Adventures in Data:
Revisiting Historical Water Quality and Streamflow Data in the Catskills

Karen Moore & James Mayfield

Catskill Environmental Research & Monitoring Conference
October 23-24, 2014

Overview

Focus: A new approach to looking at water quality and stream flow data.

Purpose: Extract new insights from old data.

In this talk, I will discuss:

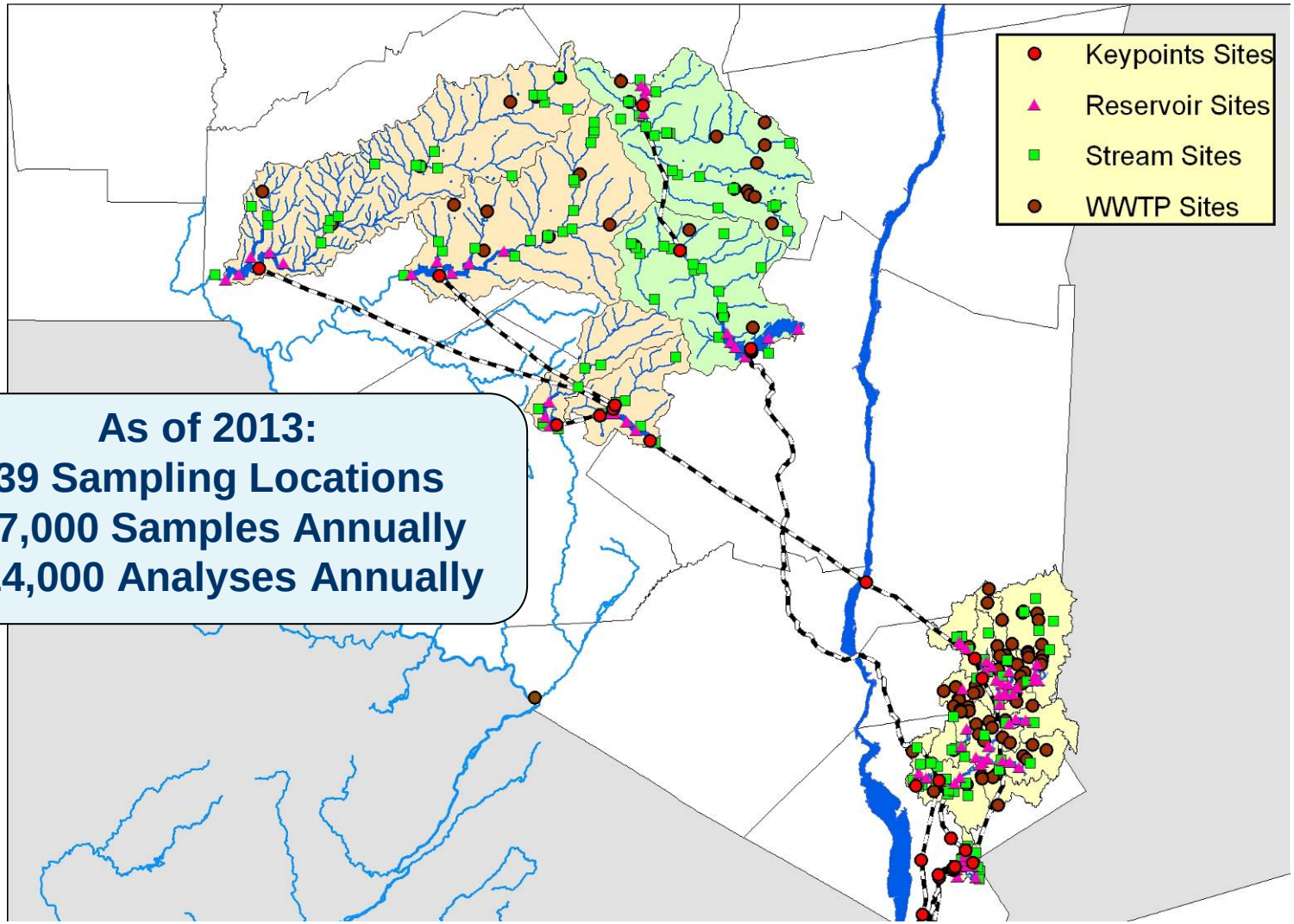
- Methods & tools developed by U.S. Geological Survey (USGS) scientists
- Examples from application of these tools to Catskill rivers

Underlying philosophy:
There is value to re-visiting the data to learn from it.

Source: http://relu.data-archive.ac.uk/media/36645/relubrief16.pdf

Watershed Water Quality Monitoring

DEP has a long-term investment in water quality monitoring throughout the watershed with 27 years of data for major river inflows to West of Hudson reservoirs.



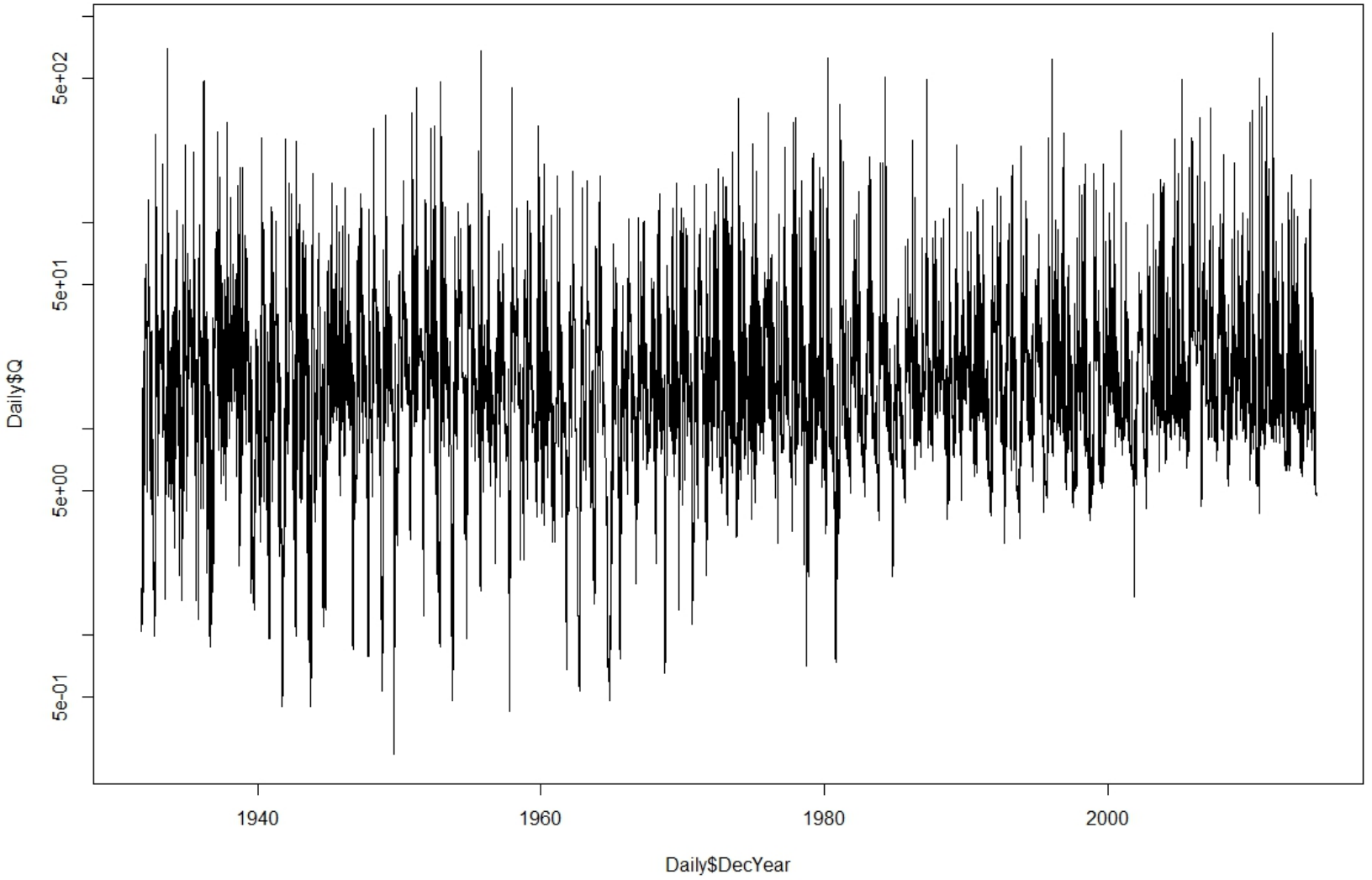
As of 2013:
439 Sampling Locations
>17,000 Samples Annually
>214,000 Analyses Annually

New Tools Developed by USGS

- R (free, “open source” software) packages courtesy of the U.S. Geological Survey:
- **dataRetrieval:** An R package that facilitates rapid acquisition of USGS daily streamflow data, water quality data, and meta data from USGS Web-repositories or from user-supplied files.
- **EGRET** (**E**xploration and **G**raphics for **R**iv**E**r **T**rends): An R package that generates graphics and summary statistics to aid in understanding the hydrologic system in a changing world.
- A tool-within-a tool: WRTDS (Weighted Regression for Time, Discharge, and Season) for river water quality analysis is part of the EGRET package.
- Value added to historical data as these tools help with the evaluation of changes from natural and unnatural causes, as well as viewing management strategies from a big-picture perspective.

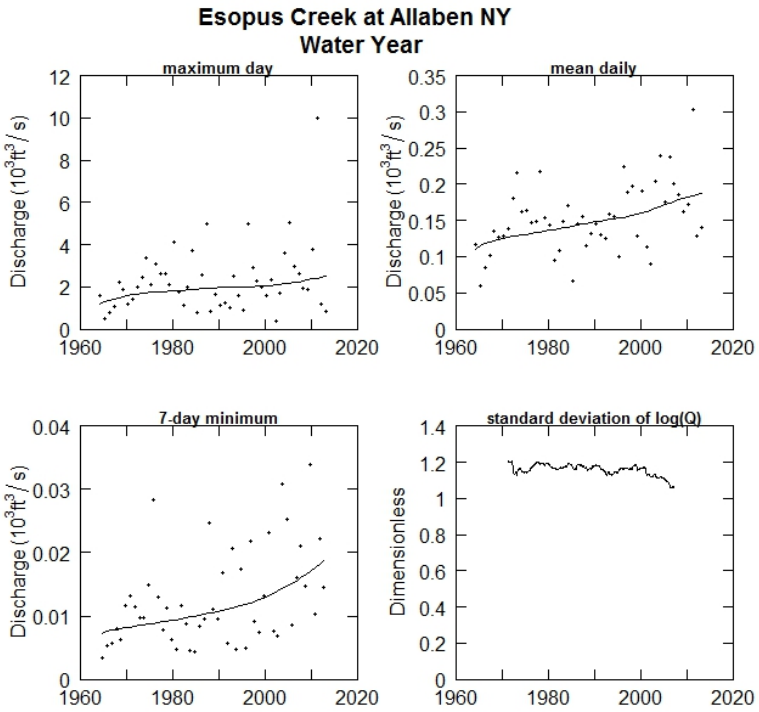
Links to some relevant papers and the program in R:
https://github.com/USGS-R/EGRET
Or type into your search engine: USGS EGRET

Esopus Creek at Coldbrook (flow in cfs)

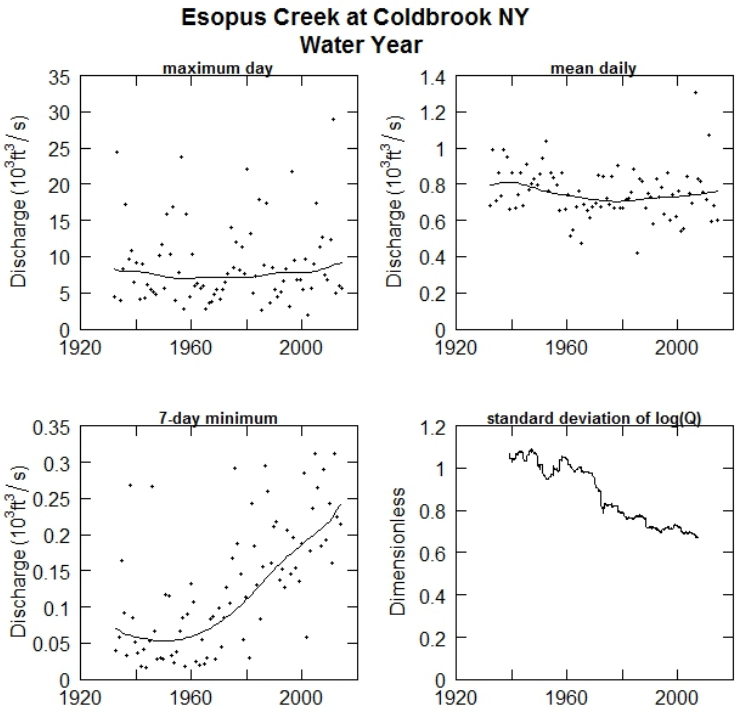


Raw output from EGRET showing mean daily flow (flow in cubic feet/sec, cfs)

Flow History Requirements for WRTDS

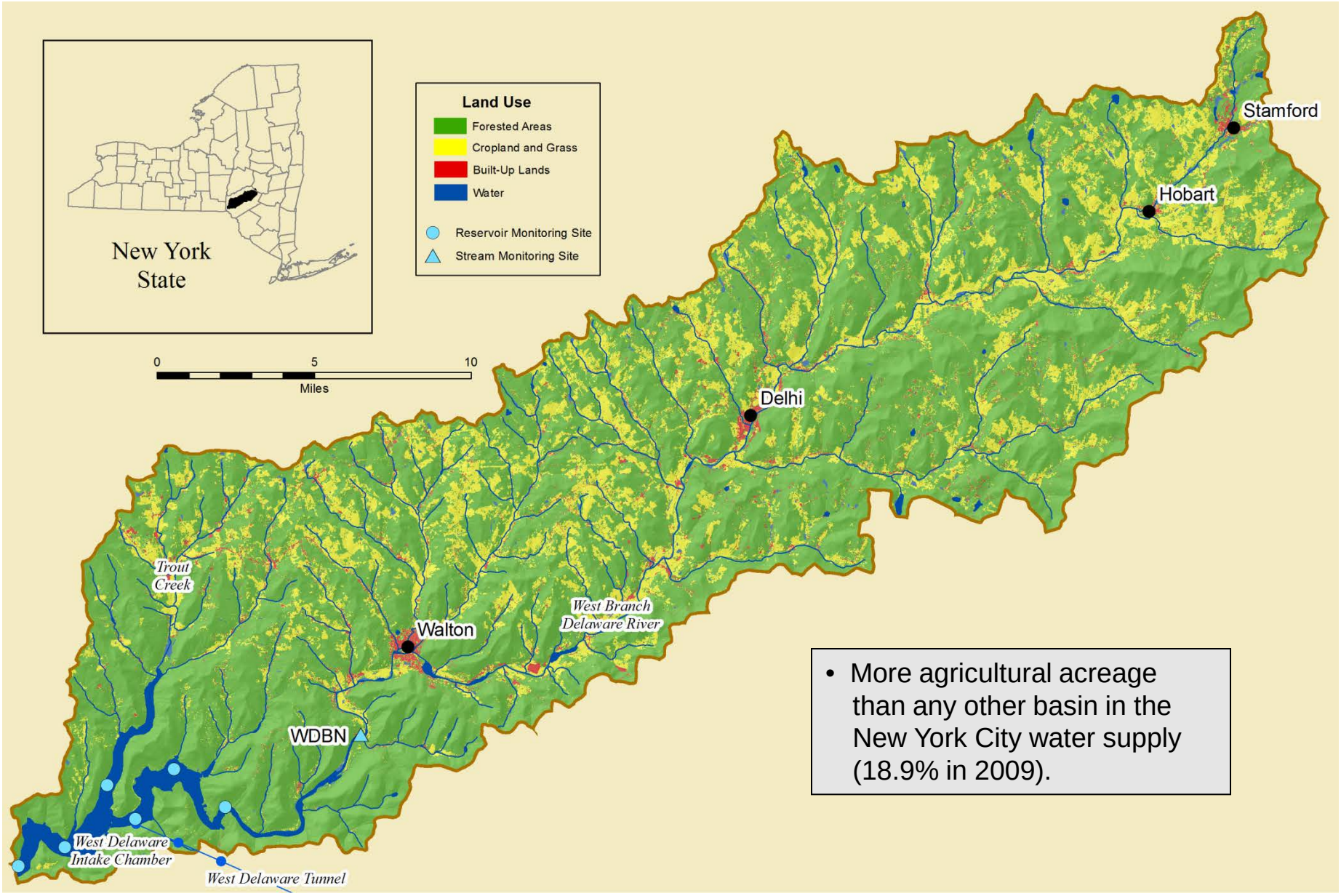


Upstream, unregulated site on Esopus Creek above Shandaken Portal. Note standard deviation plot shows stationarity, a requirement for WRTDS estimation of concentration and load (flux).



Site on Esopus Creek below Shandaken Portal. Note standard deviation plot shows non-stationarity and disqualifies it from use of WRTDS model.

Example analysis from Cannonsville



- More agricultural acreage than any other basin in the New York City water supply (18.9% in 2009).

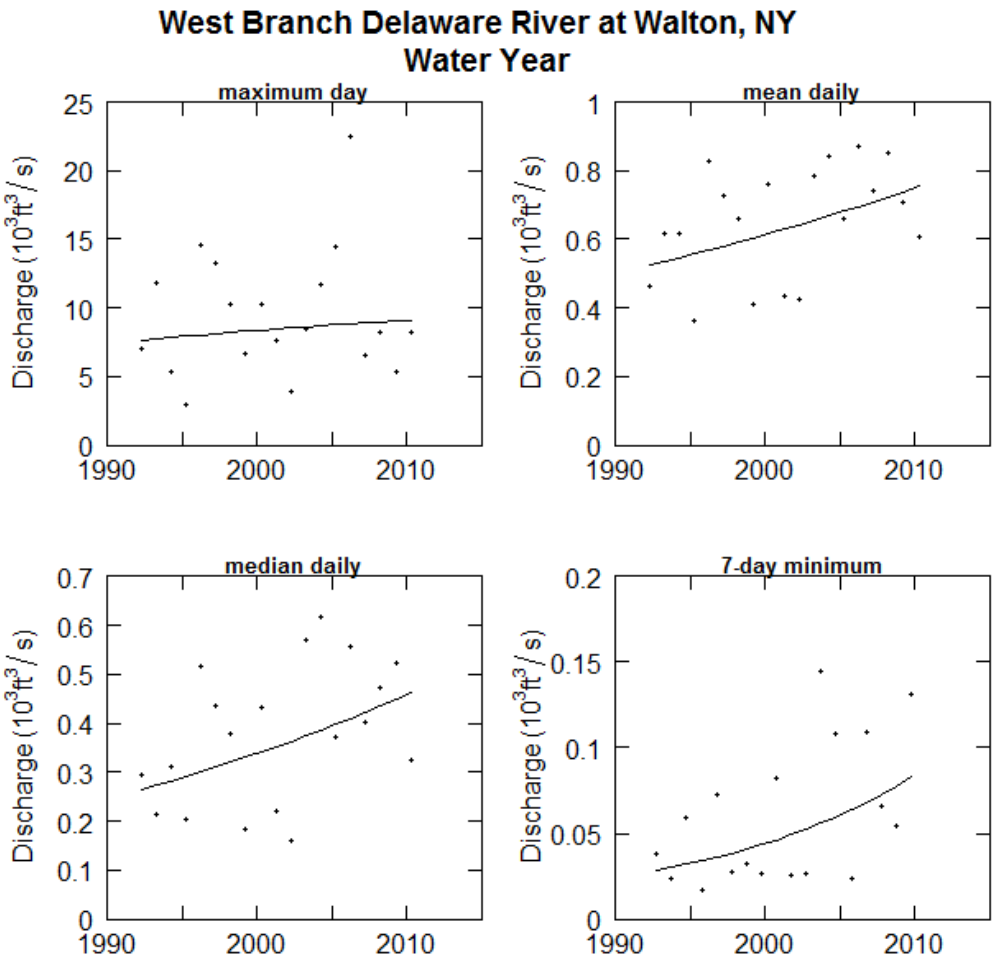
Watershed Partnership Programs

DEP manages and funds a broad spectrum of Cannonsville watershed initiatives including but not limited to:

Program	2005	2010 (Δ from 2005)
Farms with Whole Farm Plans		
Large farms	157	123 (-34)
Small farms	22	52 (+30)
Agricultural structural BMP's		
Large farms	1617	2066 (+449)
Small farms	172	365 (+193)
WWTP upgrades	5	5 (0)
Sand/Salt storage	10	10 (0)
Stream management projects (linear ft)	1200	4900
Acquired land + easements	12, 429 acres	27,165 acres (+14,736 acres)

Phosphorus reduction programs began in 1993; analyses show by 2002 Cannonsville was no longer considered a phosphorus-restricted basin.

Flow History Statistics from EGRET



- Curves show a LOWESS smooth (locally weighted scatterplot smooth), which indicates an upward trend at all levels of flow summary.
- All statistics are based on water year (Oct. 1 – Sep. 30) except the 7-day minimum, which is based on "climate year" (Apr.1 – Mar. 31) to minimize the probability that a droughts will be represented across multiple water years. Data source: USGS (NWIS).

Weighted Regression (WRTDS)

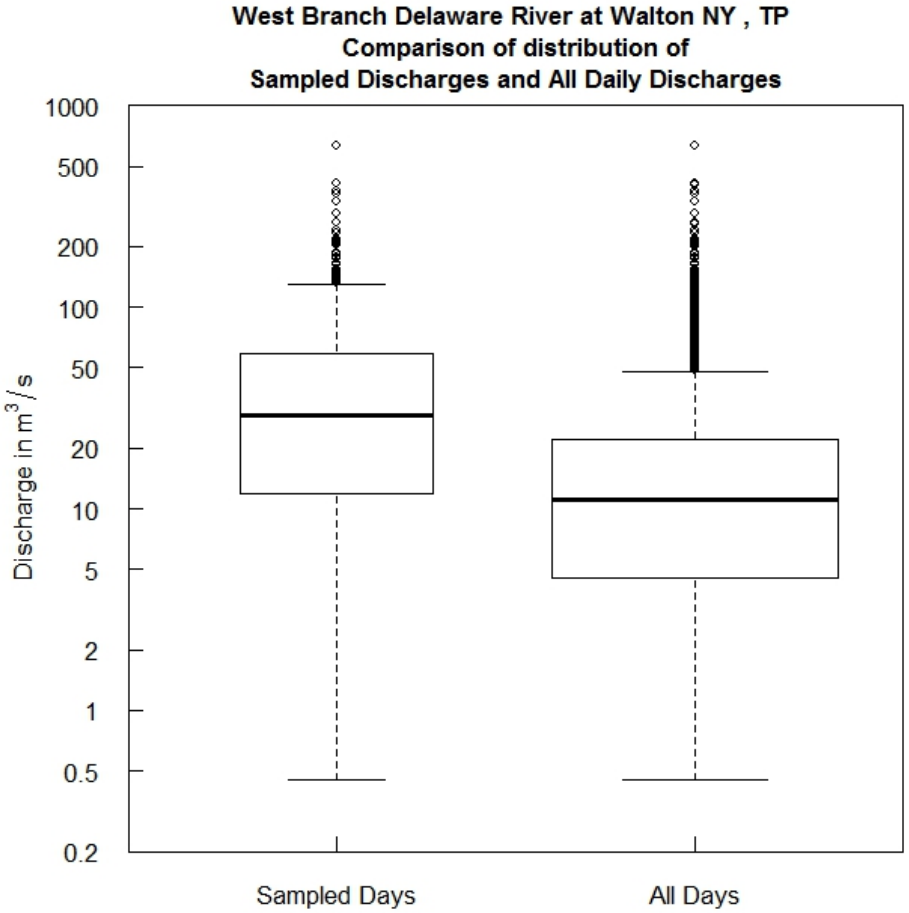
- Original paper describing approach: Hirsch et al. (2010) with more recent applications to Chesapeake Bay, Lake Champlain, and Mississippi R. watersheds.

- Requires long-term (20 yrs+ mean daily river discharge data and minimum of 200 water quality samples. Gaps in water quality should not exceed 2 yrs.

- Conceptual view: WRTDS is a locally-weighted regression where weights are based on time, discharge, and season.

- Basic idea: the weighted regression model can give a picture of the system for any given day. In addition to simple flux (load) and concentration estimates, the model calculates a “flow-normalized” flux and concentration.

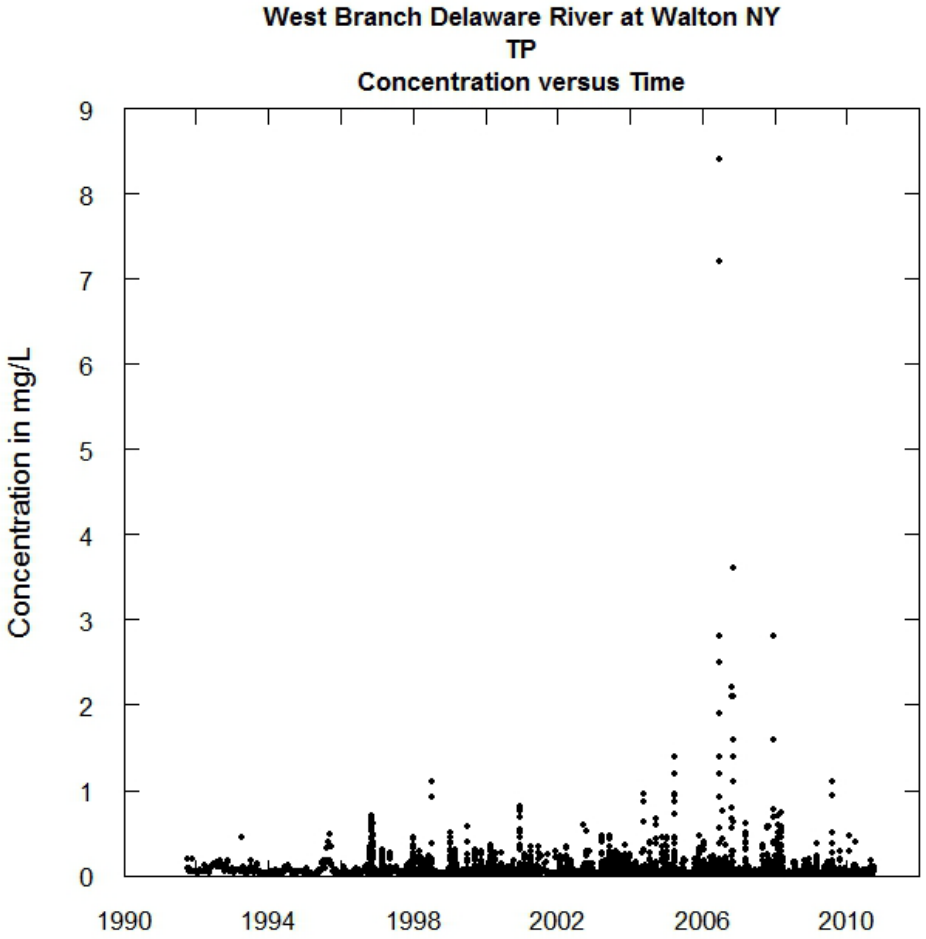
Sampling considerations



- Collecting representative samples across the range of flows is important. For this data set, high-flow events were sampled intensively and this is reflected in the distribution of discharge for sampled days (NYSDEC samples, 1991-2010).

- To obtain accurate estimates of flux (load), the best sampling strategy is to collect more samples at high discharge.

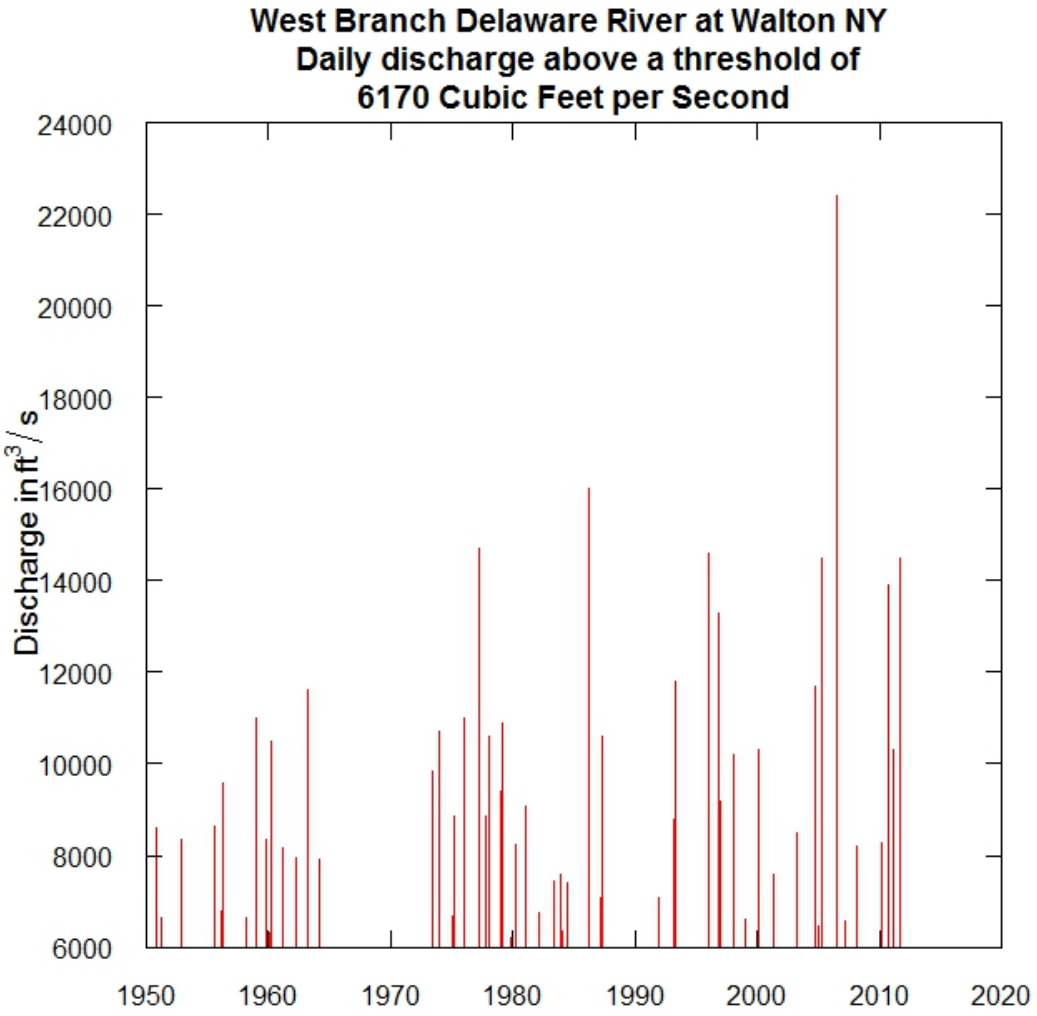
Total Phosphorus Concentration



- Flood of record on June 26-29, 2006 resulted in highest observed TP concentration. Since TP reflects sediment-bound P, it isn't surprising that the flood yielded the highest TP, even after watershed protection programs had a demonstrated impact on P-reduction.

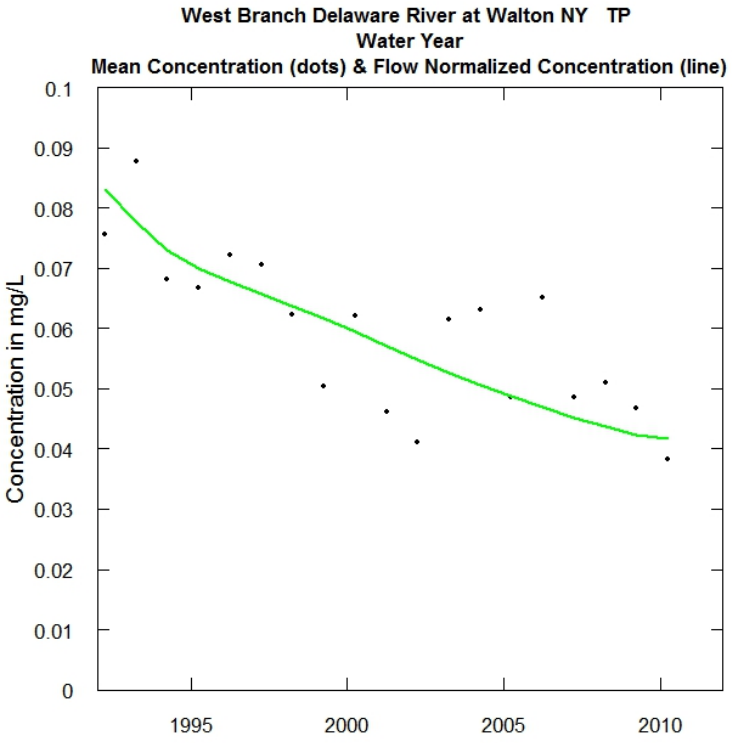
- Note scale is influenced by this anomalous event in 2006. The EGRET package allows the user to change the units. TP is typically expressed in µg L^{-1} units.

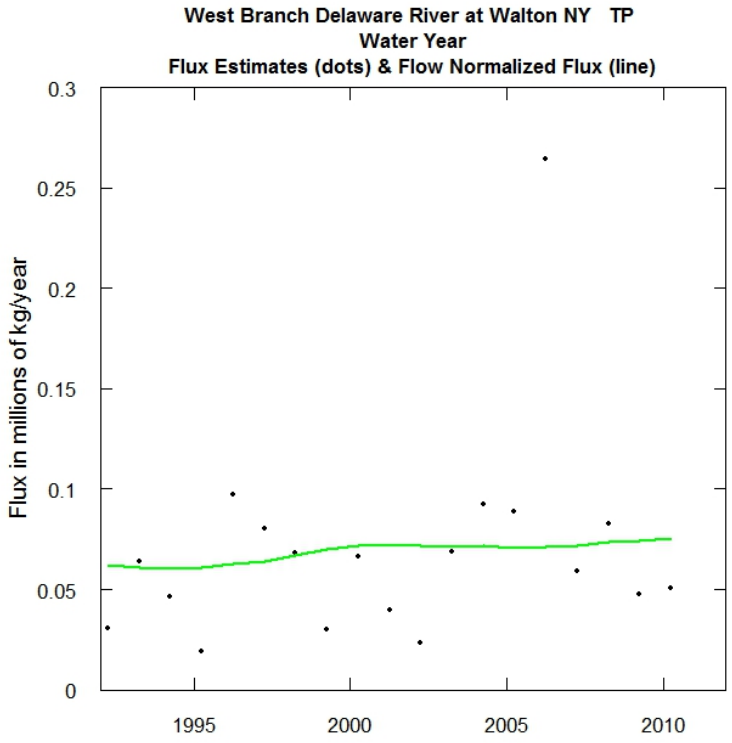
Flow History in EGRET



EGRET allows flexibility in screening data. Only those events at or above flood stage are shown here.

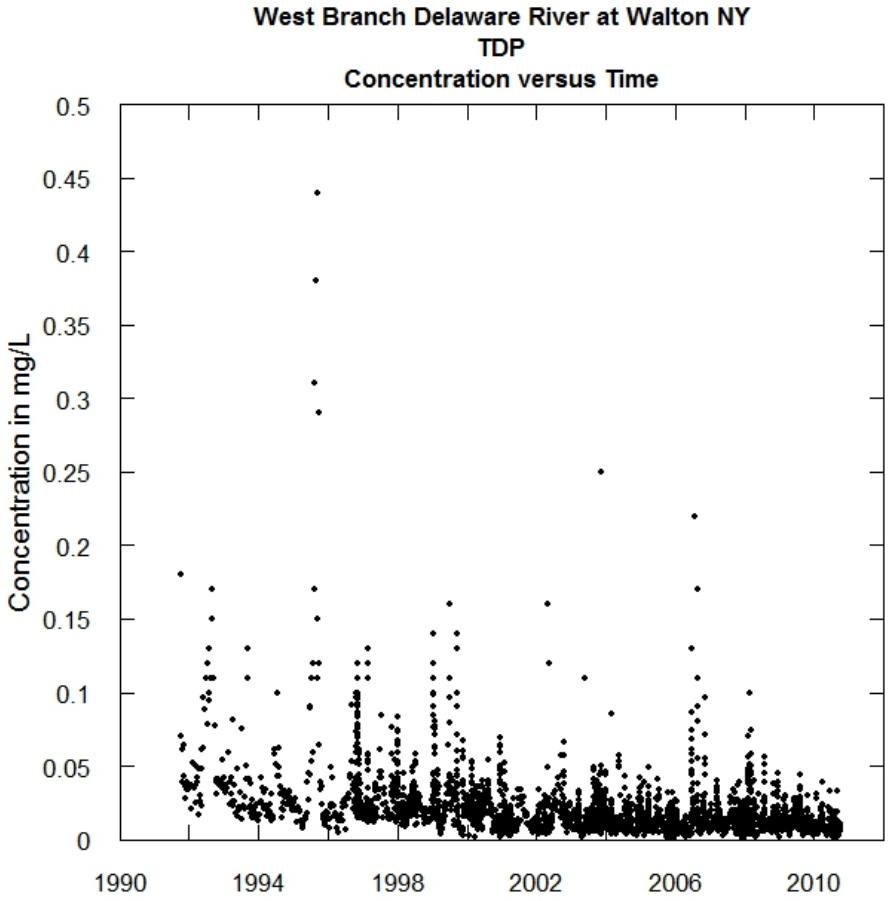
Total Phosphorus – Flow-Normalized





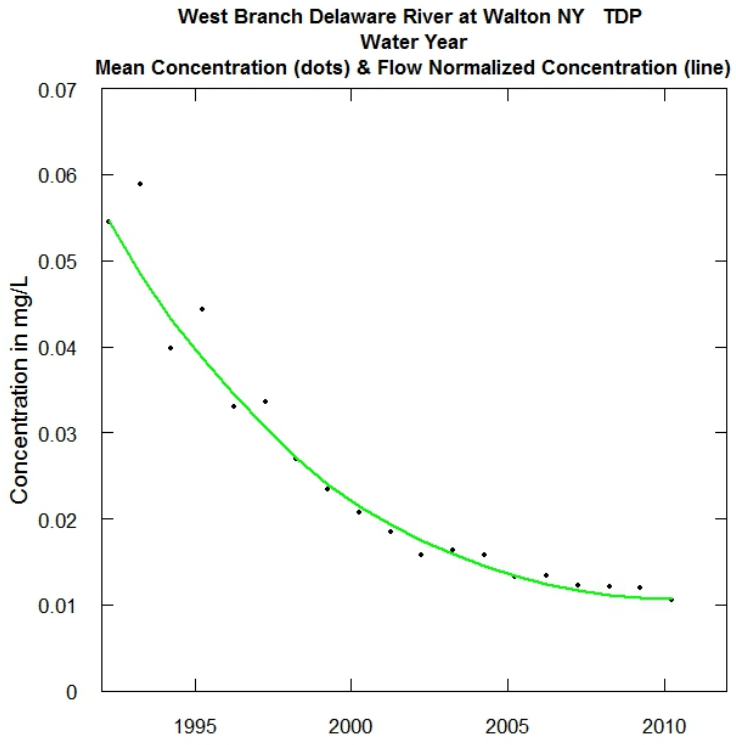
Flow-normalized concentration and flux give a picture of progress or effectiveness of programs by filtering out the year-to-year variations in discharge (green line), while estimates of concentration and flux alone (dots showing mean annual concentration and flux) are important for understanding the actual history of river water quality.

Total Dissolved Phosphorus Concentration

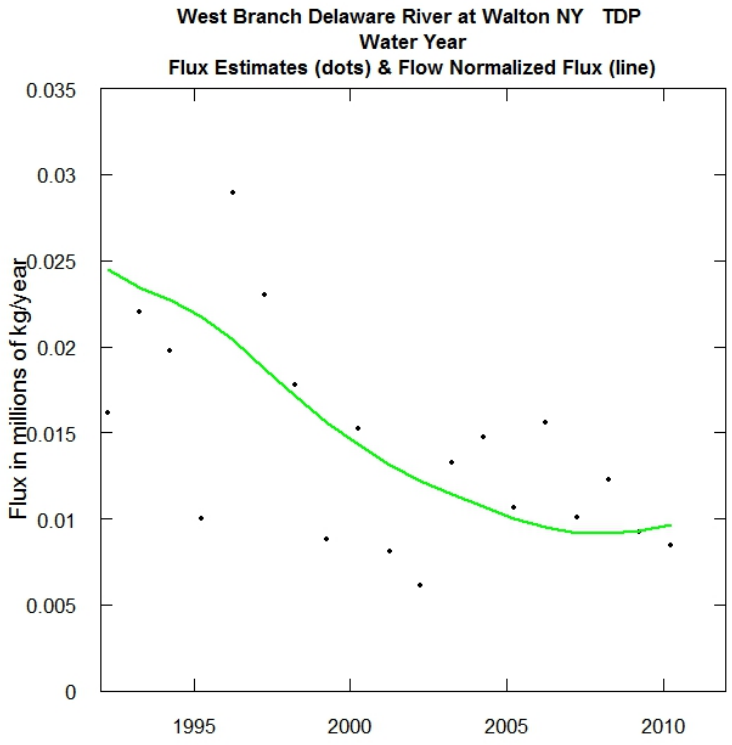


- Total Dissolved Phosphorus gives a clearer picture of the impacts of point-source control measures. The storm event of June 2006 still shows a pulse of incoming P, but it is lower than some events in the period before phosphorus reduction measures were implemented.

Total Dissolved Phosphorus – Flow-Normalized



Flow-normalized TDP concentration dropped by 78% from 1993-2010, a decline of 4.6% yr^{-1}.



Flow-normalized TDP flux dropped by 59% from 1993-2010, a decline of 3.5% yr^{-1}.

Total Dissolved Phosphorus Summary

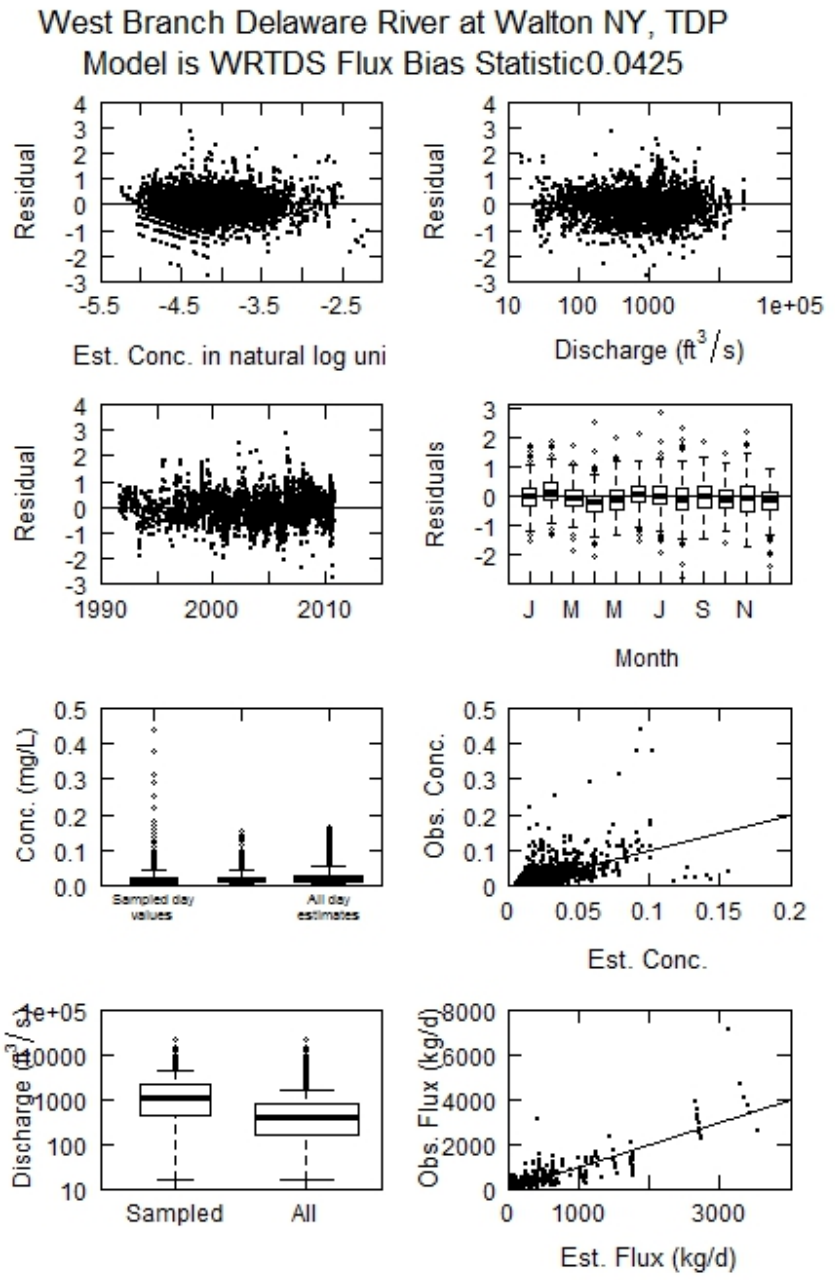
Change in Flow-normalized Concentration and Flux for selected periods as estimated by WRTDS for West Branch Delaware River at Walton, NY

Period	ΔConc. ($\mu g\ L^{-1}$)	% change	ΔFlux ($10^3\ kg\ yr^{-1}$)	% change
2003-2005	-2.5	-16%	-1.4	-12%
2005-2007	-1.7	-6.5%	-0.85	-8.5%
2007-2010	-0.9	-2.6%	+0.48	+5.2%

- It is possible to see trends in flow-normalized concentration and flux that are opposite in sign.
- If, for example, there were large decreases in point sources but increases in non-point sources of TDP associated with high flow events, we could expect to see a negative trend in concentration and a positive trend in flux.

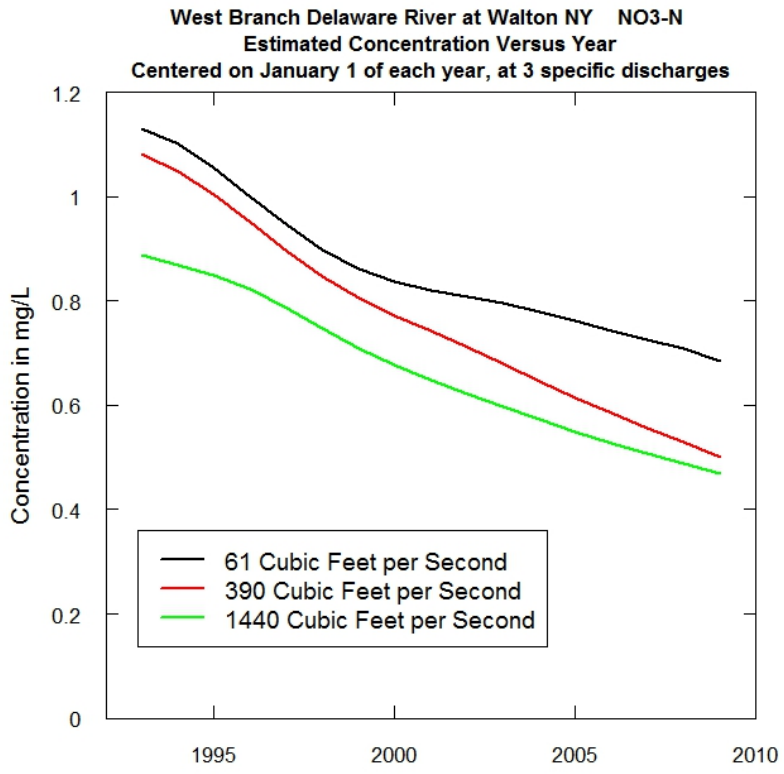
WRTDS Model Diagnostics

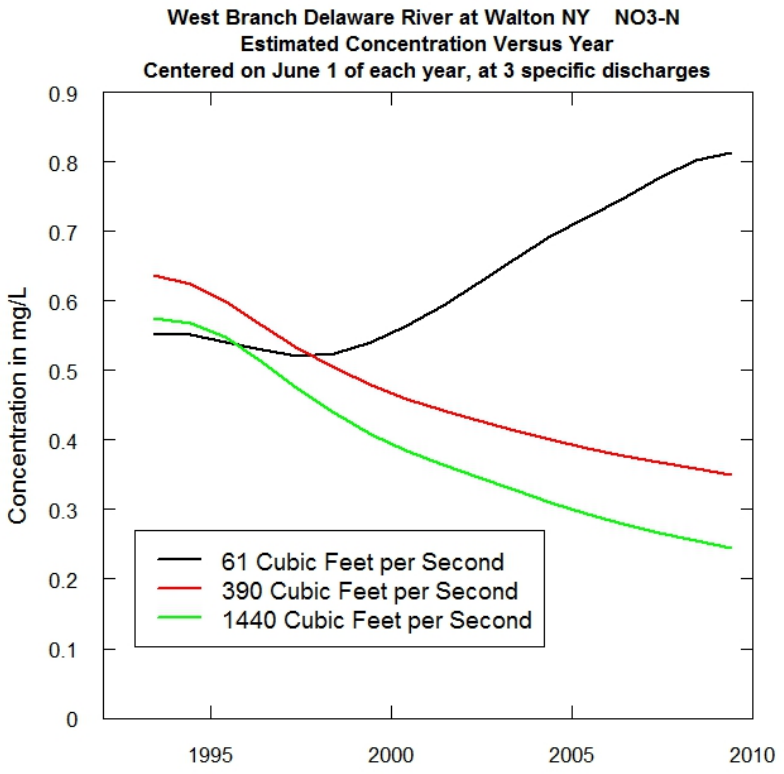
- All models have biases.
- The WRTDS diagnostic plots give clues to how well the model performs.
- The Flux Bias Statistic indicates that flux is overestimated by about 4%.
- If bias is too large, then consider using a different model.



Concentration – Flow Relationships

EGRET has flexible time periods for data viewing. In this view January (left) and June (right) nitrate concentration at low flow (10%ile in black), median flow (50%ile in red), and high flow (90%ile in green) is averaged for a 60-day period for WY 1992-2010.





The explanation for increases in nitrate at baseflow in the summer months needs to be pursued further by looking at additional data.

Where to Next?

- There are at least 15 additional sites with a 27-year record of NYCDEP water quality data that can be examined using EGRET.
- Further exploration of connections between water quality patterns, changes in the watershed, and program implementation is expected to give insights relevant to management.
- One of the advantages of this approach is that the data can be quickly re-visited as the record builds over time.