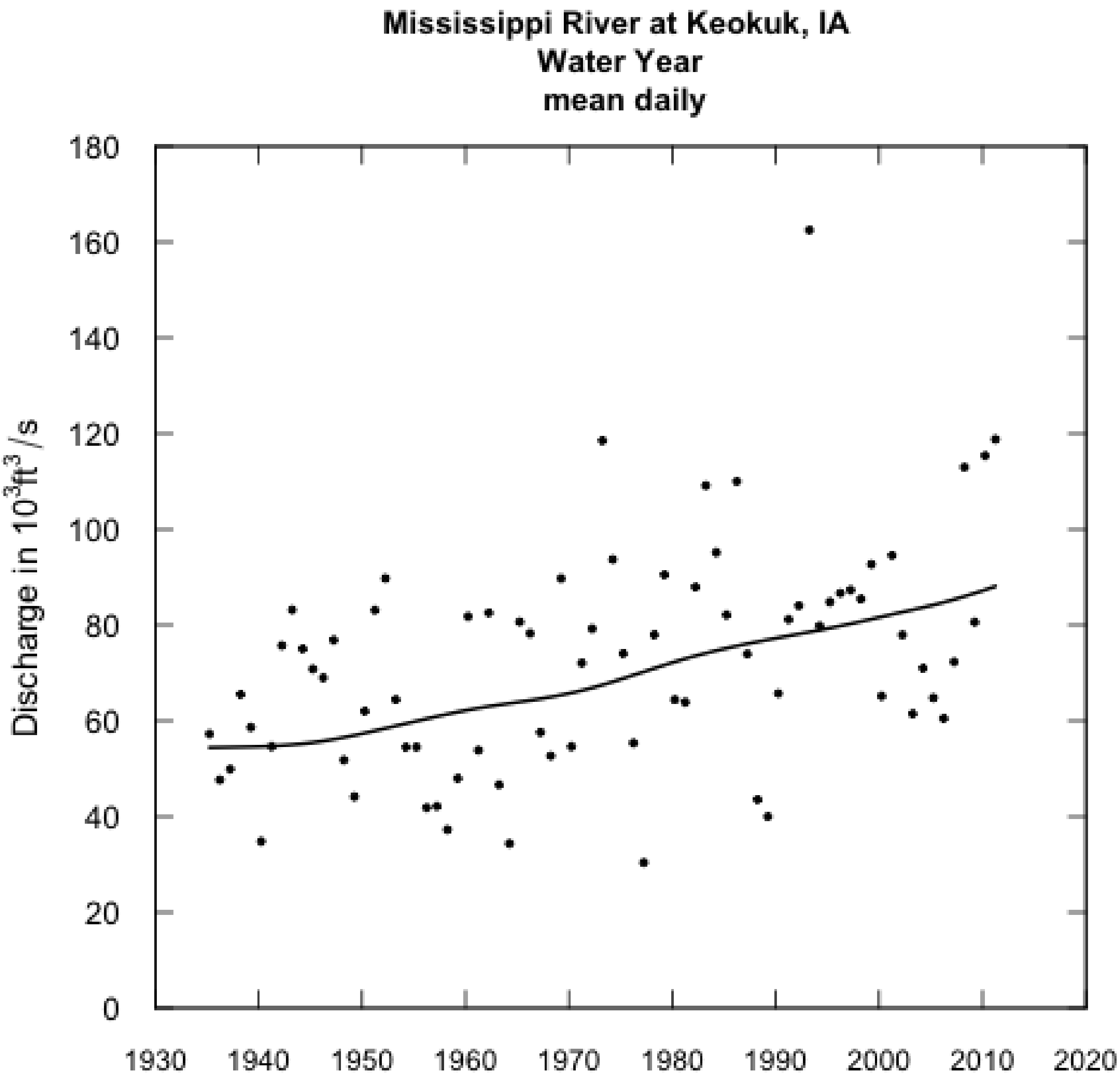
**Then I retransform it back to real space. Think of it as a median smooth (not a mean smooth).**

**By design, it is very smooth because I use logs and because of the very wide window. But, the window width can be changed at the option of the user.**

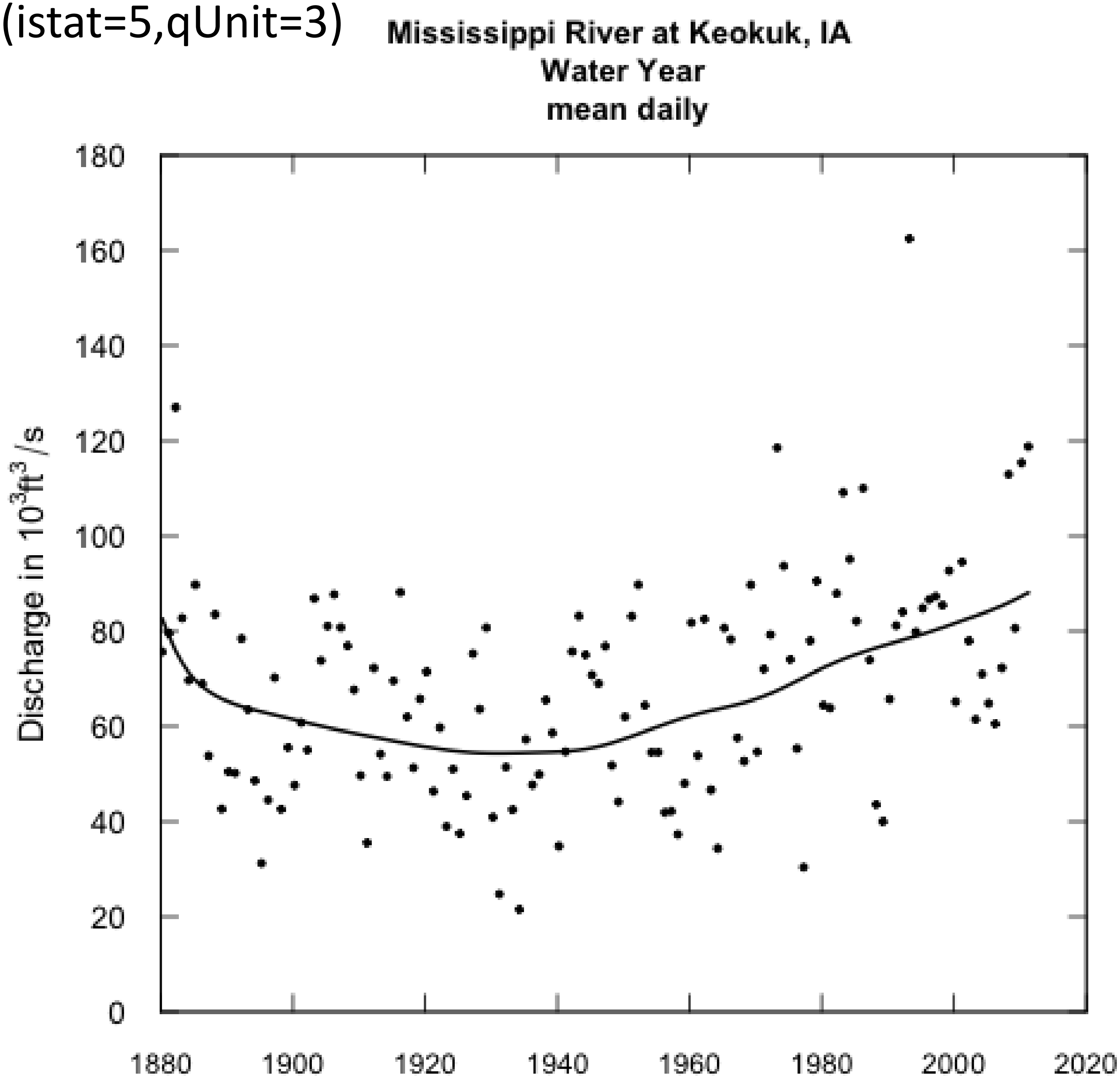
# Enough of the mechanics, lets get back to exploratory hydrology, here's the first command

```
> plotFlowSingle(istat = 5 , yearStart = 1935 , qUnit = 3)
```

```
> plotFlowSingle(istat=5,yearStart=1935,qUnit=3)
```

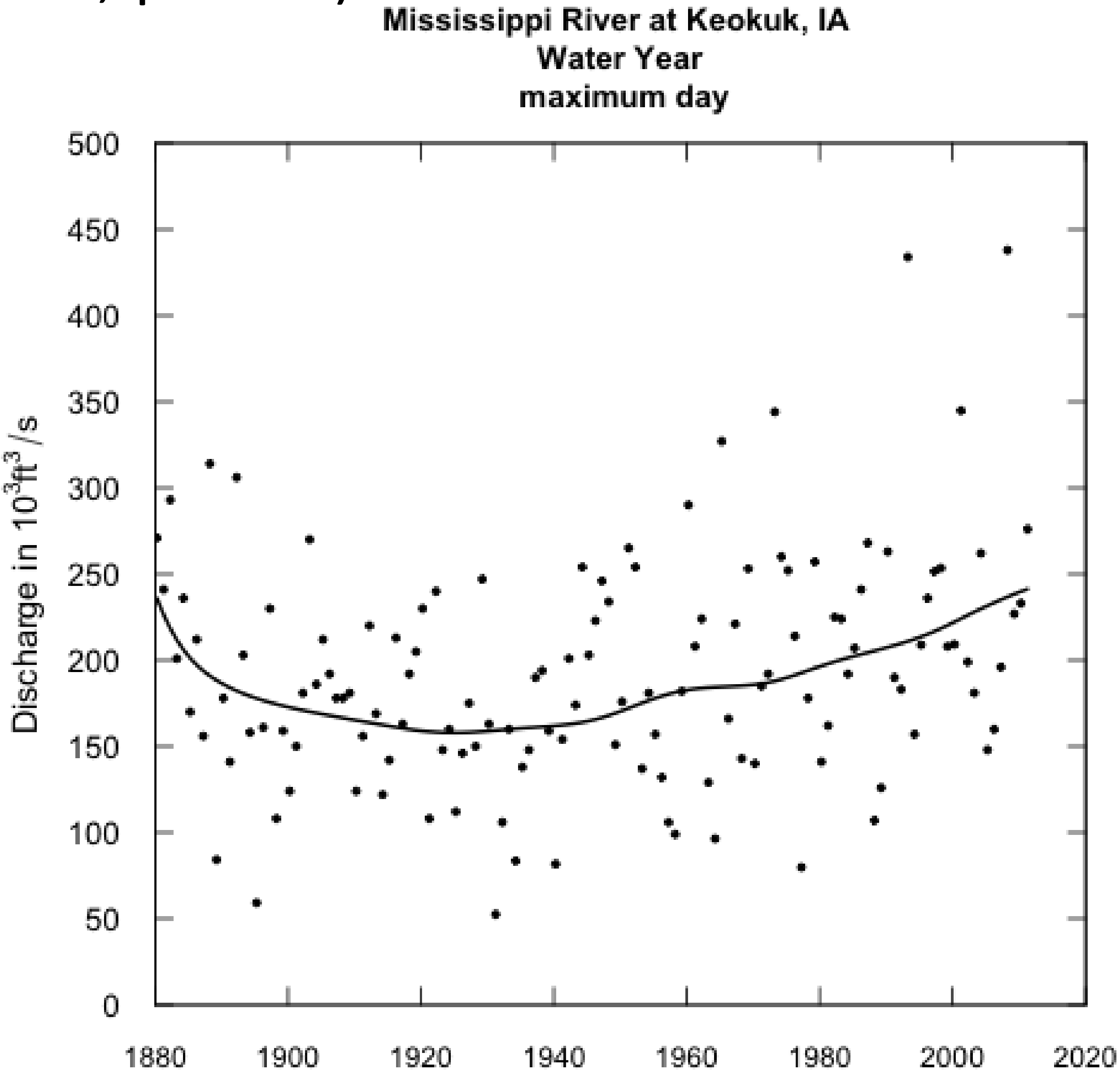


> plotFlowSingle(istat=5,qUnit=3)



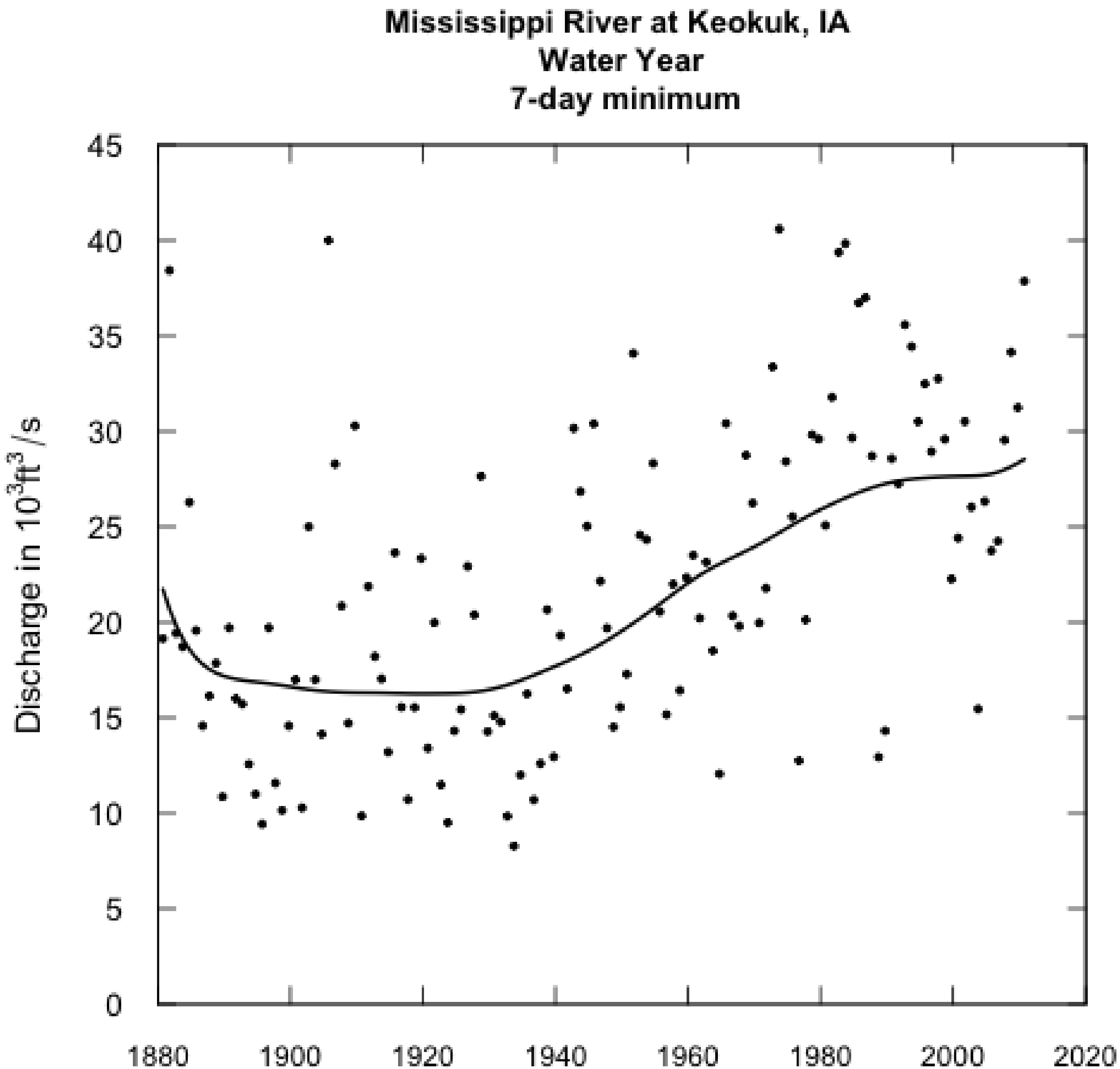
Lets look at floods

> plotFlowSingle(istat=8,qUnit=3)

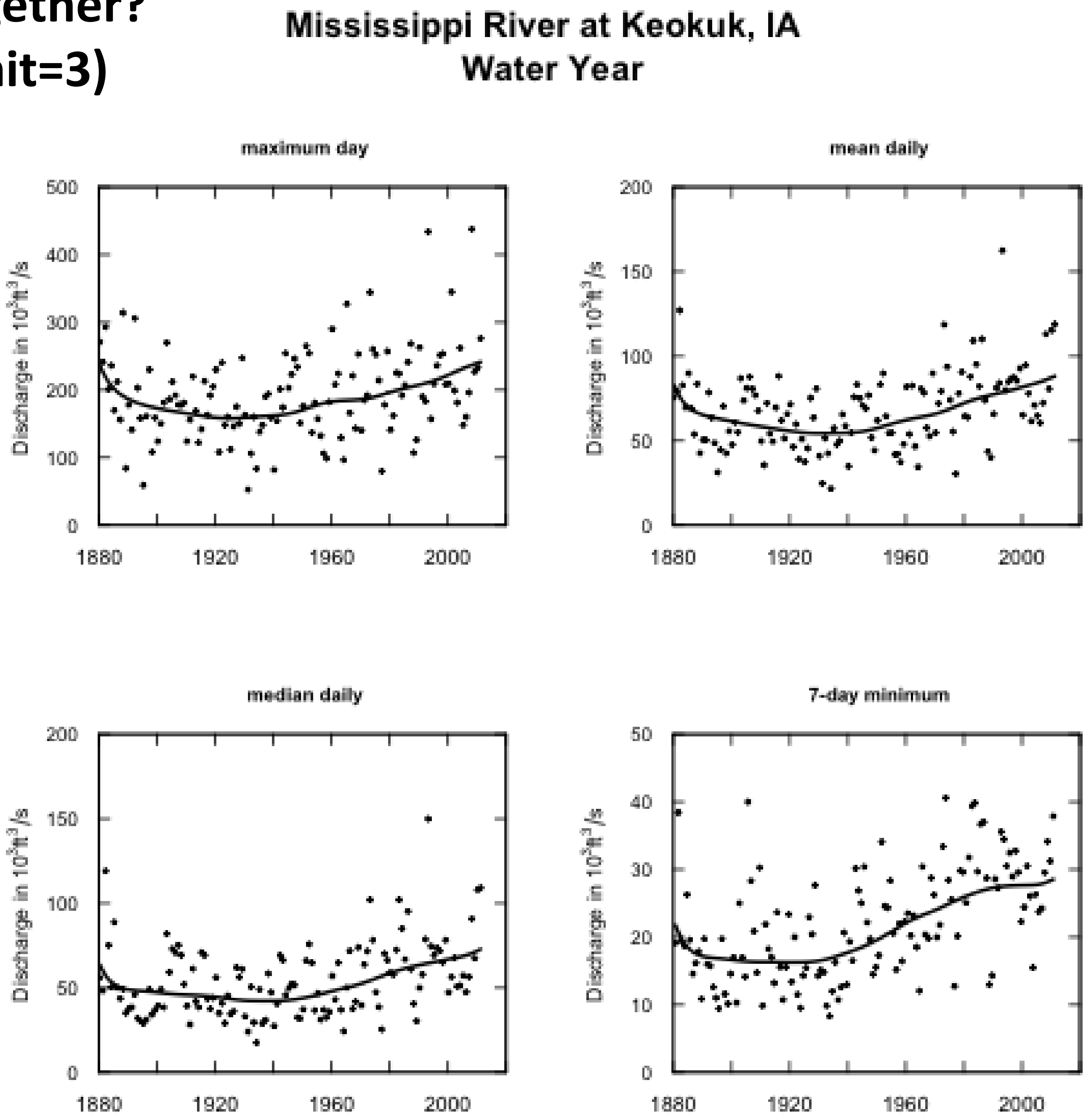


Lets look at 7-day low flows

> plotFlowSingle(istat=2,qUnit=3)



# Can we put it all together?
# > plotFourStats(qUnit=3)



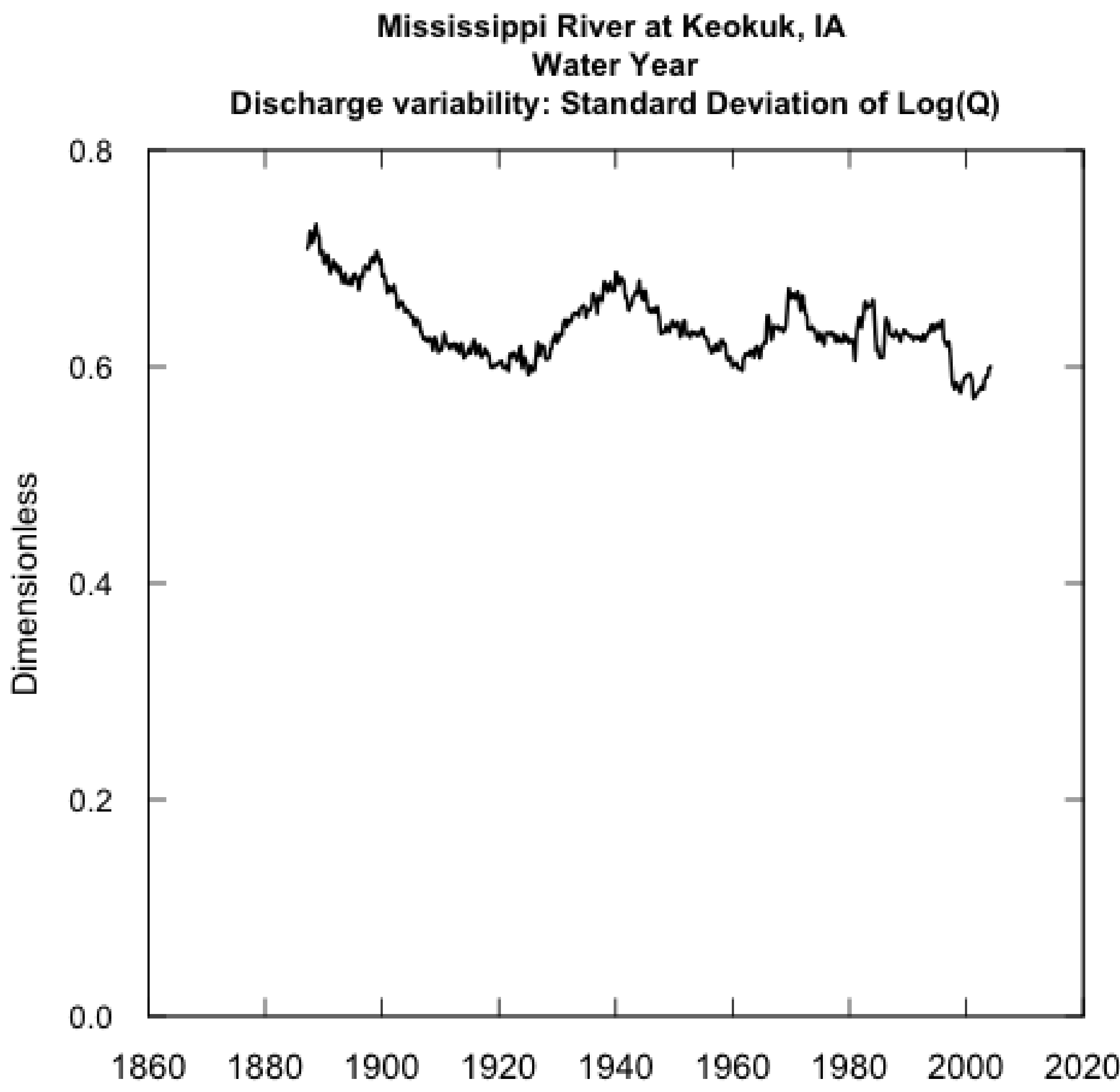
They say that variability is supposed to increase with increased greenhouse forcing

Here's one reasonable measure of that

The running standard deviation of the log discharge

Computed here on a 15-year moving window

> plotSDLogQ()



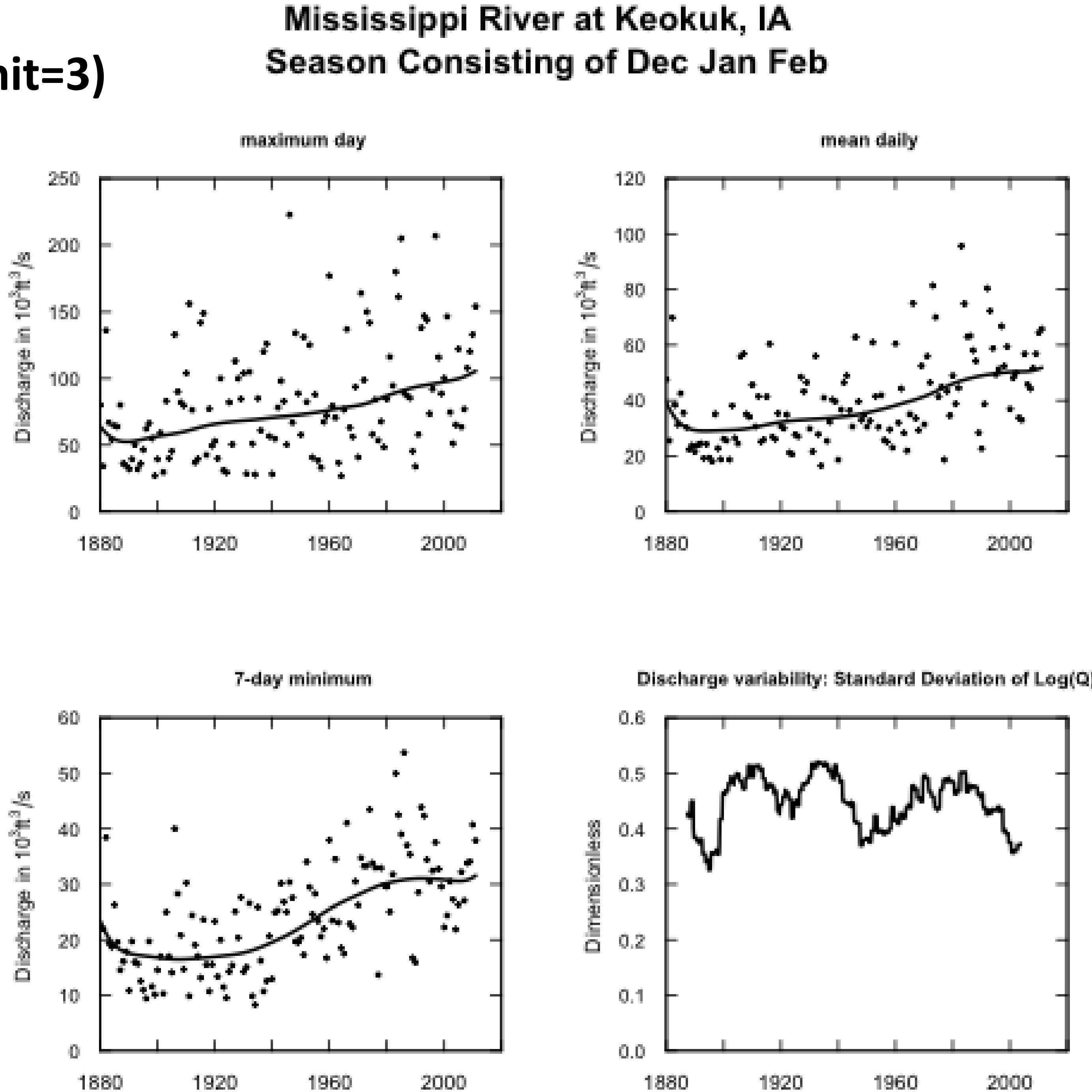
**A number of studies suggest that we are seeing warming, particularly in the winter, and that could have an impact on winter streamflow variability.**

**I will define winter as December, January, and February (you can pick your own definition).**

```
> INFO <- setPA(paStart = 12, paLong = 3)
> annualSeries <- makeAnnualSeries()
```

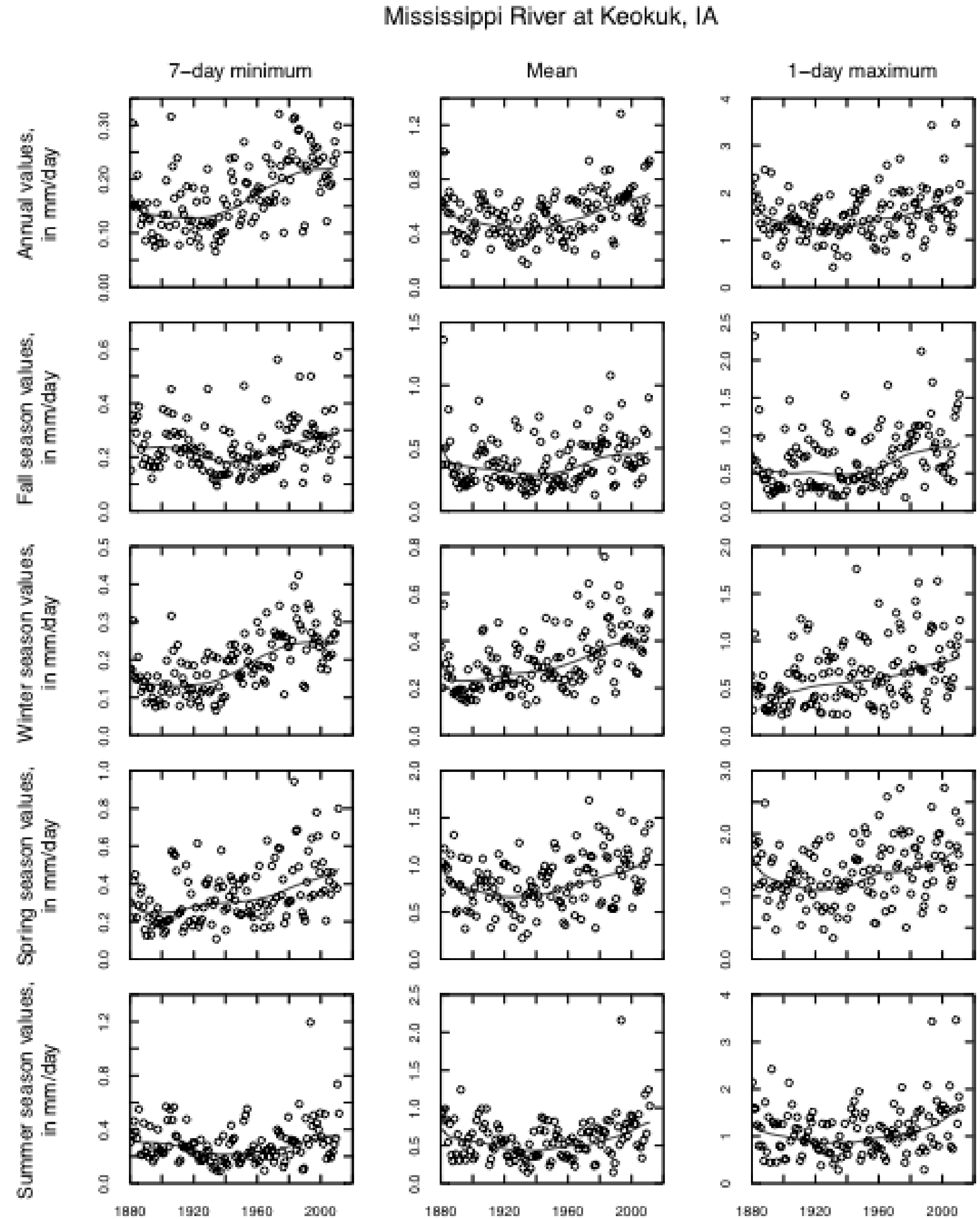
**So now the data frame "annualSeries" contains annual values of these flow statistics for this 3 month period only.**

> plotFour(qUnit=3)



**Designed for publication**
**Makes a pdf**
**Annual plus 4 seasons**
**7-day min, mean, max day**

> plot15(savePath,yearStart=1880, yearEnd=2012)




Streamflow statistics (circles) in units of millimeters per day, annual values and seasonal values
Fall (Sept., Oct., and Nov.), Winter (Dec., Jan., and Feb.), Spring (Mar., Apr., and May), and Summer (June, July, and Aug.)
and locally weighted scatterplot smooth (solid curve) for Mississippi River at Keokuk, IA for 1880 – 2012.

# I want to describe changes over time for May, June, and July:

```
> INFO<-setPA( paStart = 5 , paLong = 3)
> annualSeries<-makeAnnualSeries()
```

```
> tableFlowChange(istat=6,qUnit=2,yearPoints=c(1880,1940,1970,2011))
```

# I want to describe changes over time:

```
> tableFlowChange(istat=6,qUnit=2,yearPoints=c(1880,1940,1970,2011))

   Mississippi River at Keokuk, IA
   Season Consisting of May Jun Jul
    30-day maximum

              Streamflow Trends
   time span          change       slope        change      slope
                        cms     cms   /yr         %         %/yr
 1880  to  1940       -1974          -33        -37        -0.62
 1880  to  1970       -1365          -15        -26        -0.29
 1880  to  2011          87         0.67        1.7        0.013
 1940  to  1970         609           20         18         0.61
 1940  to  2011        2061           29         62         0.87
 1970  to  2011        1453           35         37          0.9
```

**I want numbers!!**

**What’s happened to 30-day maximum flows in May, June and July?**

```
> INFO<-setPA(paStart=5,paLong=3)
> annualSeries<-makeAnnualSeries()
> printSeries(istat=6,qUnit=4)
```

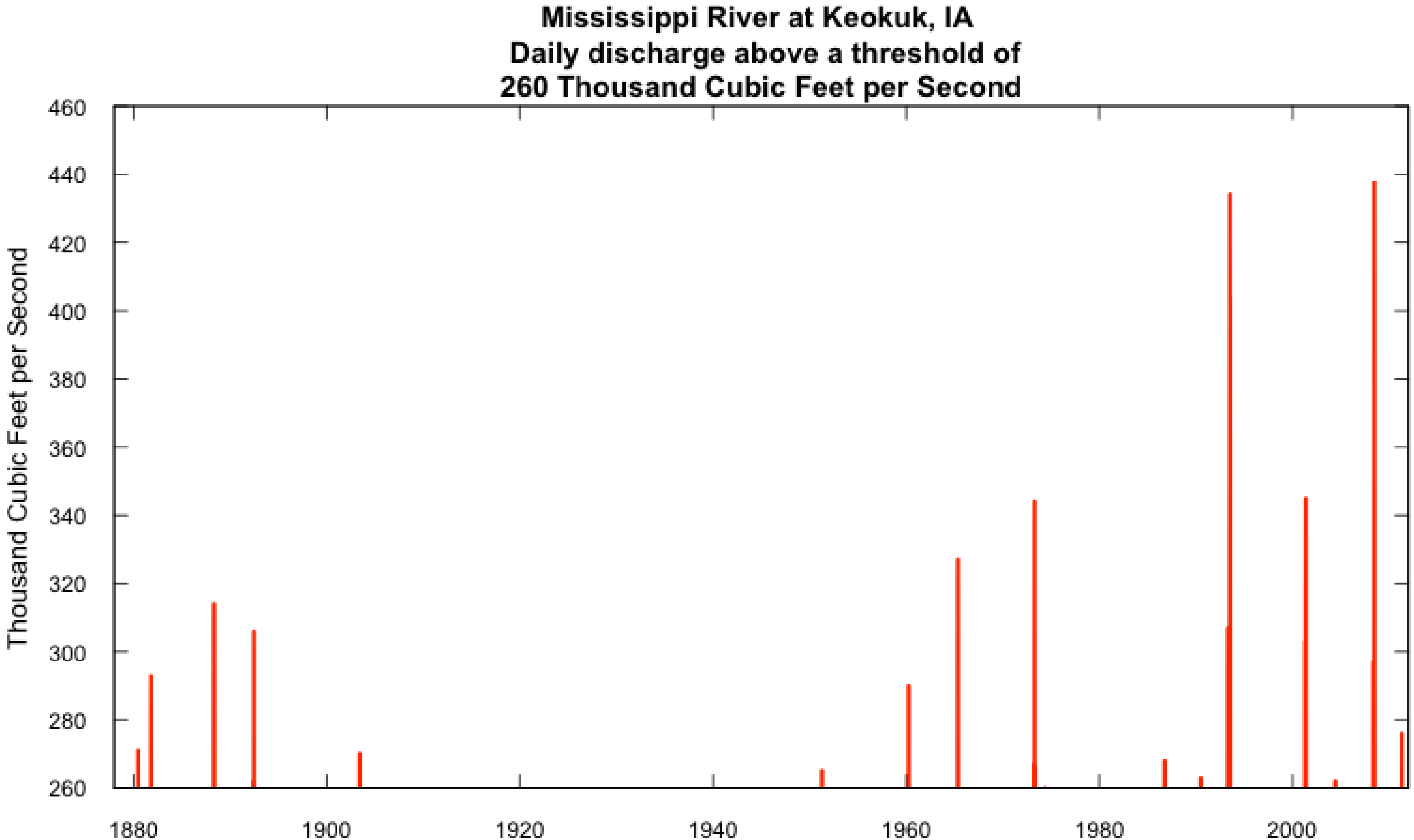
```
> printSeries(istat=6,qUnit=4)

 Mississippi River at Keokuk, IA
 Season Consisting of May Jun Jul
    30-day maximum
    Thousand Cubic Meters per Second
   year    annual     smoothed
            value      value


   1880      5.71       5.29
   1881      5.42       5.12
   1882      5.62       4.96
   1883      5.06       4.82
   1884      4.53       4.71
   1885      4.00       4.61
   1886      5.13       4.52
   1887      3.59       4.44

   1991      4.76       4.56
   1992      4.04       4.59
   1993     10.99       4.62
   1994      3.59       4.65
   1995      5.00       4.69
   1996      5.40       4.73
   1997      5.87       4.78
   1998      5.76       4.83
   1999      5.15       4.89
   2000      4.87       4.95
   2001      8.30       5.01
   2002      4.45       5.07
   2003      4.29       5.13
   2004      6.12       5.18
   2005      3.58       5.24
   2006      4.01       5.29
   2007      4.59       5.33
   2008      7.66       5.36
   2009      4.10       5.38
   2010      5.15       5.39
   2011      6.60       5.38
```

Let's look at the question about extreme high flows becoming more common.

To pick a threshold to consider use flow duration curve.

```
> flowDuration(qUnit=3)

Flow Duration for Mississippi River at Keokuk, IA

Flow duration is based on full year

Discharge units are Thousand Cubic Feet per Second

    min      5%     10%     25%     50%     75%     90%     95%     max
    5.0    20.0    23.5    32.7    52.1    89.0   137.0   167.0   438.0
```

> plotQTimeDaily(1878,2012,qLower=260,qUnit=3)



**This approach could be formalized into a statistical test, using Mann-Kendall approach**

**Null hypothesis: expected number of events per year is constant over time**

**Alternative hypothesis is that it is changing monotonically**

**Test will ignore the natural tendency towards serial correlation, thus it will be a little too likely to lead to a rejection of the null hypothesis**

# We can also look at seasonal flow duration curves, not just annual ones.

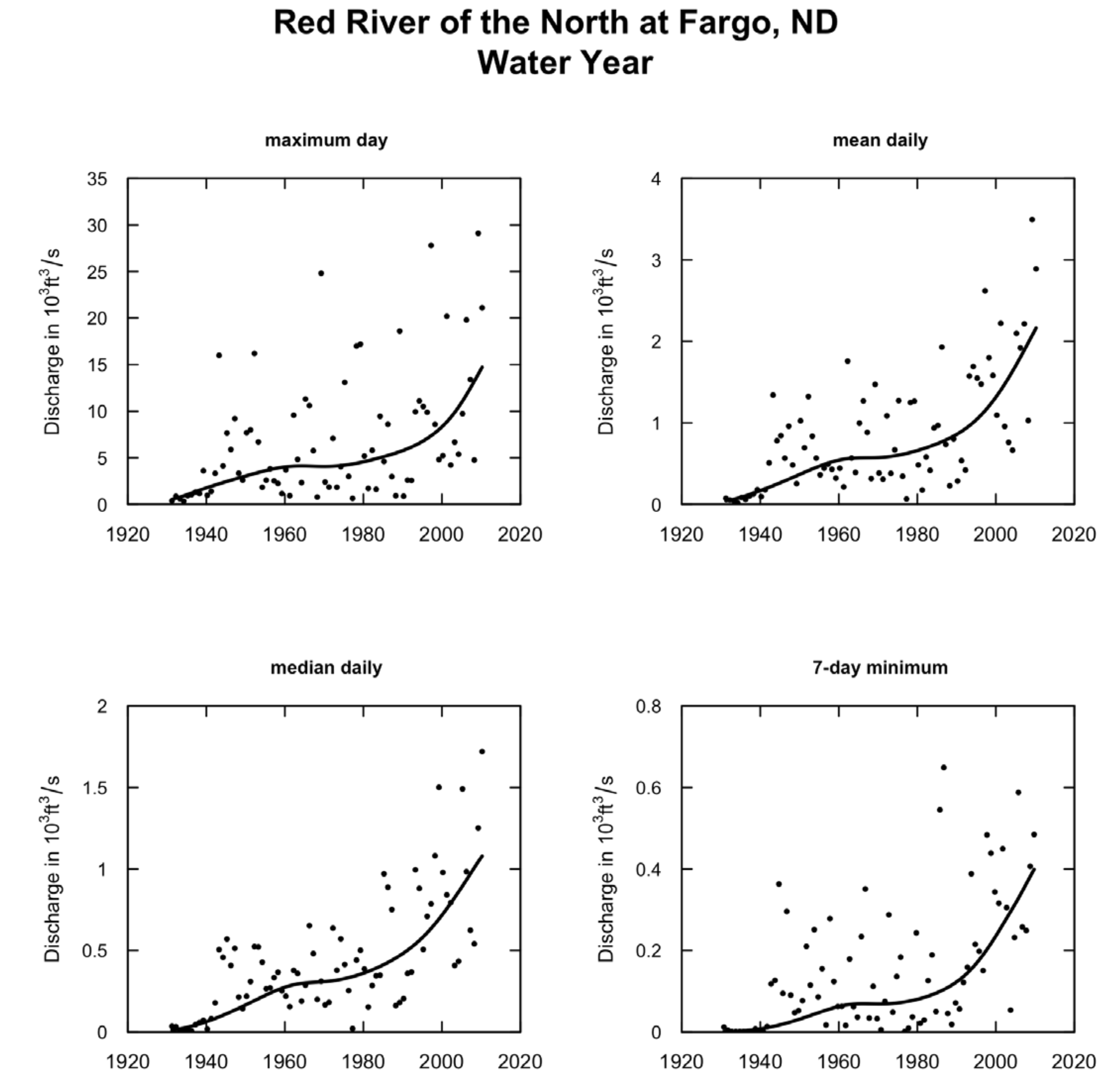
```
> flowDuration(centerDate="05-15",qUnit=3,span=45)

Flow Duration for Mississippi River at Keokuk, IA

Flow duration period is centered on May 15
And spans the period from March 31  To June 29

Discharge units are Thousand Cubic Feet per Second

     min      5%     10%     25%     50%     75%     90%     95%     max
    15.4    36.9    47.0    69.2   102.0   144.3   184.7   209.4   438.0
```

# Look at the last three decades

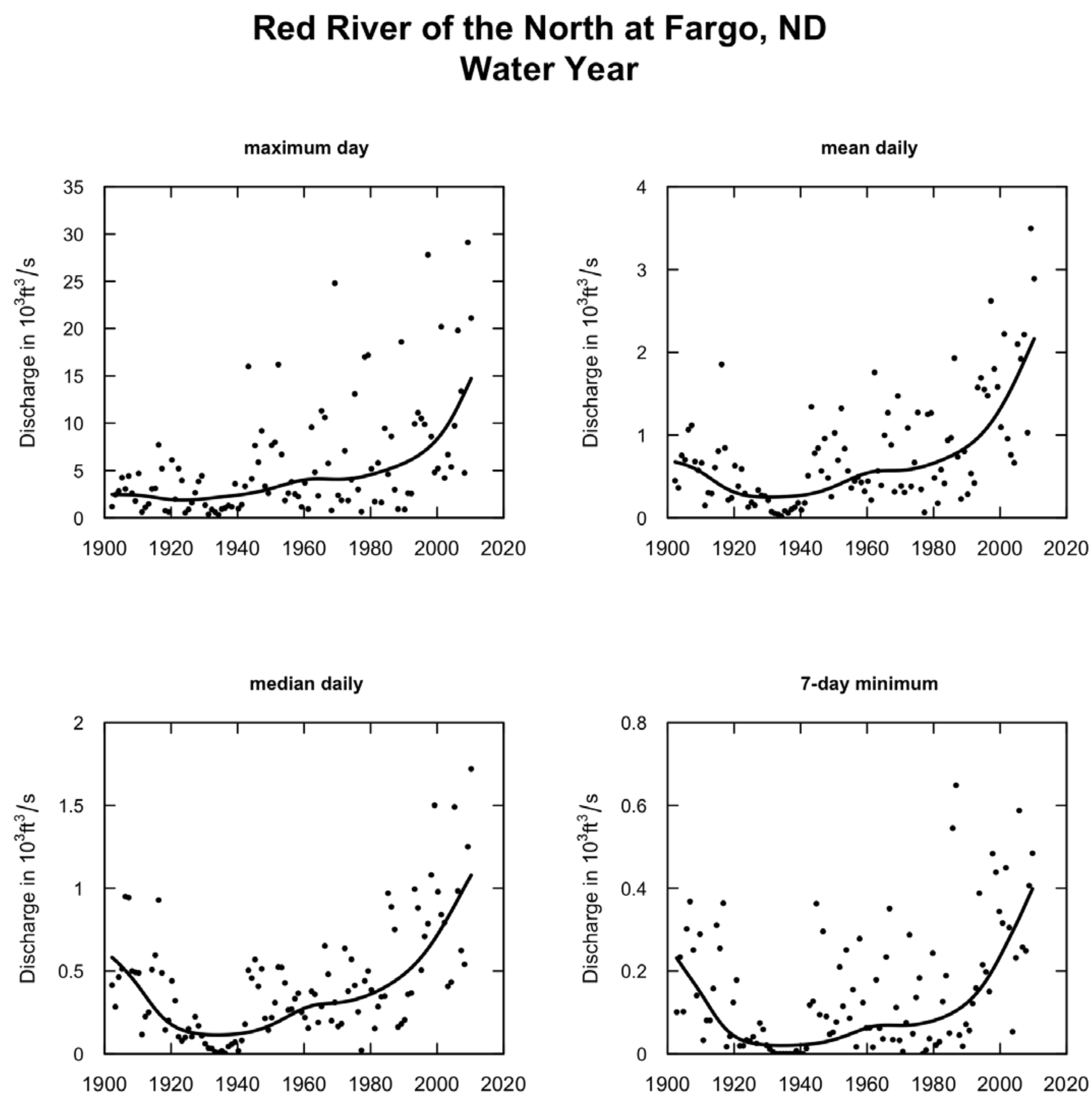
**1-day max**
**+ 67% per decade**

**Mean**
**+ 71% per decade**

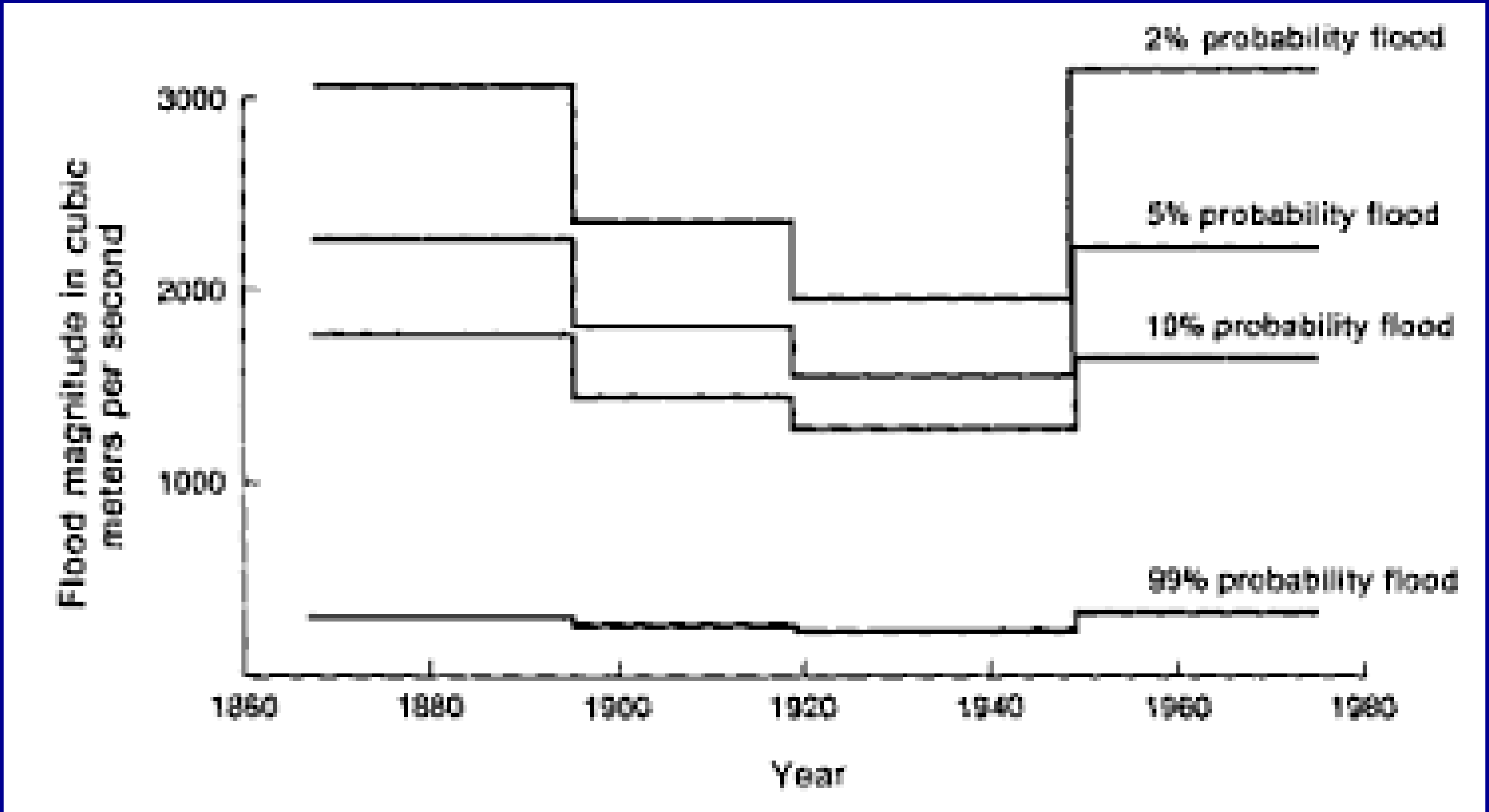
**7-day min**
**+ 130% per decade**



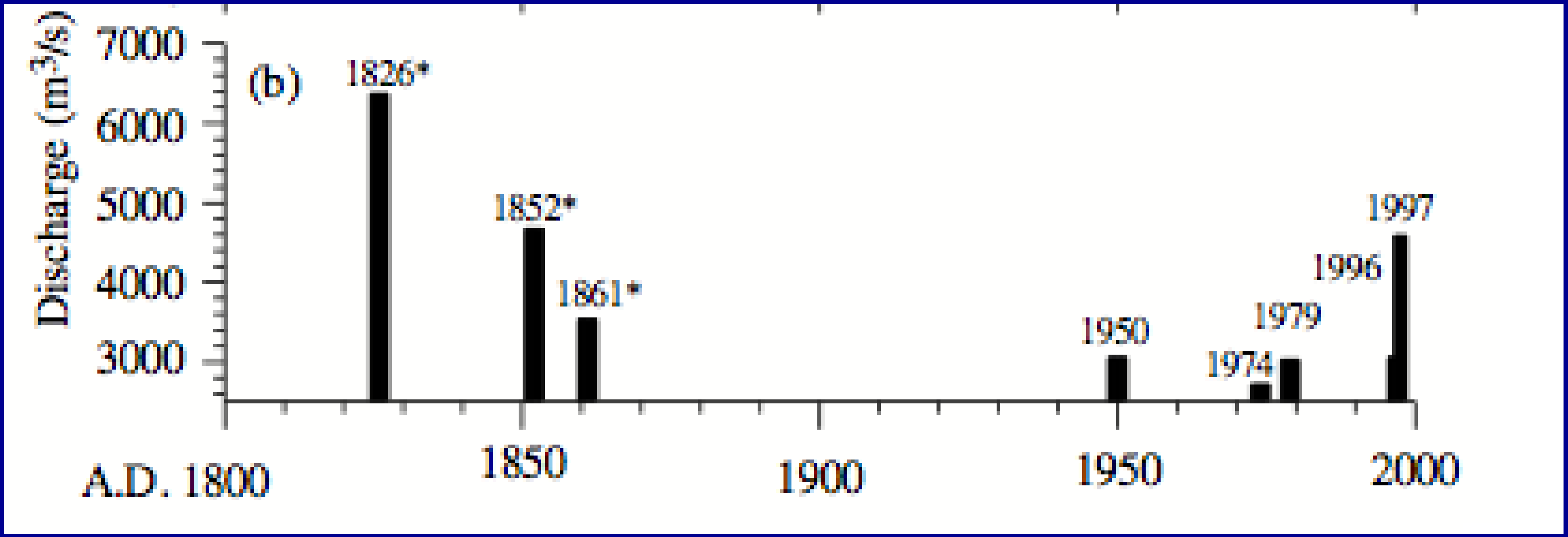
**Does it make sense to explain this with greenhouse forcing or land drainage or cropping patterns?**



# Paleo reconstruction of flood frequency, Mississippi River at St. Paul: Knox, 1983

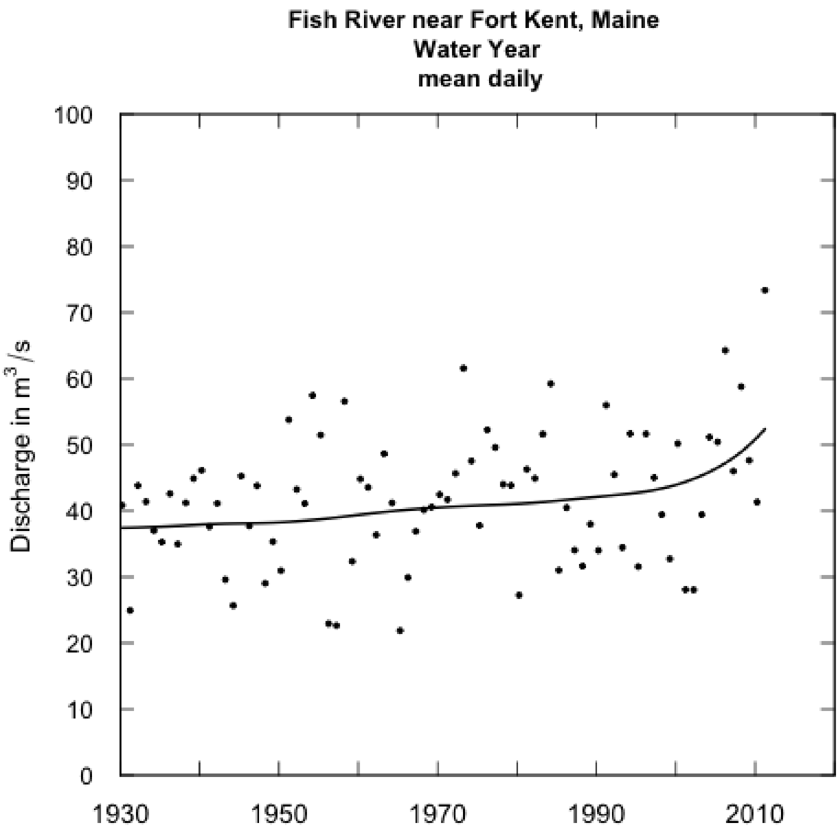


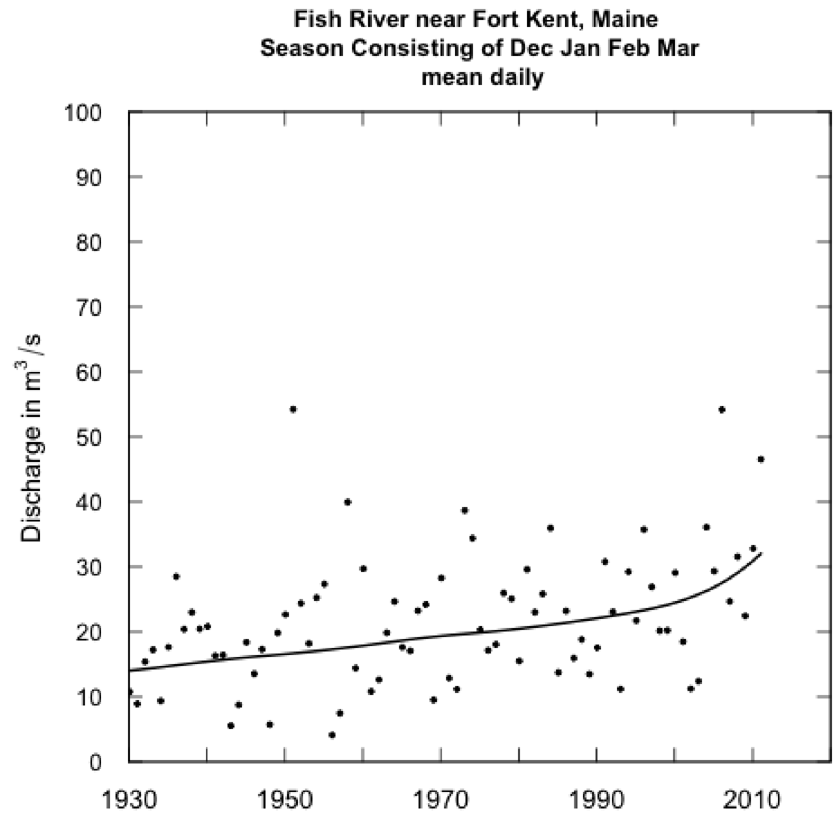
# Tree ring based flood magnitudes, Red River near Winnipeg



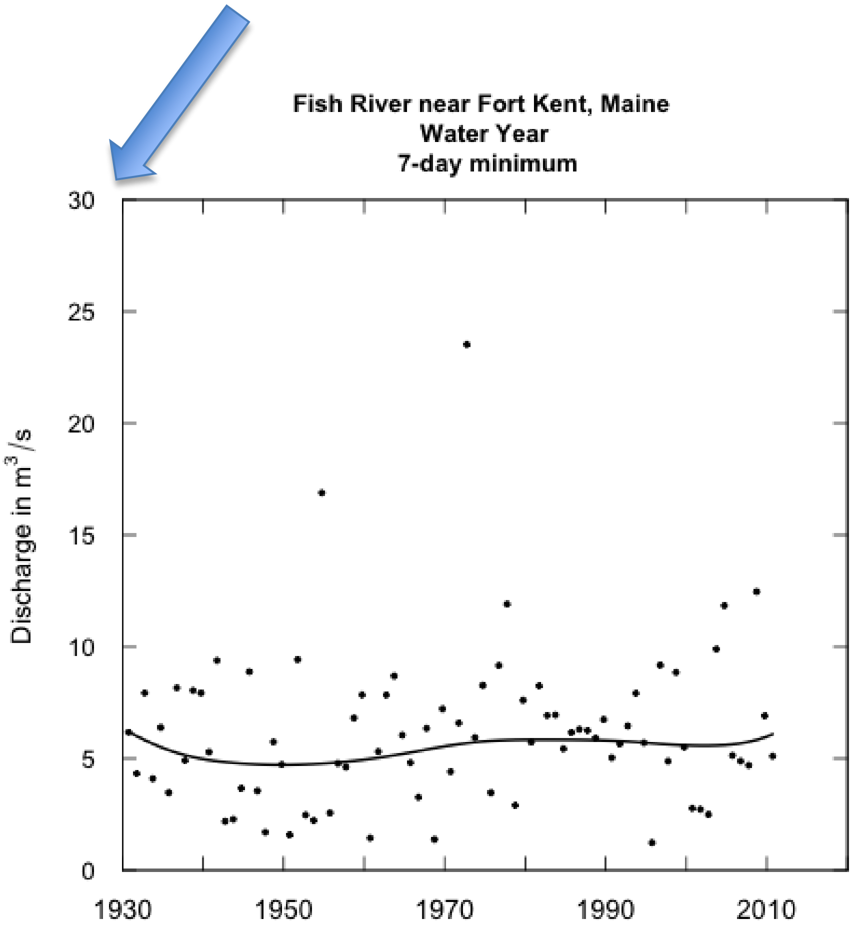

From Scott St. George and Erik Nielsen, Geology 28, 899-902, 2000, data credited to Rannie (1998)

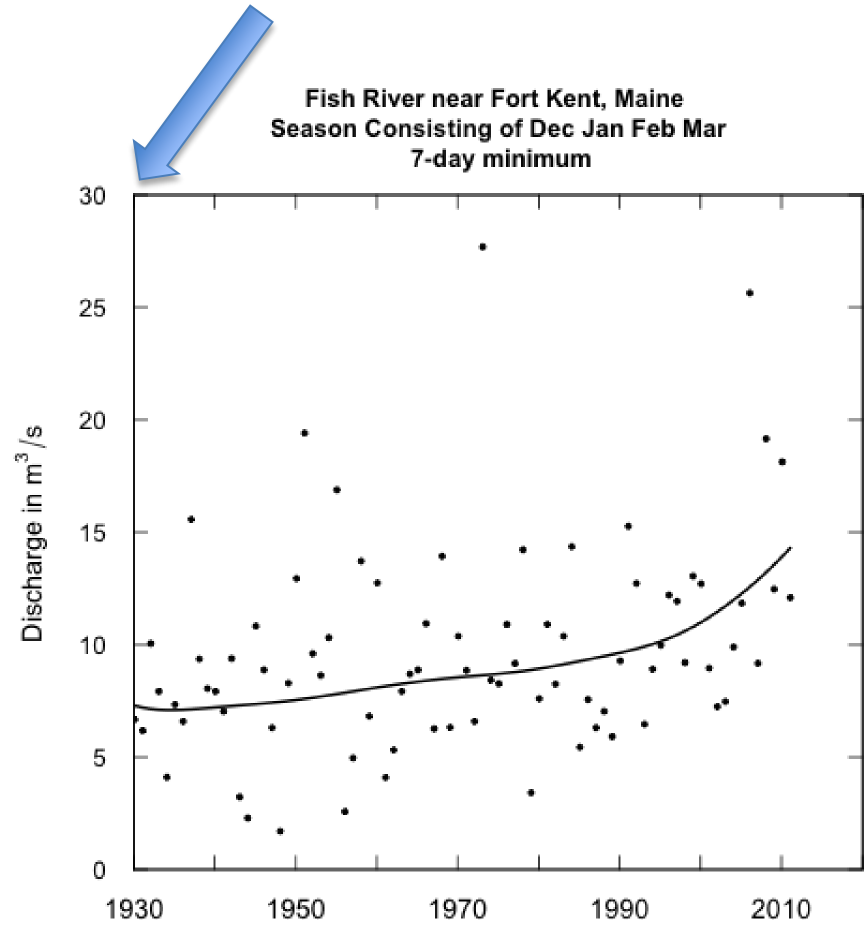
# Let’s look at a very cold place. Northern Maine. Mean Daily Q



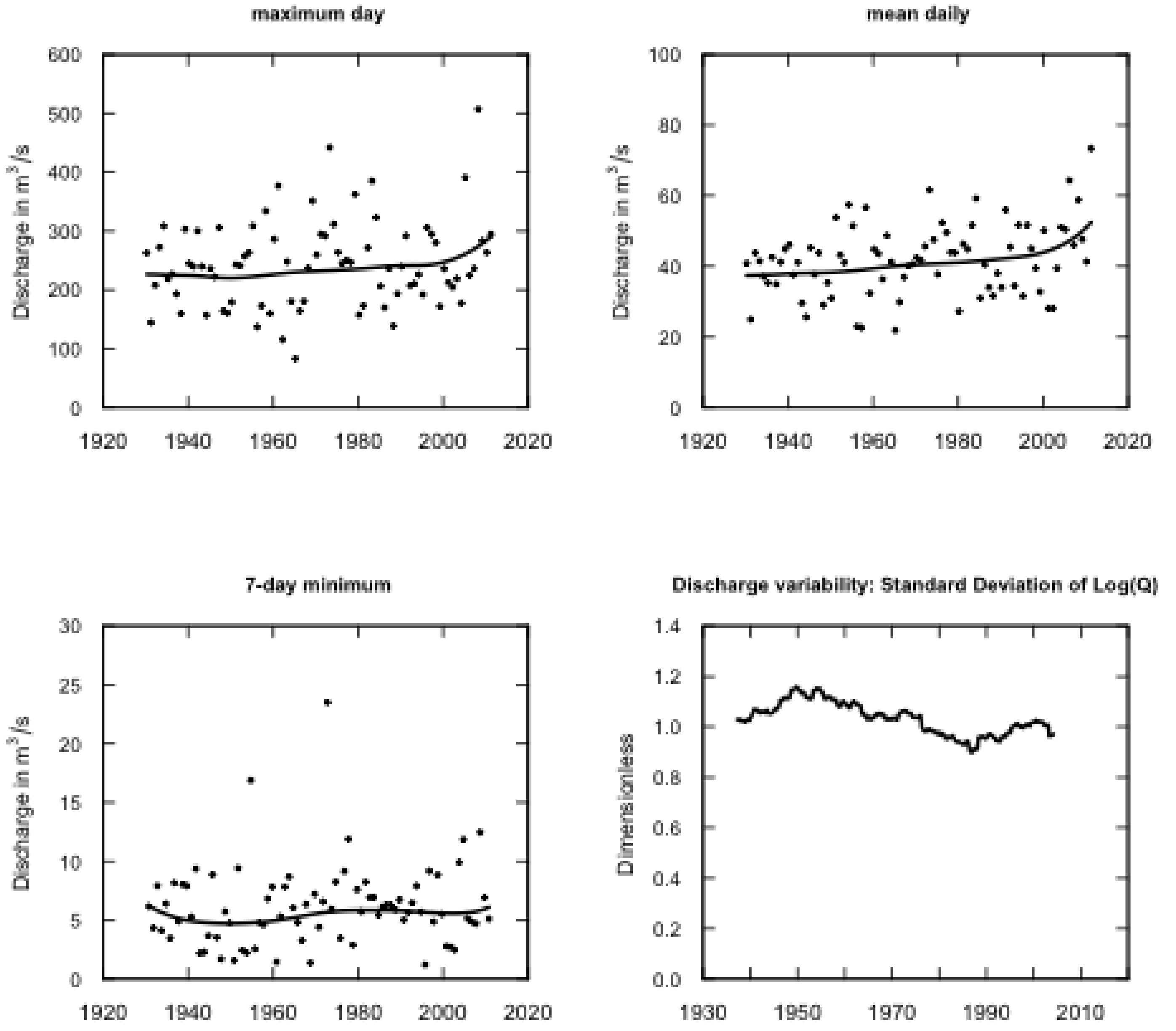
# Let's look at changes in annual minima.



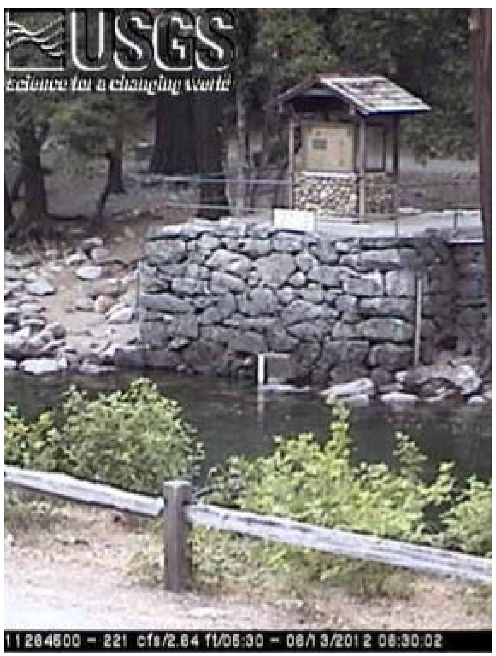



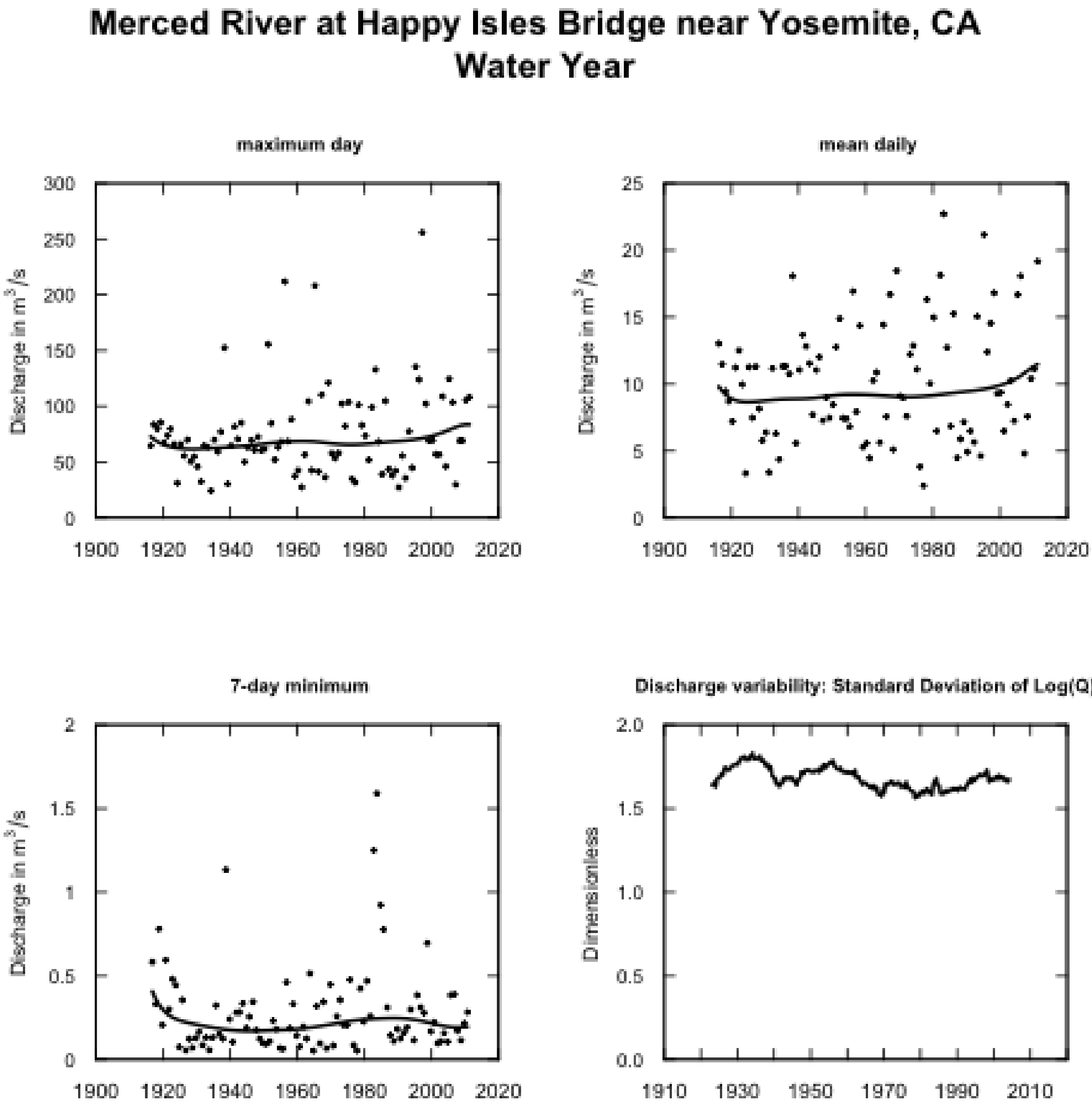

**On an annual basis it is unchanged, but for winter it is going up.**

## Fish River near Fort Kent, Maine
## Water Year

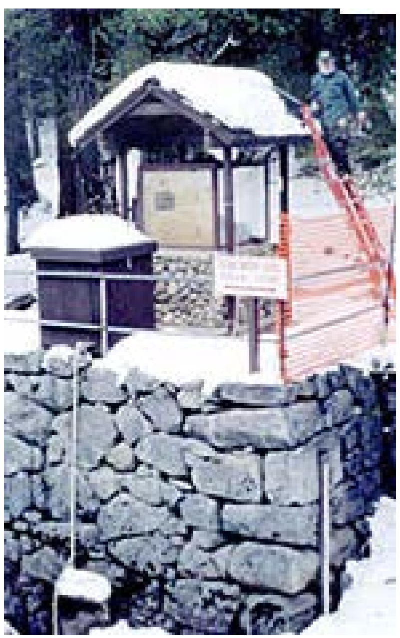


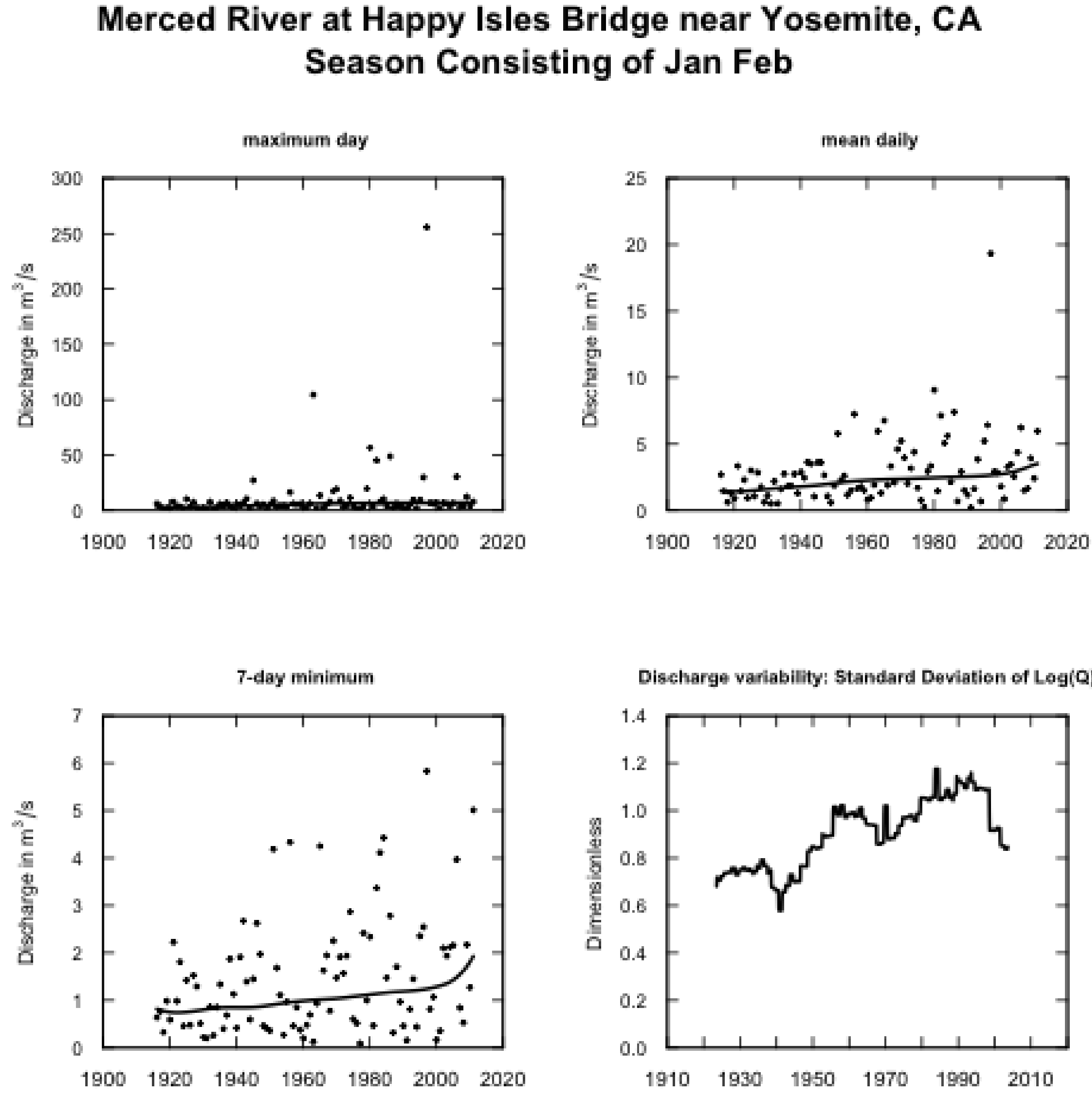
# Let's go west to a very pristine basin high in the Sierra Nevada



# What if we look at the coldest months of the year





# *Relationship of streamflow & greenhouse forcing*

**Approaches to the issue:**

1) **Use climate models to drive hydrologic models and describe the changes in the simulated hydrologic variables.**
2) **Consider the past century to be an unplanned global experiment. Use streamflow records in many watersheds as "experimental subjects" and look for the "effect".**
3) **Use climate models in hindcast mode to create multiple realizations of the past century. Then see if what actually happened in the last century falls within the envelope of what the climate models would hindcast.**

# An example of approach 2

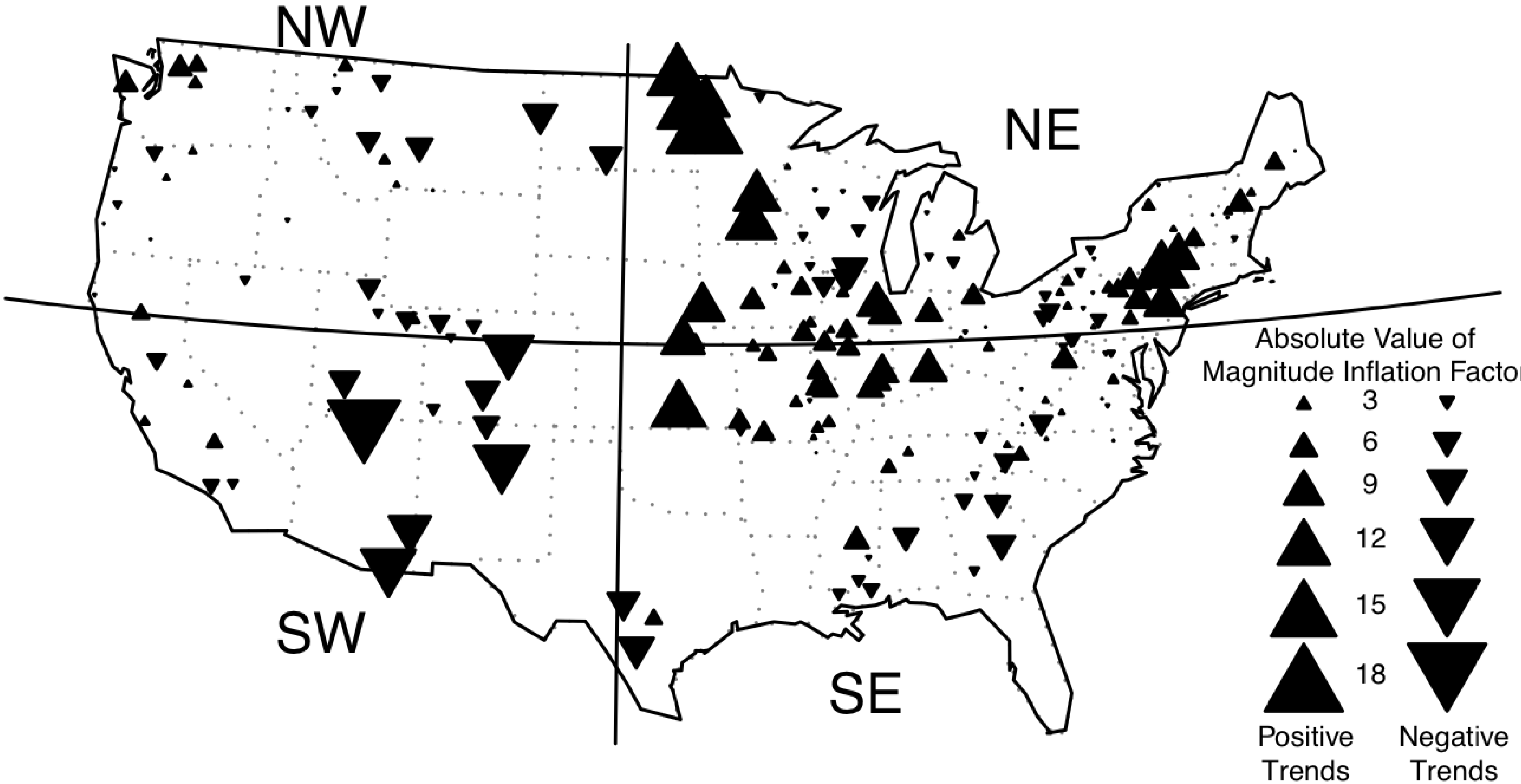
**Hirsch and Ryberg, 2011, Hydrologic Sciences Journal**

**Picked 200 very long streamflow records with limited human intervention.**

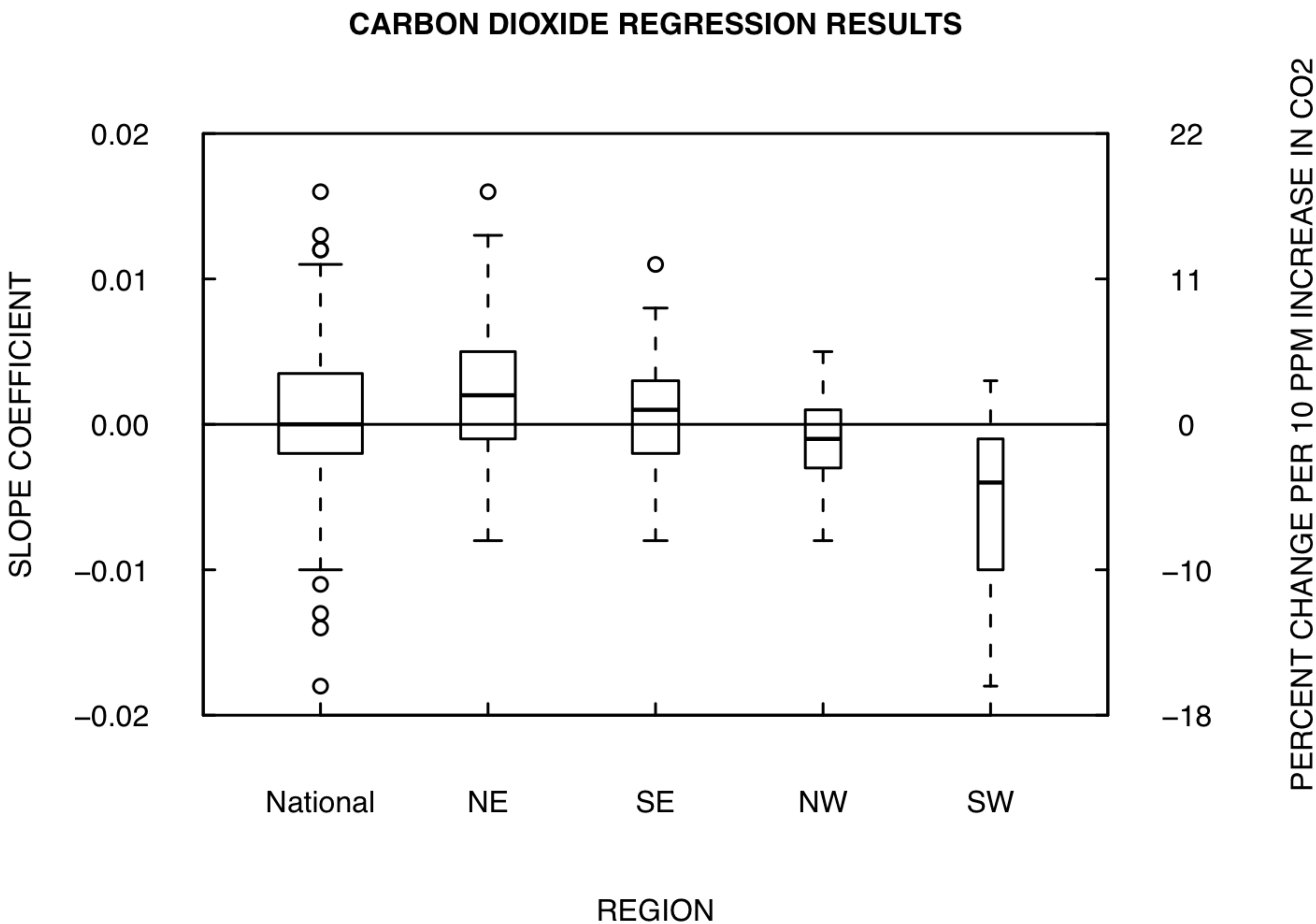
**Regress log(annual peak discharge) on global mean $CO_2$ for that year and take the regression slope as a measure of "effect" without regard to "statistical significance"**

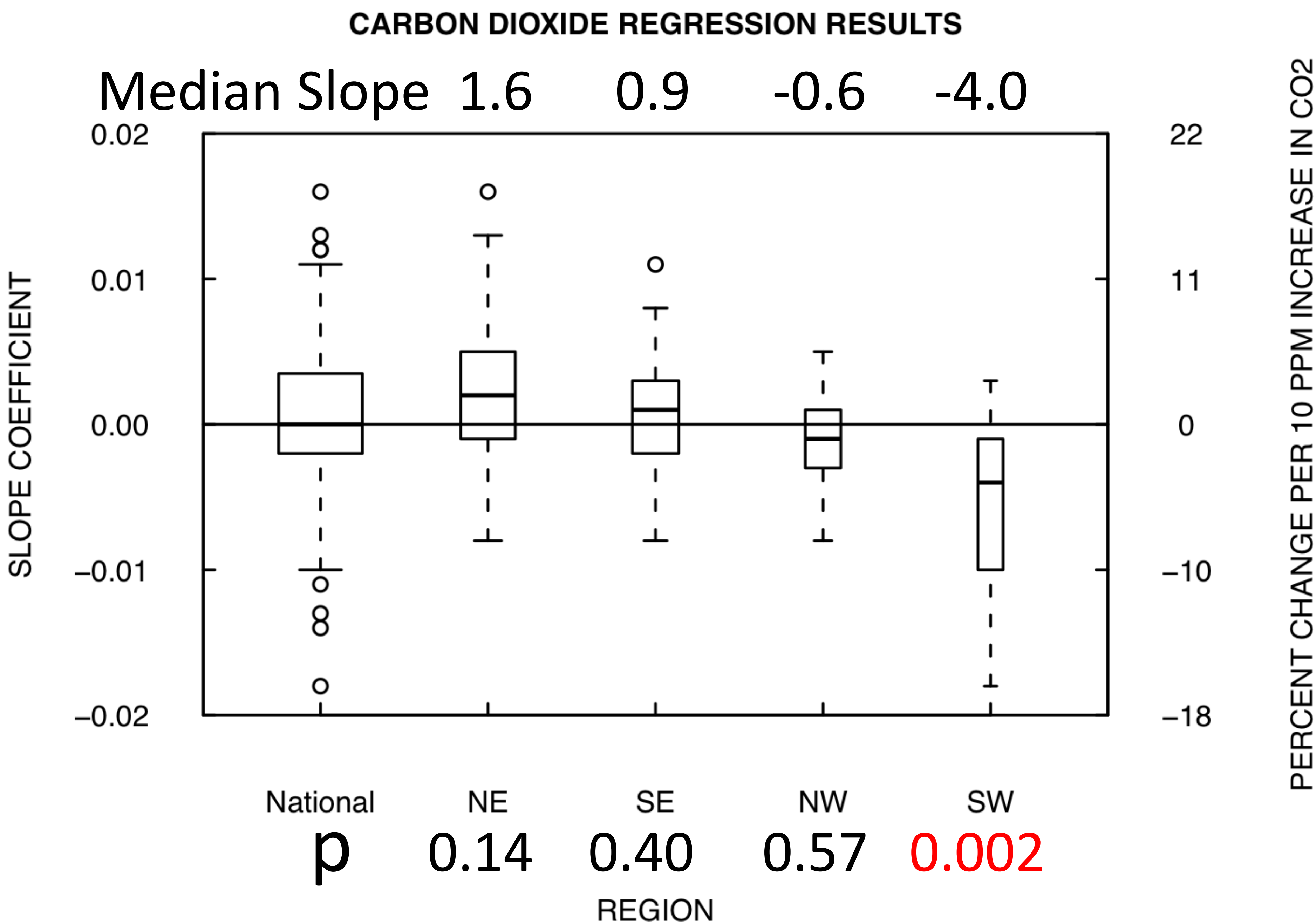
**What's the pattern, nationally, regionally, by drainage area?**

# National results: 200 streamgage records




**Units are % change per 10 ppm change in $CO_2$**
**(current increase is about 10 ppm every 5 years)**


CARBON DIOXIDE REGRESSION RESULTS
SLOPE COEFFICIENT
0.02
0.01
0.00
−0.01
−0.02
PERCENT CHANGE PER 10 PPM INCREASE IN CO2
22
11
0
−10
−18
National
NE
SE
NW
SW
REGION


CARBON DIOXIDE REGRESSION RESULTS
Median Slope 1.6 0.9 -0.6 -4.0
0.02
0.01
0.00
−0.01
−0.02
SLOPE COEFFICIENT
22
11
0
−10
−18
PERCENT CHANGE PER 10 PPM INCREASE IN CO2
National NE SE NW SW
p 0.14 0.40 0.57 0.002
REGION

# Take away messages:

- The only region in which there is strong statistical evidence of an association between floods & global $CO_2$ is in the southwest, and the relationship there is negative.
- All approaches to understanding the streamflow/greenhouse gas connection have flaws. But we need to look at the data regularly and with diverse approaches to see what might be emerging.

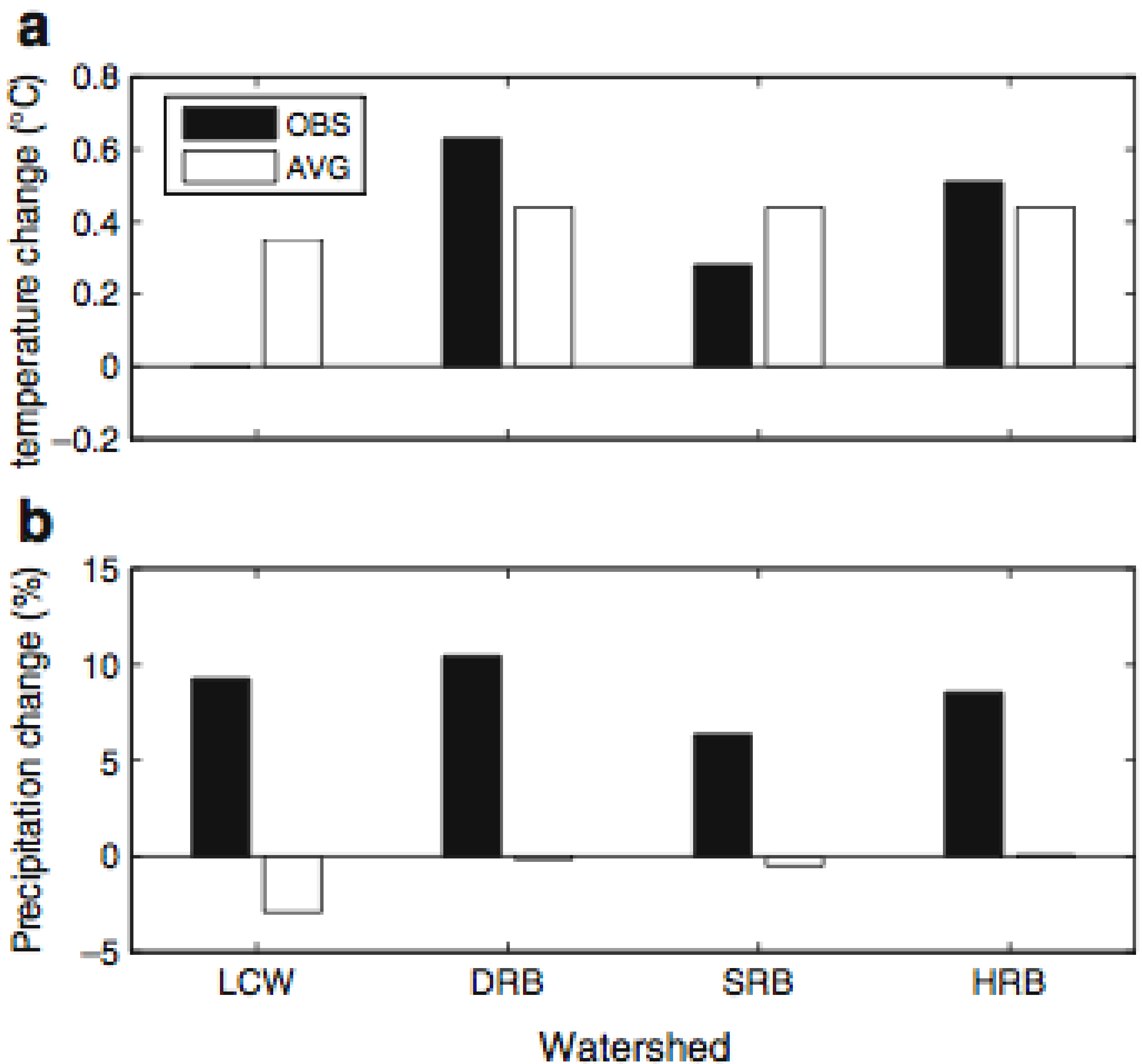
# Approach 3 (hindcast studies)

- **First is Najjar, et.al. 2009, on temperature and precipitation in the Chesapeake, Delaware River and Hudson River watersheds.**
- **Used multiple climate model hindcasts of 20th century and captured the average change between the 1911-1940 period and the 1971-2000 period.**
- **Compared these changes to changes in the observed records for the same watersheds.**

From Najjar et al. 2009,

Comparing 1911-1940 to 1971-2000





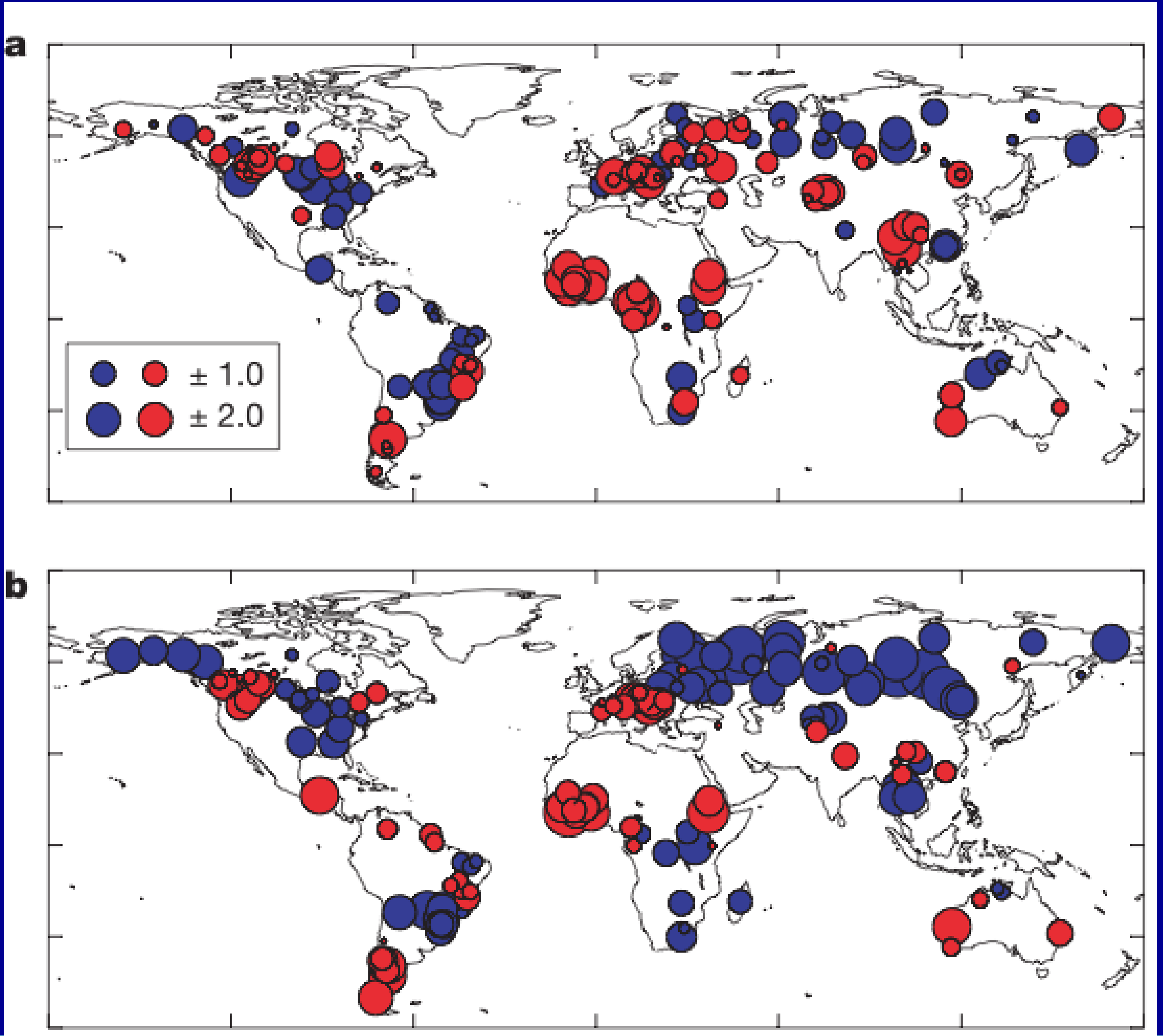
# What are we to conclude?

- **Looking at real hydrologic records my refrain always seems to be:**
- **"And you know something's happening but you don't know what it is…..do you, Mr. Jones?"***
- **** Words and music by Bob Dylan, "Ballad of a Thin Man", Highway 61 revisited.***

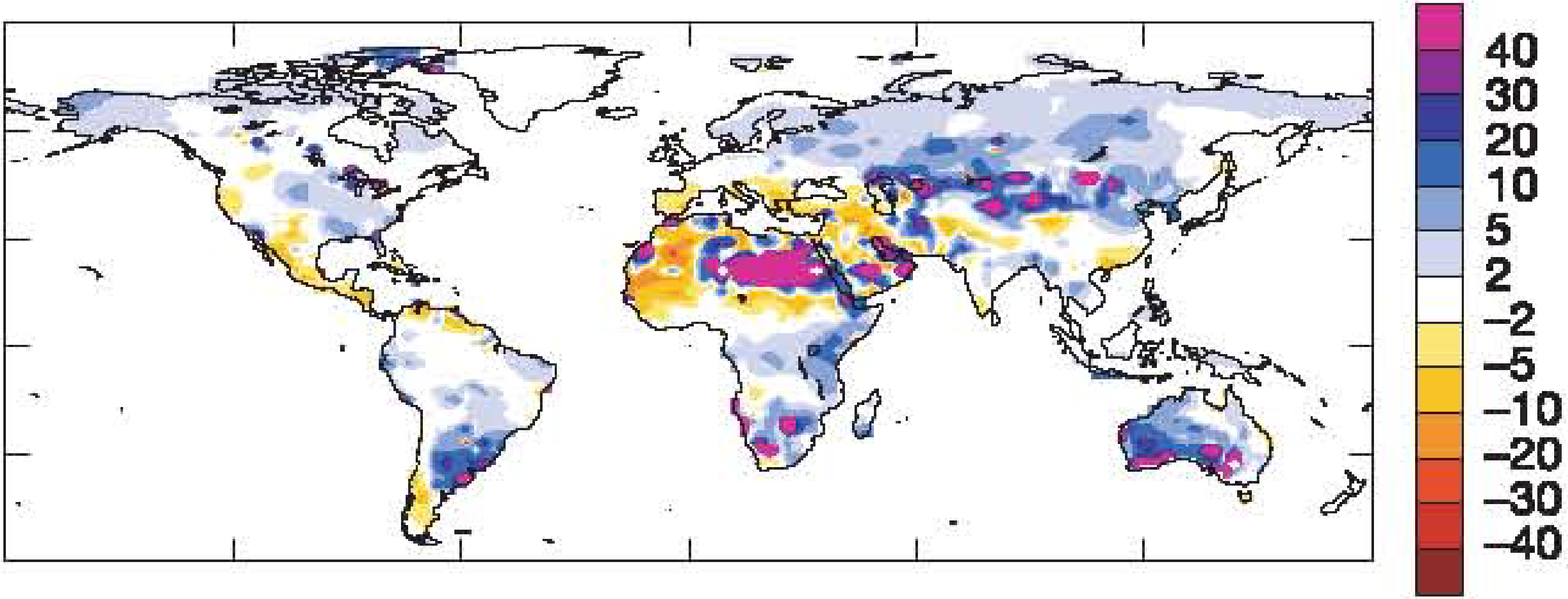
# Milly, Dunne, and Vecchia, Nature, 2005: Comparison of streamflow: 1900-1970 to 1971-1998

**Streamgage Data**

**Averaged results of 35 GCM runs**

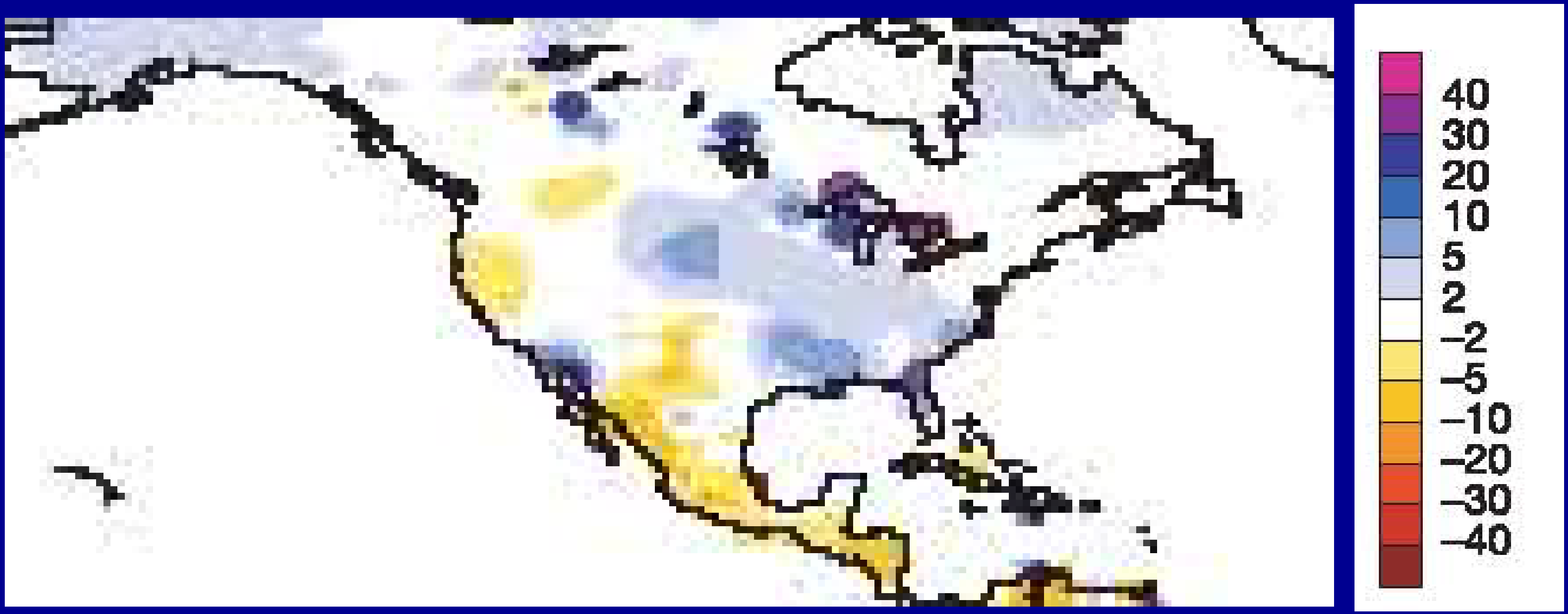


# Milly, et al. (2005, Nature)




Estimated percentage change in runoff for 1971-98 vs. 1900-1970 due to global atmospheric forcing, ensemble of GCM model runs

# Milly, et al. (2005, Nature)



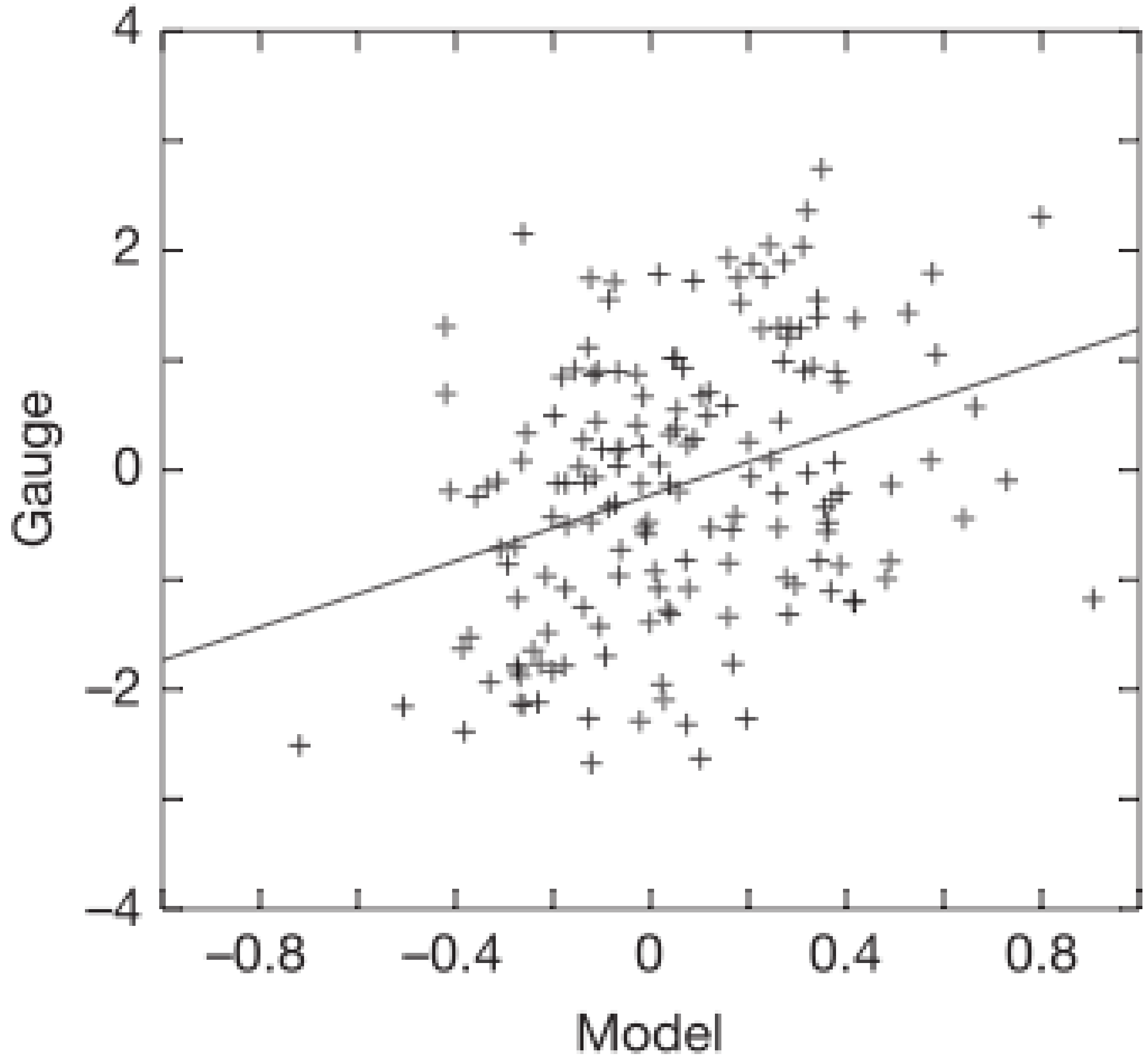

**Actual examples:**

| | |
|---|---|
| Potomac River, Point of Rocks, MD | +23% |
| Mississippi River, Keokuk, IA | +24% |
| Red River of the North, Grand Forks, ND | +65% |

# Milly, Dunne, and Vecchia, Nature, 2005: Comparison of streamflow: 1900-1970 to 1971-1998

**Plotting all those pairs of model versus streamgage data.**

**Results are "statistically significant" but $R^2$ = 12%**



# Take away messages

- Studies of type 2 and 3 are very few and far between.  The USGS needs to do more.
- Empirical hydroclimatic studies MUST be increased.  The climate models must be judged against actual history.
- No one is better suited to this than the USGS. We collected the data & we understand it.

# Why don't we generally see the impacts that the modelers suggest we should see?

- **Maybe the models are wrong.**
- **Maybe the models are right, but the signal is still small compared to the noise.**
- **Maybe the models are right, but there is a threshold.**
- **Maybe direct human impacts on watersheds overwhelm these climate-driven impacts.**

# Back to EGRET

- **We can use the tools of EGRET to depict the profound non-climatic influences on streamflow.**

- **We need to learn more about these, and from that experience, improve our approaches to doing frequency estimation in a non-stationary world.**

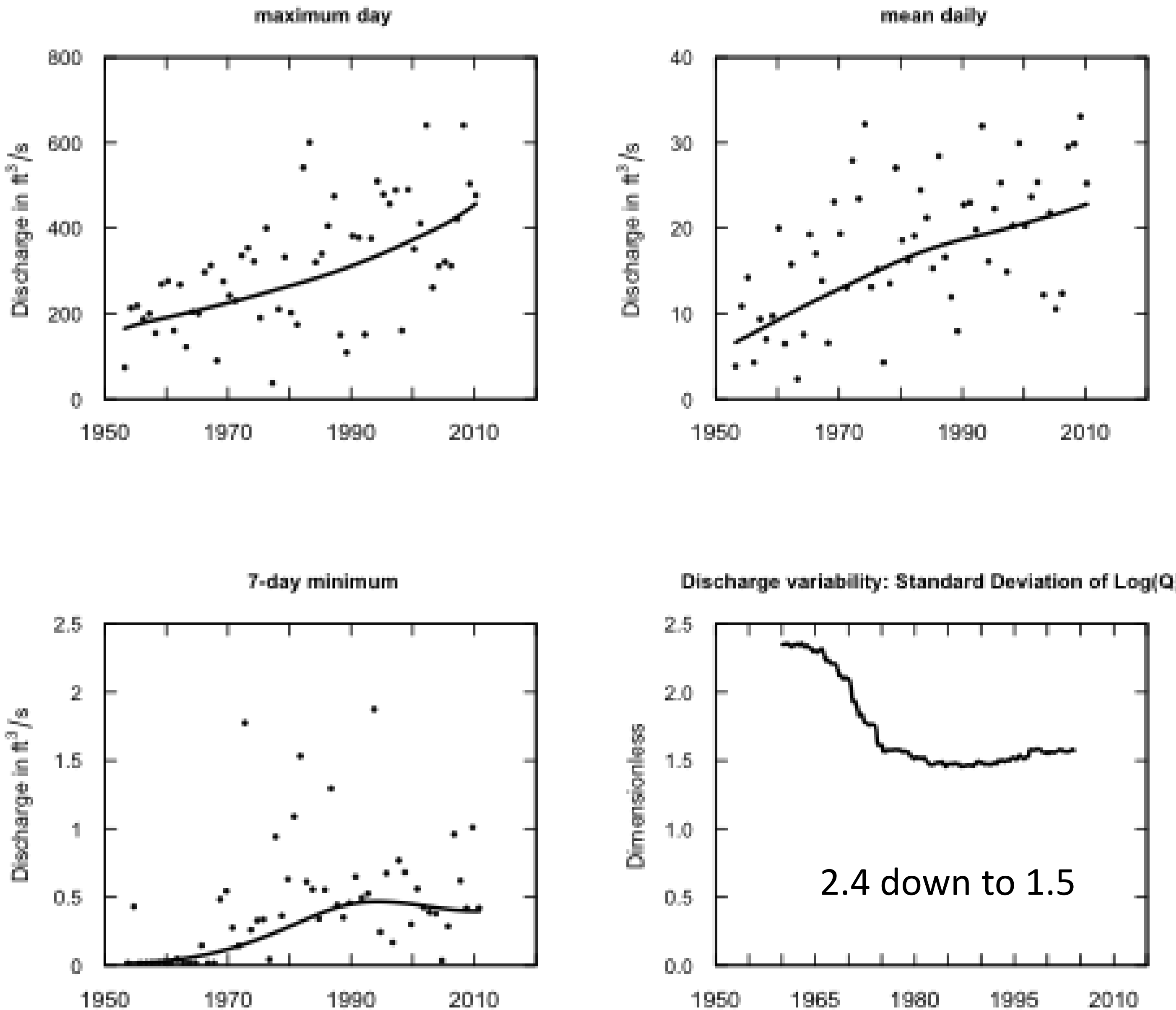
# Urbanization

## North Branch Chicago River at Deerfield, IL
## Water Year

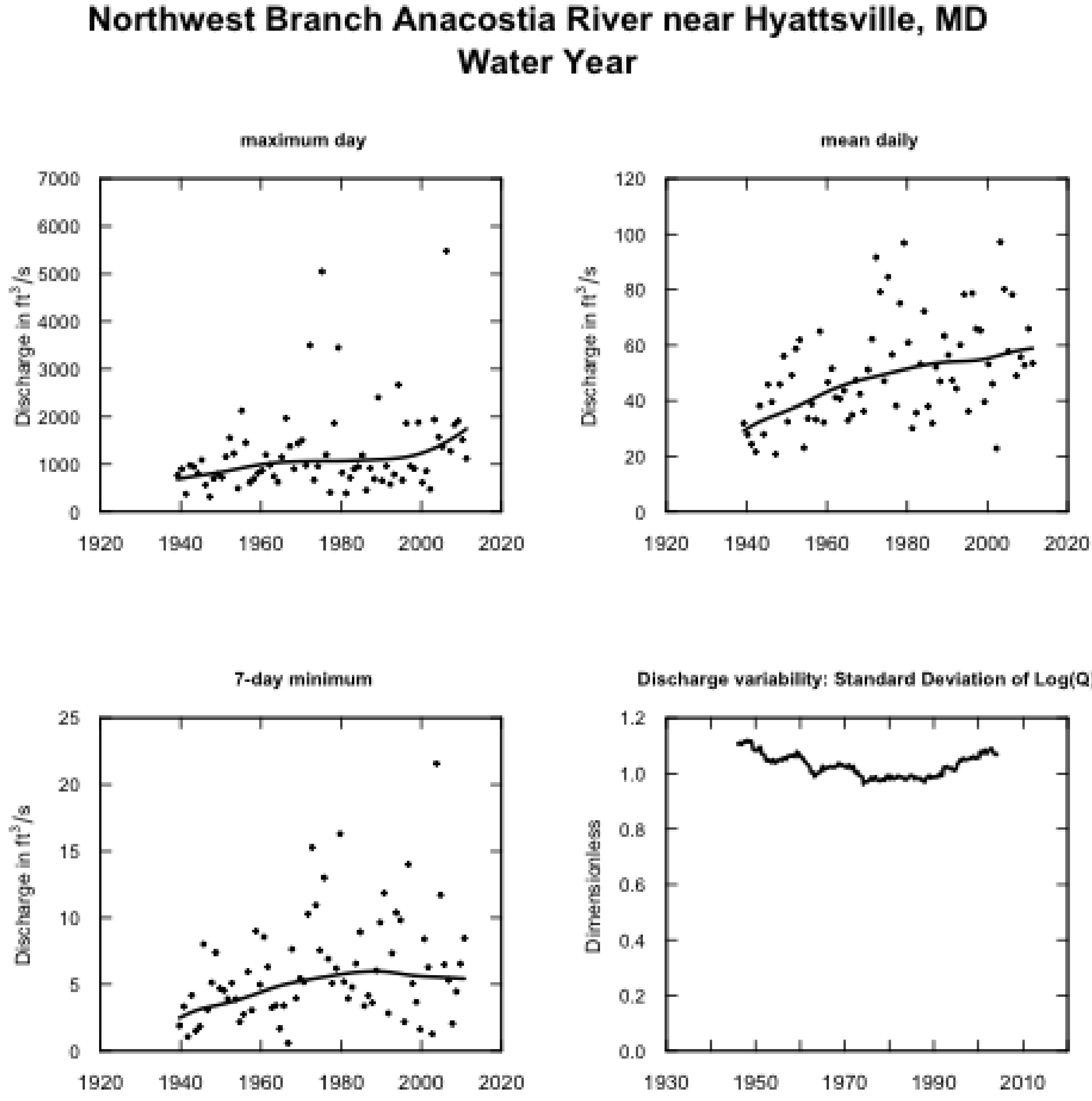
Up 222%
1955-2010

Up 163%
1955-2010

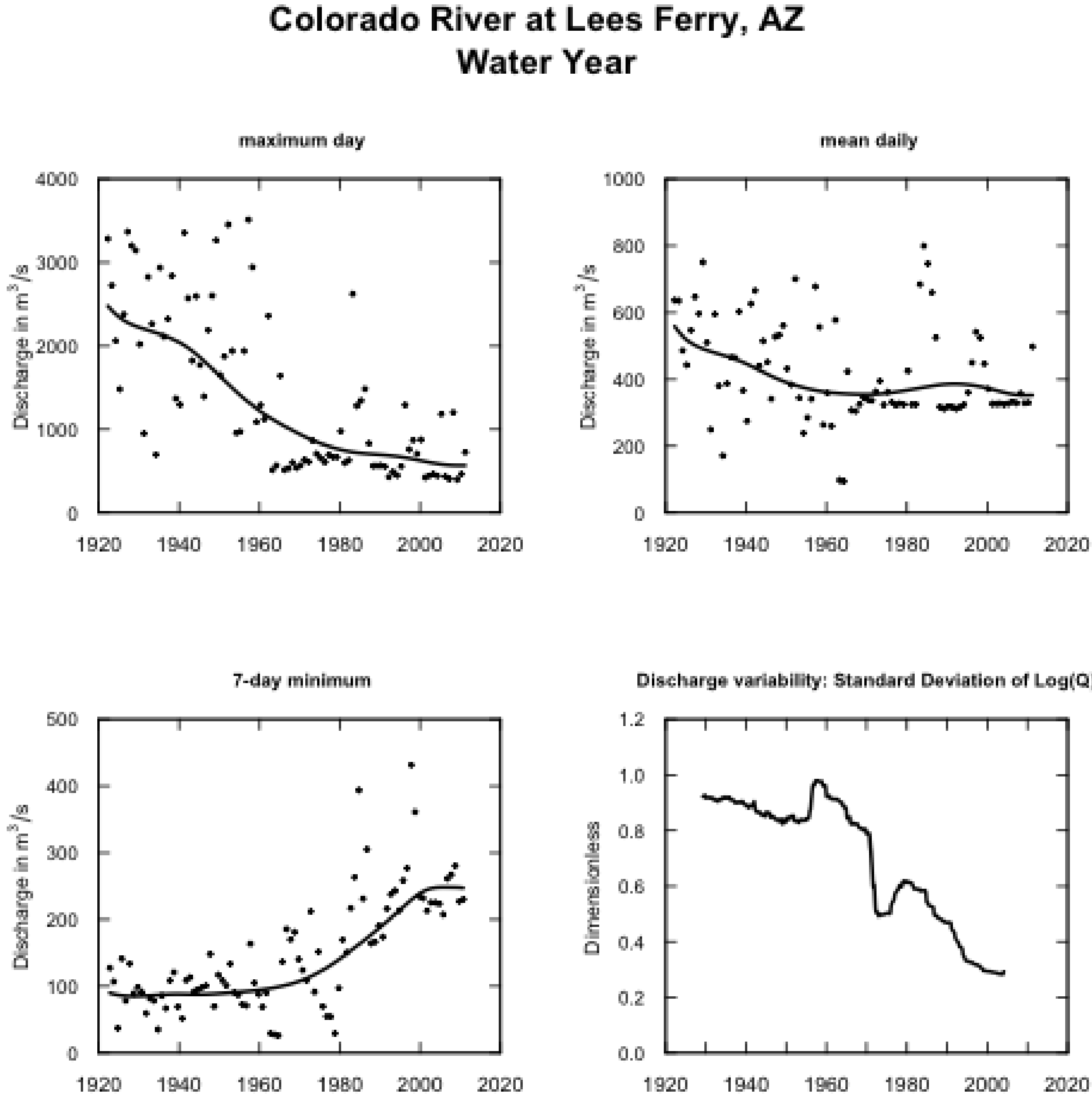
Up 245%
1970-2010



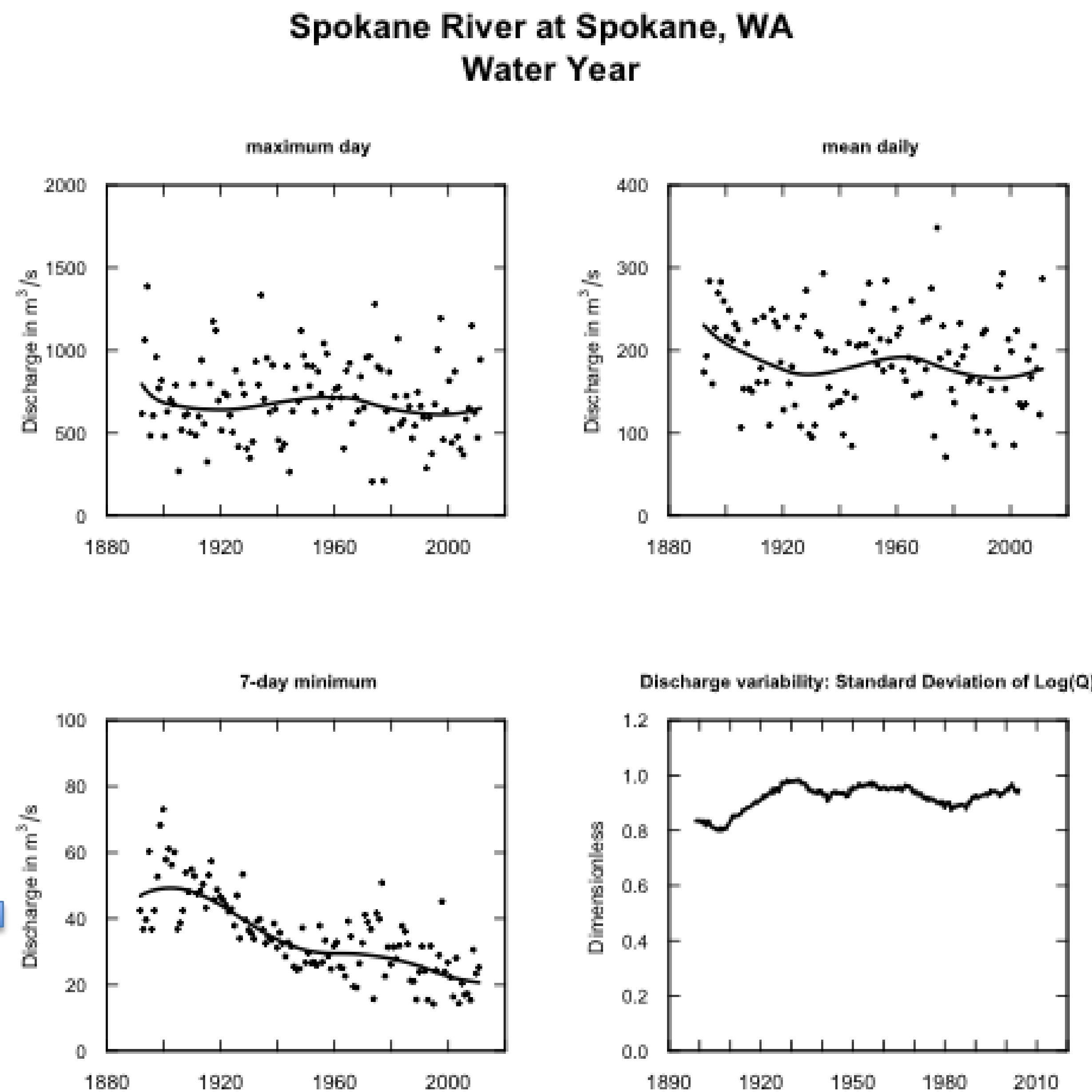
# Urbanization



# Flow regulation and diversion



# Groundwater development




**1900 – 1950**
**-0.8 % / year**

**1950 – 2010**
**-0.5 % / year**

# EGRET functionality

- Graphs and tables are self-labeling and suitable for presentation or publication.
- Reporting units selected by user.
- Works interactively or in batch.
- Structures the data and results so it is easy to go back and ask further questions.
- Data structure opens up options for many other kinds of analysis using a wide-range of functions that are part of R.
- Data frames can be easily shared among users.

# EGRET philosophy

- Get the data easily and organize it for analysis.
- Don't **only** drive to get numbers or significance levels.
- Drive to understanding, hypotheses, & descriptions of our changing world.

USGS
science for a changing world

# EGRET philosophy

There are changes all around us, now describe them to help guide how we manage our water resources!

USGS
science for a changing world

## We welcome your questions and feedback on: *the methods, the outputs, the manual*

**The R-packages, draft manual and this presentation are on the EGRET web site:**

**https://github.com/USGS-CIDA/WRTDS/wiki**

**Send your questions and feedback to:**

**egret_comments@usgs.gov**

USGS
science for a changing world