Quantile Regression

Caren Marzban

Applied Physics Lab., Department of Statistics
Univ. of Washington, Seattle, WA, USA 98195
CAPS, University of Oklahoma, Norman, OK

Abstract

The prediction from most regression models - be it multiple regression, neural networks, trees, etc. - is a point estimate of the conditional mean of a response (i.e., quantity being predicted), given a set of predictors. However, the conditional mean measures only the "center" of the conditional distribution of the response. A more complete summary of the conditional distribution is provided by its quantiles. The 0.5 quantile (i.e., the median) can serve as a measure of the center, and the 0.9 quantile marks the value of the response below which resides 90quantile and the 0.05 quantile serves as the 90thereby conveying uncertainty. Quantiles arise naturally in environmental sciences. For example, one may desire to know the lowest level (e.g., 0.1 quantile) of a river, given the amount of snowpack; or the highest temperature (e.g., the 0.9 quantile), given cloud cover. Recent advances in computing allow the development of regression models for predicting a given quantile of the conditional distribution, both parametrically and nonparametrically. The general approach is called Quantile Regression, but the methodology (of conditional quantile estimation) applies to any statistical model, be it multiple regression, support vector machines, or random forests. In this talk, the principles of quantile regression are reviewed and the methodology is illustrated through several examples. The technique and the examples display many of the features common in both machine learning and statistics.

Motivation

Q: Given data $y_i, i = 1, 2, ..., n$, what prediction α minimizes MSE

$$E = \frac{1}{n} \sum_i^n \epsilon_i^2 = \frac{1}{n} \sum_i^n (y_i - \alpha)^2 \ ?$$

A: The sample mean of y. Proof:

$$\frac{dE}{d\alpha} \sim \frac{1}{n}\sum_i^n (y_i - \alpha) \sim \frac{1}{n}\sum_i^n y_i - \alpha \quad \Longleftrightarrow \quad \hat{\alpha} = \frac{1}{n}\sum_i^n y_i \ .$$

Q: Given data $y_i, i = 1, 2, ..., n$, what prediction α minimizes the MAE

$$E = \frac{1}{n}\sum_i^n |\epsilon_i| = \frac{1}{n}\sum_i^n |y_i - \alpha| \ ?$$

A: The sample median of y. Proof: Similar.

Q: Given data (x_i, y_i), what prediction $\hat{y}(x) = \alpha_0 + \alpha_1 x$ minimizes MSE

$$E = \frac{1}{n}\sum_i^n \epsilon_i = \frac{1}{n}\sum_i^n [y_i - (\alpha_0 + \alpha_1\ x_i)]^2?$$

A: The conditional sample mean of y. I.e.,

$$\hat{y}(x) = \hat{\alpha}_0 + \hat{\alpha}_1 x = \text{mean of y, given } x$$

$E[y|x]$ = conditional mean = location of conditional distribution.
Width of conditional distribution $\sim$ uncertainty.

Why conditional mean? Why not conditional median?

And if median (i.e., 0.5 quantile), why not other quantiles?

Why not output = 0.1, ..., 0.5, ..., 0.9 conditional quantile (given x)?

Hence Quantile Regression (QR).

Effectively, QR produces the whole conditional distribution of y.

With it one can estimate and conduct inference on conditional quantile functions. ($\sim$ Extreme Value Theory).

90% Prediction Interval (not CI) = 0.95 quantile - 0.05 quantile.

History

Google counts:

meteorology regression	655,000
meteorology "artificial intelligence"	495,000
meteorology "neural networks"	153,000
meteorology "quantile regression"	296

QR: Roger Koenker and Gilbert Bassett (1978), Econometrica.

QR Neural Nets: James W. Taylor (2000), Journal of Forecasting.

QR Forests: Nicolai Meinshausen (2006), Jour. of Machine Learning Research.

Nonparametric QR: Takeuchi, Le, Sears, Smola, (2006), Jour. of Machine Learning Research.

_____

In meteorology:

QR: John Bjrnar Bremnes (2004): Probabilistic forecasts of precipitation in terms of quantiles using NWP model output. Mon. Wea. Rev.

Friedrichs and Hense (2007): Statistical Downscaling of Extreme Precipitation Events Using Censored Quantile Regression. Mon. Wea. Rev.

Jeremie Juban , Lionel Fugon, George Kariniotakis (2007): Probabilistic short-term wind power forecasting based on kernel density estimators. European Wind Energy Conference. (QR Forests)

Quantile Regression (Forests)

Instead of minimizing $\Sigma_i^n[y_i - (\alpha_0 + \alpha_1\ x_i)]^2$, minimize

$$\sum_i^n f(y_i - (\alpha_0 + \alpha_1\ x_i))$$

where

$$f(y - q) = \begin{cases} \beta\ (y - q) & y \geq q \\ (1 - \beta)\ (q - y) & y < q \end{cases}$$

to obtain the βth quantile.

Instead of

$$E[y|x] = \alpha_0 + \alpha_1 x$$

QR gives, for each quantile

$$Q[y|x] = \alpha_0 + \alpha_1 x$$

Nonparametric QR gives

$$Q[y|x] = \text{splines}$$

QR Forests* give

$$Q[y|x] = \text{piece-wise constant}$$

* Breiman (2001): Take a bootstrap sample, develop a regression tree, repeat, average the outputs over the samples.

<center>Examples</center>

1) No. tornadoes/month, 1950-1998 (technically, not regression).
12-month period filtered out with running median filter.
Courtesy of Joseph Schaefer, Storm Prediction Center.

2) This month's vs. last month's number of tornadoes.

3) Global temperature, 1851-1996 (Ibid).
http://www.image.ucar.edu/GSP/Data/Climate/observed.dat

4) Max water-level (above NAVD88) vs. Max 1-hr change, 23 sites in south-western Louisiana and southeastern Texas following Hurricane Rita, Sep 2005.
http://pubs.usgs.gov/ds/2006/220/ (Table 3).

5) Six-minute obs. water-level vs. water temperature, San Fransisco, CA,
12/29/07-12/31/07
http://tidesonline.nos.noaa.gov/data_read.shtml?station_info=9414290+San+Franci

6) Annual stream flow vs. prcp, 1948-1997, Big Sioux, Dell Rapids, SD.
Courtesy of Dennis Lettenmaier, Univ. of Washington.

7) Avg. (over 1931-1960) Jan. daily-min. temp. for 56 US cities, vs. lat. lon.

Code: R (packages: splines, quantreg, quantregForest).

Plot 0.05, 0.25, 0.5, 0.75, 0.95 quantiles, and the least-square fit (dashed).

Disclaimer: All substantive conclusions are cursory.

Tornado Time Series

(Intercept)   **x**
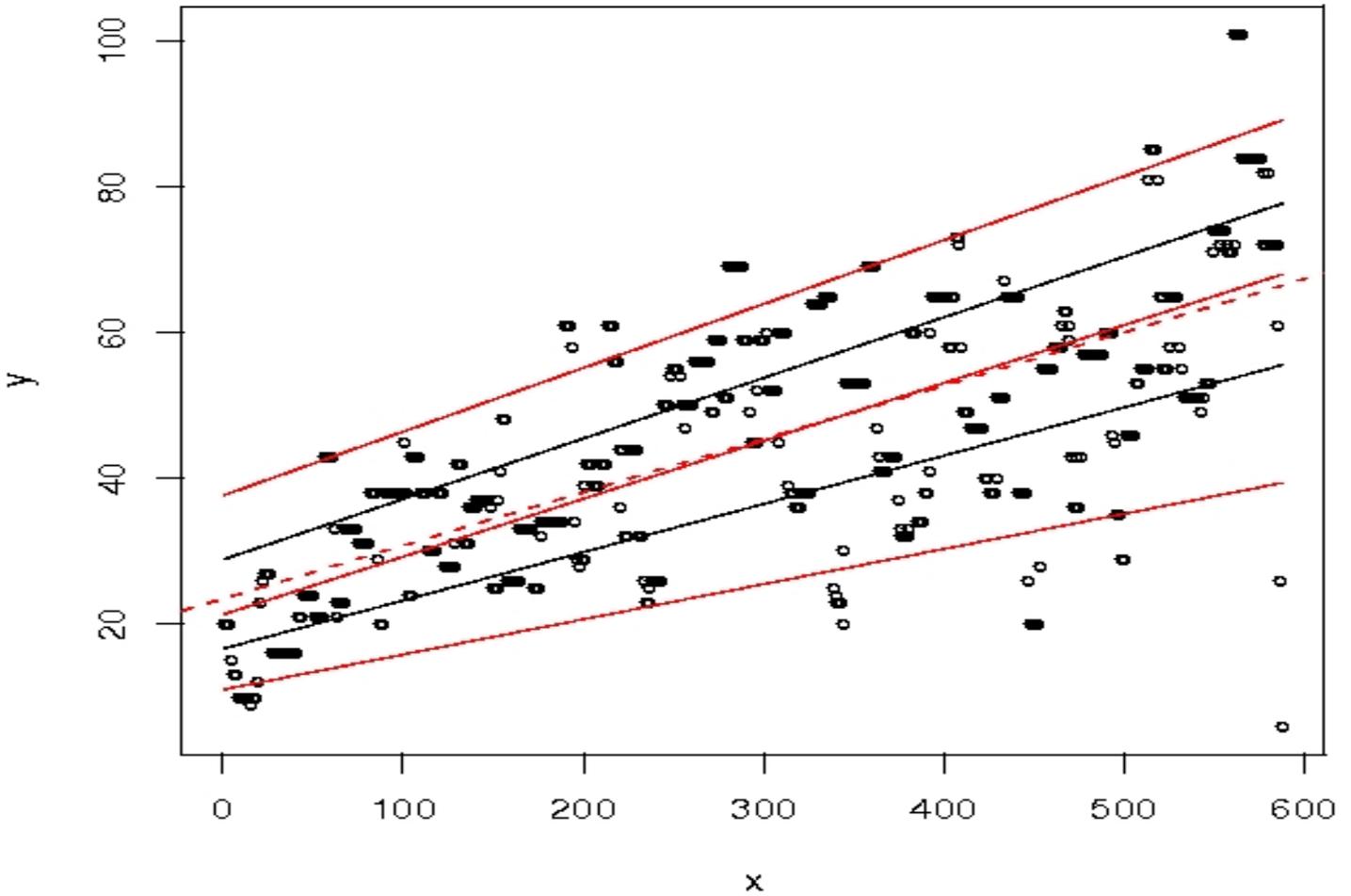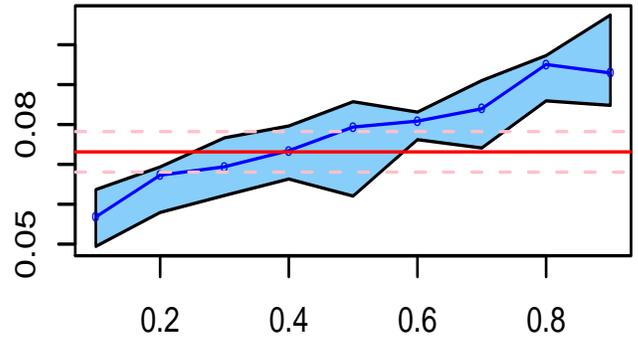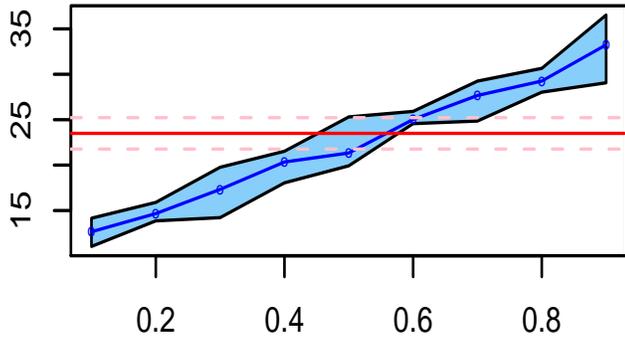
No. of tornadoes is increasing with time.
The quantiles fan out slightly.
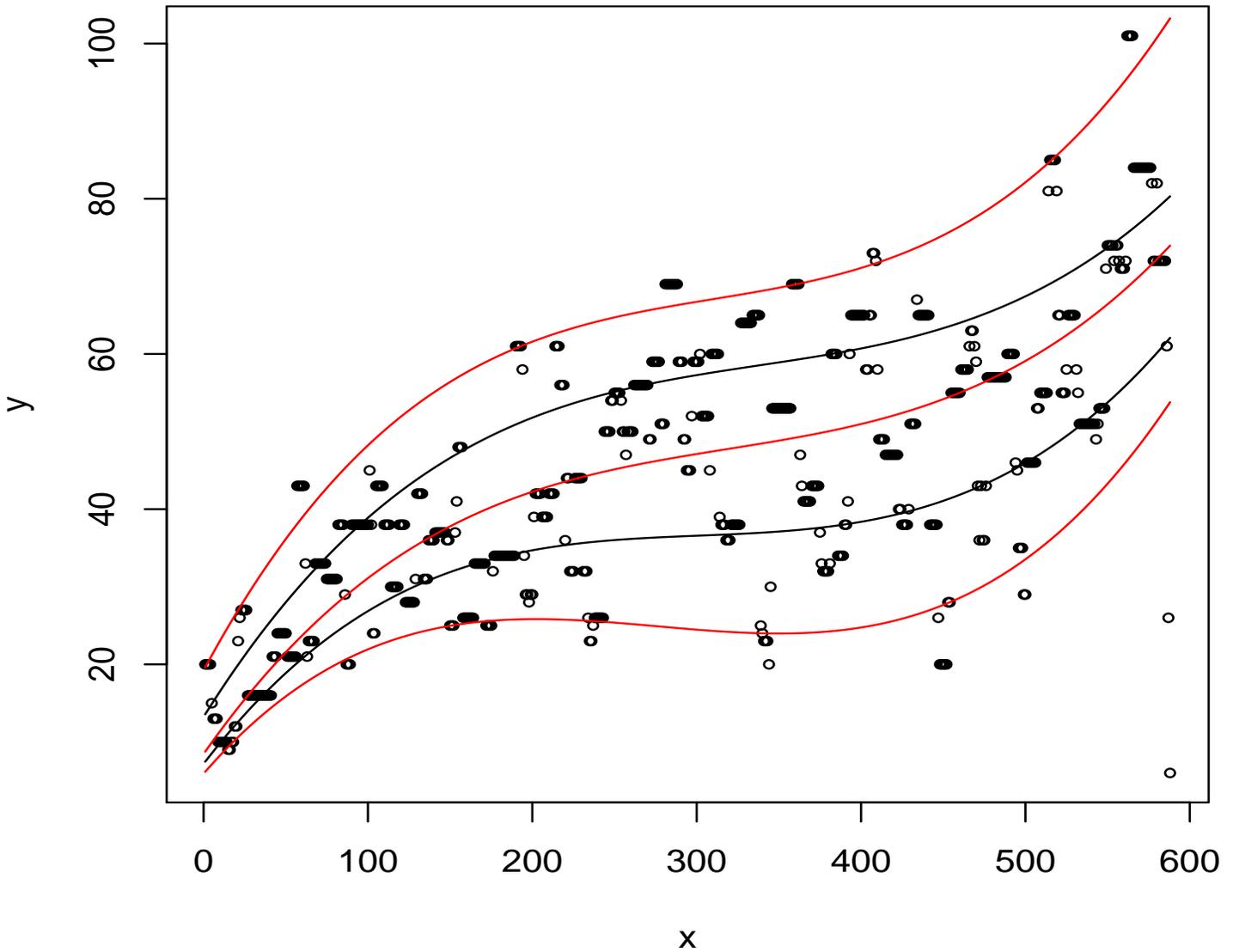
The max (i.e., 0.95 quantile) number of tornadoes grows faster than the min (i.e., 0.05 quantile) number of tornadoes.

More realistically, min is constant in time but max gows?

This Month's vs. Last Month's No. of Tornadoes
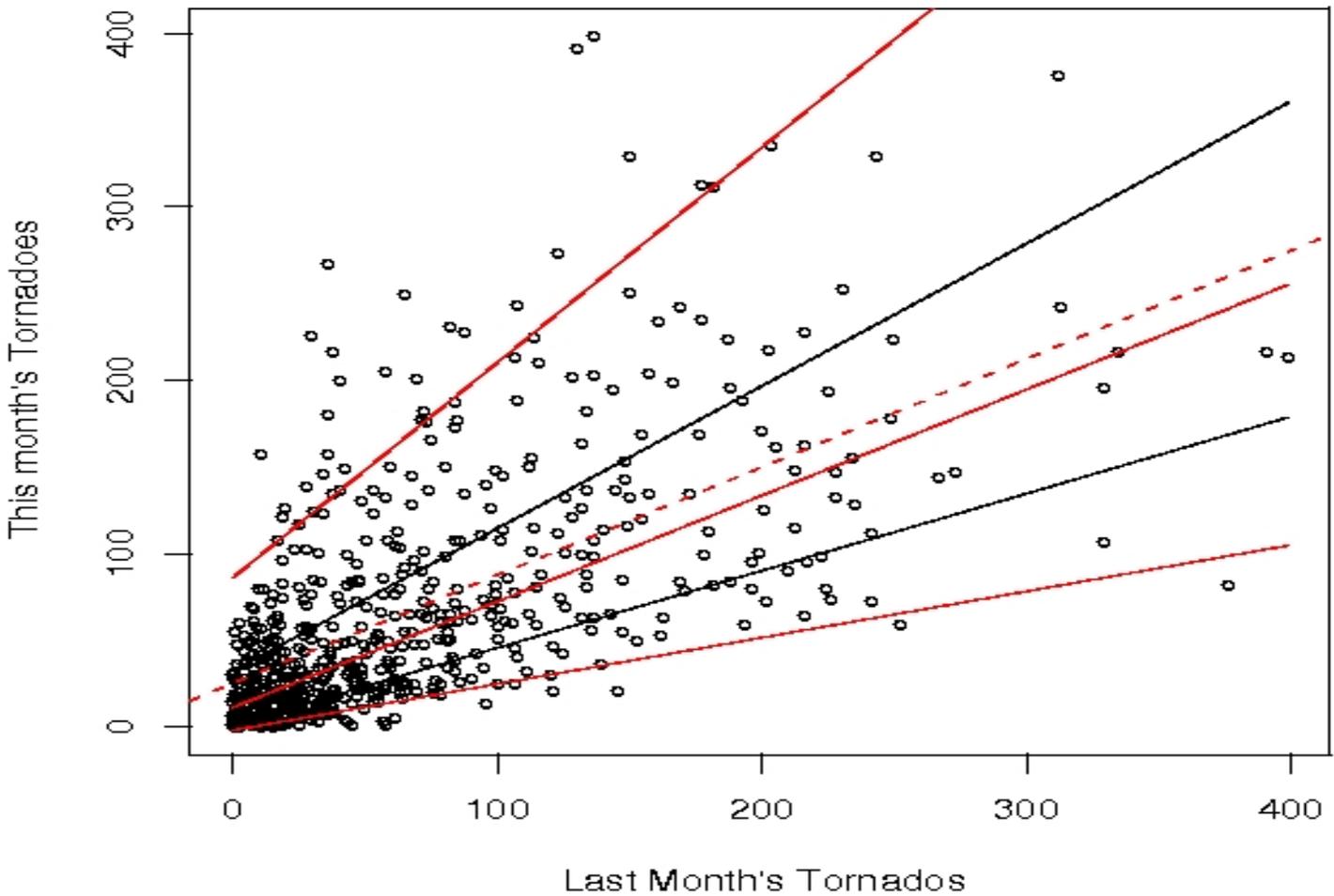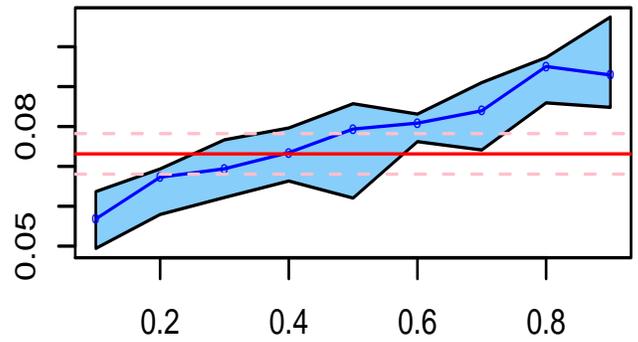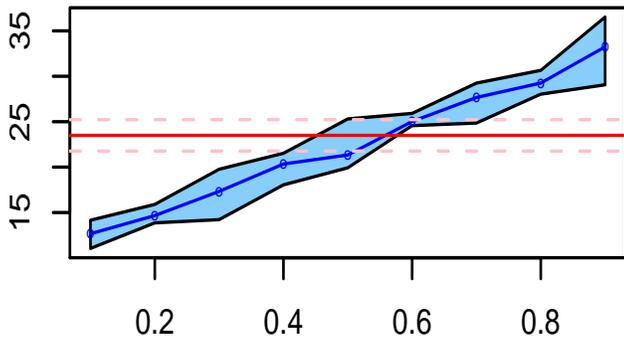
More/less tornadoes yesterday means more/less tornadoes today.

Center of distribution shifts up, but it spreads out too.

Median below mean, etc., i.e., distribution has a long tail on the high-y side.
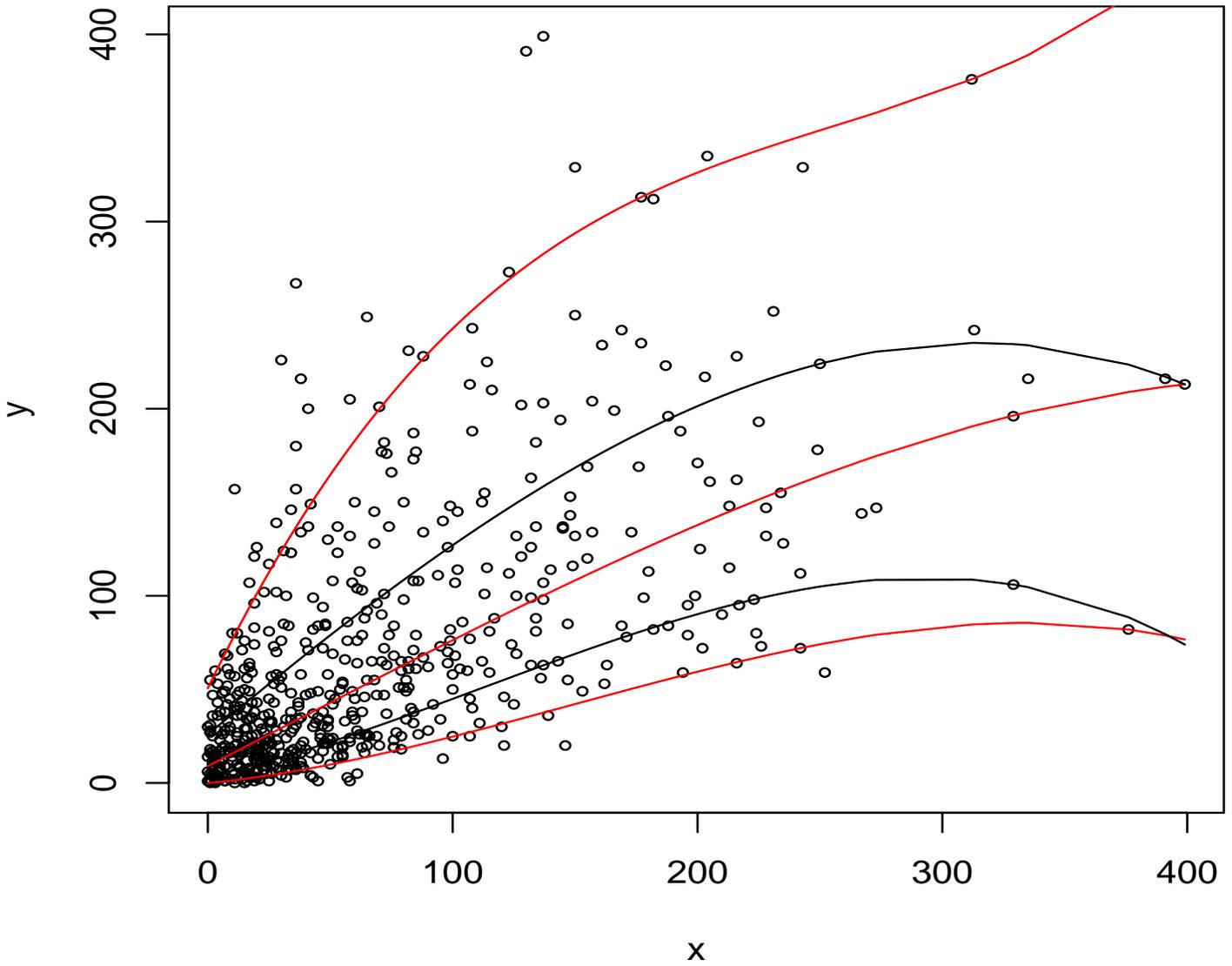
Shape of conditional distribution of y changes with x.

(Intercept)

X
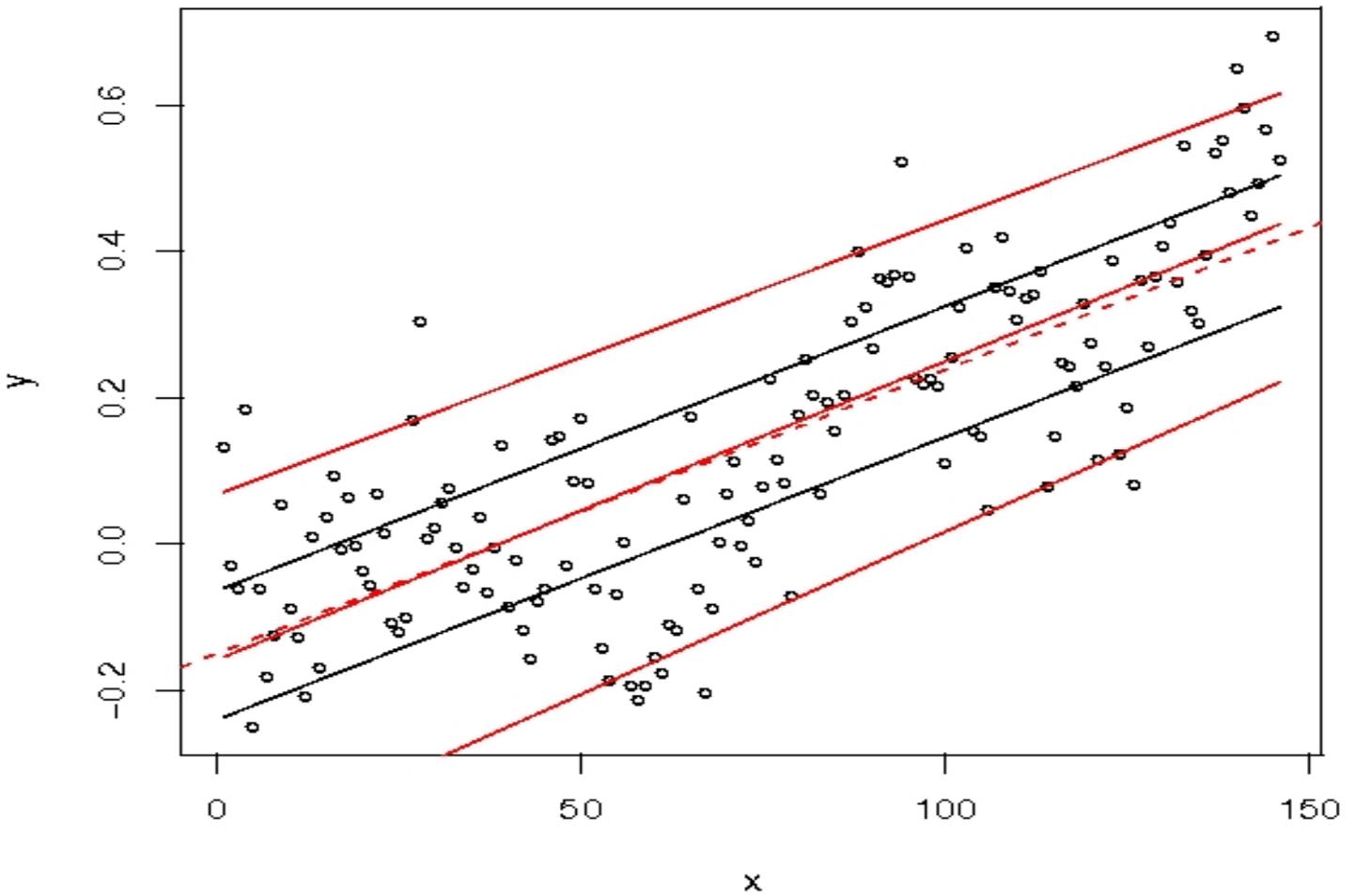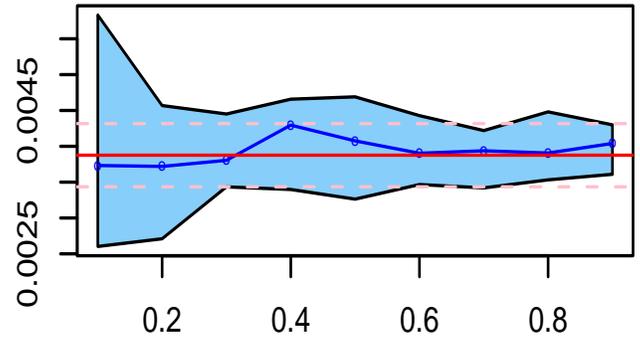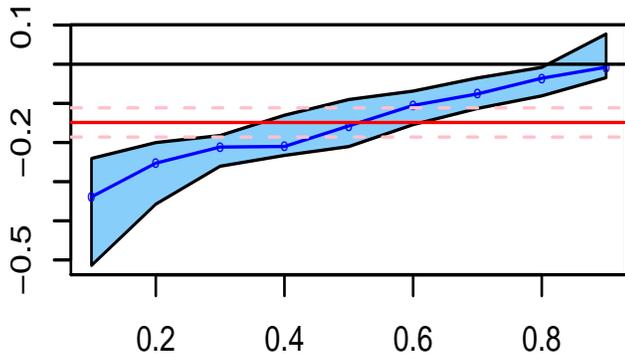
Global temperature is increasing with time.

No fanning, i.e., whole conditional distribution shifts up.

See next.
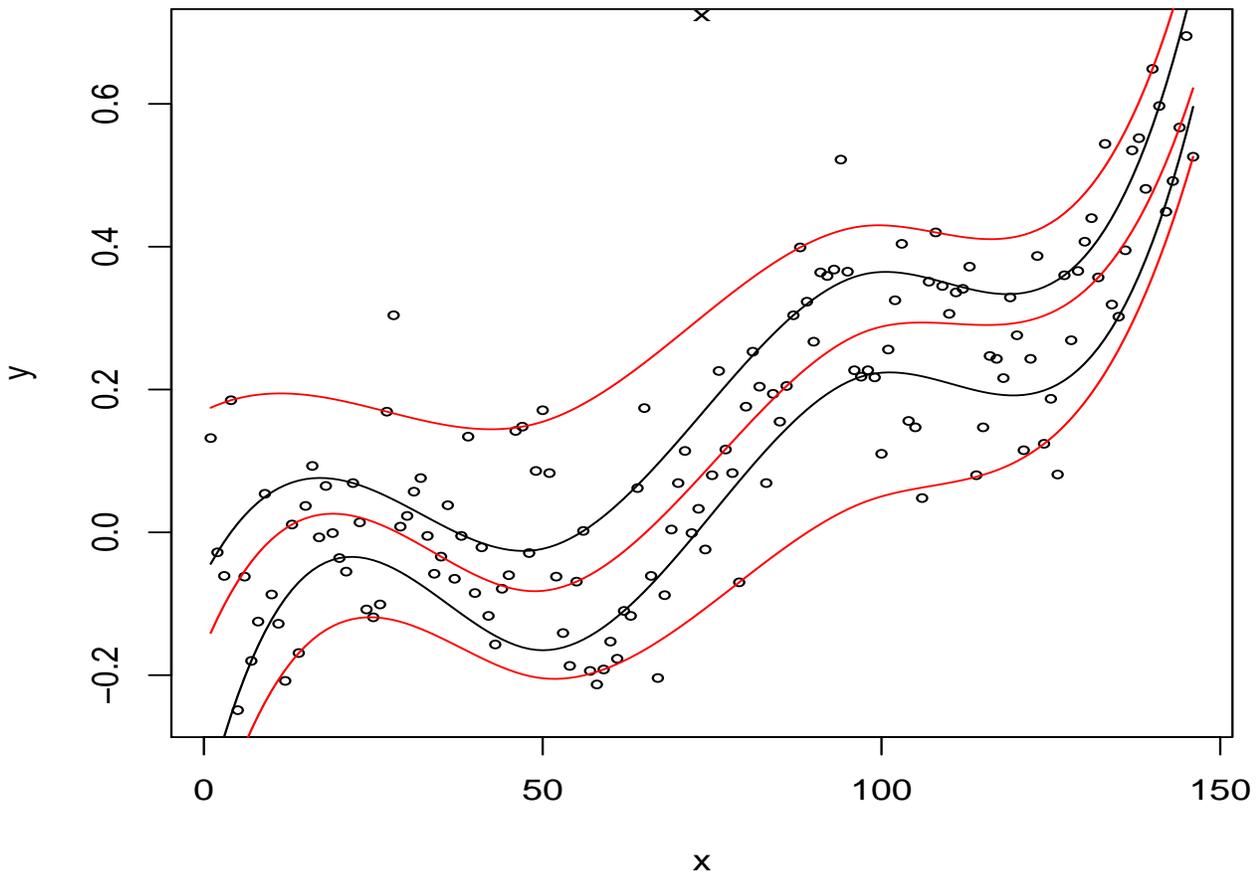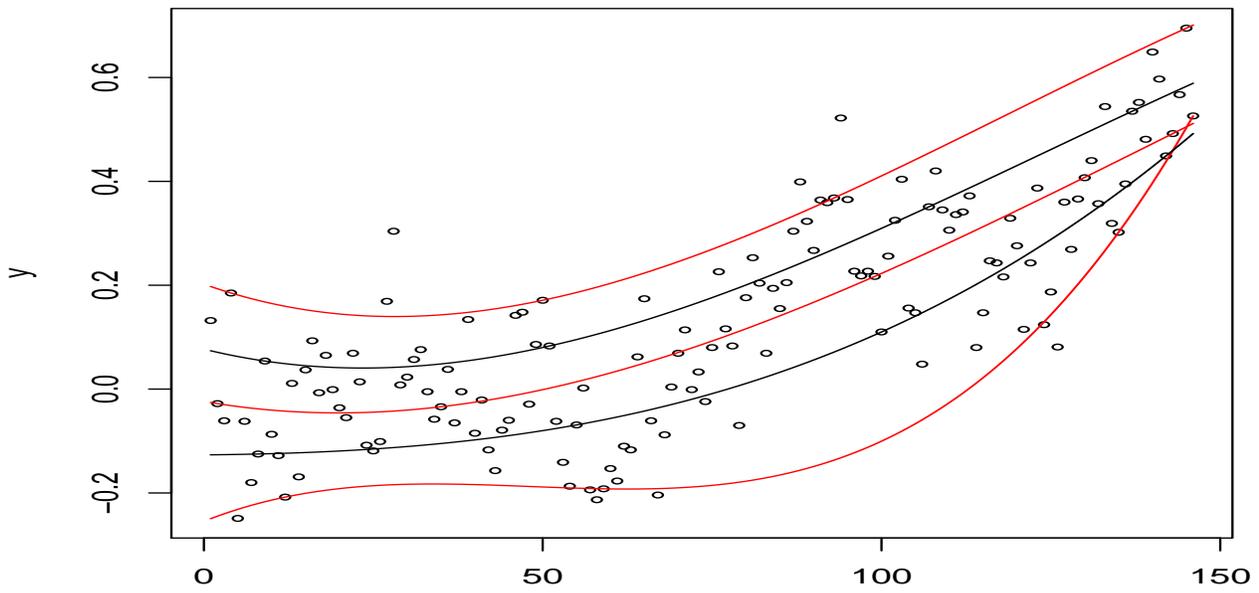
Global temperature is increasing/undulating.

The central spread (i.e., 0.75 - 0.25 quantile) is constant.

The full spread (i.e., 0.95 - 0.05 quantile) shrinks with time.

Hurricane Rita
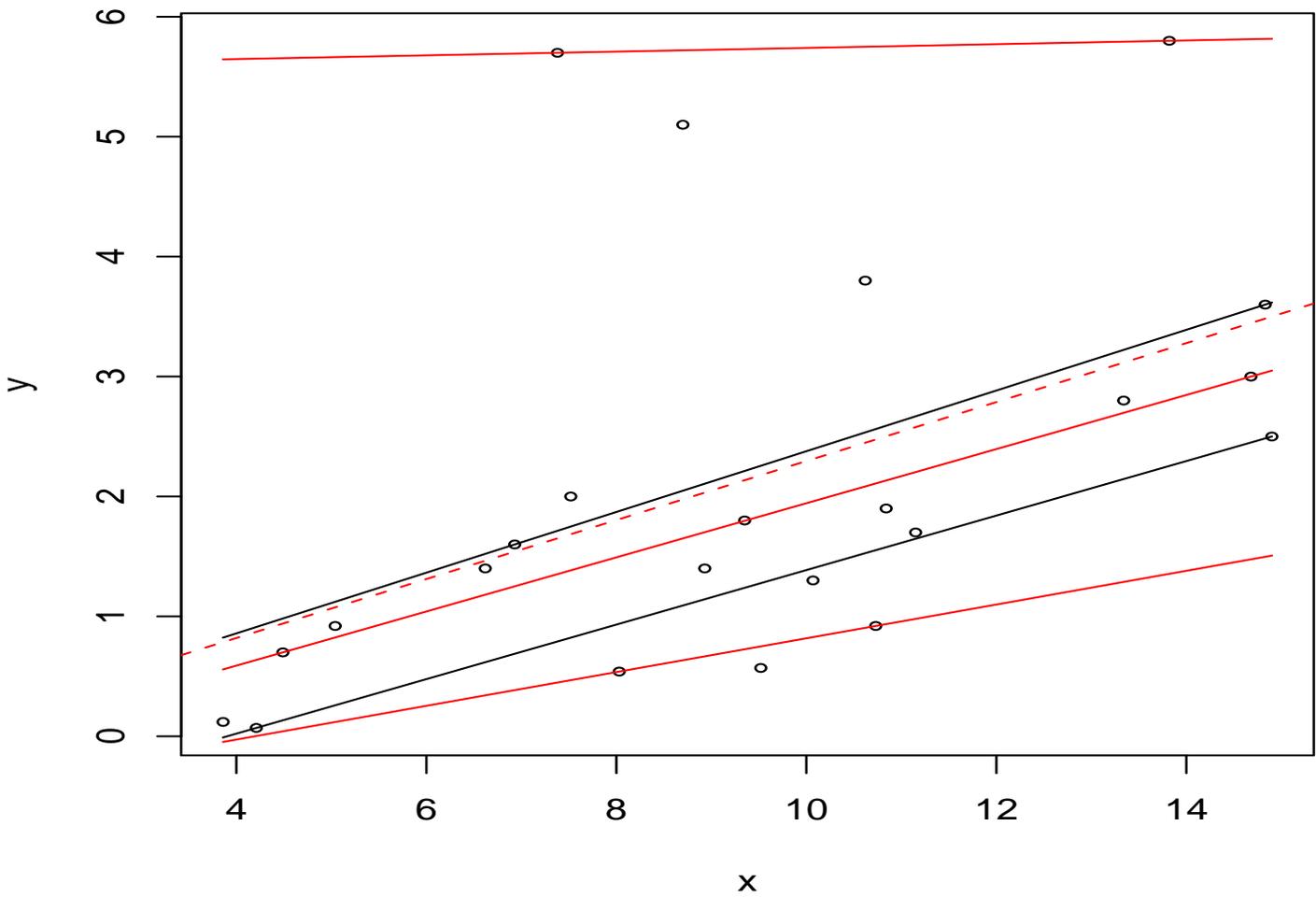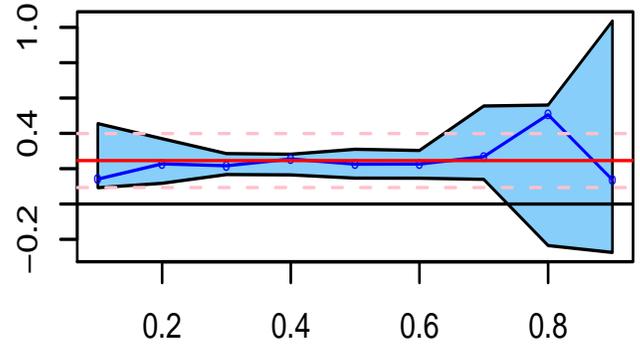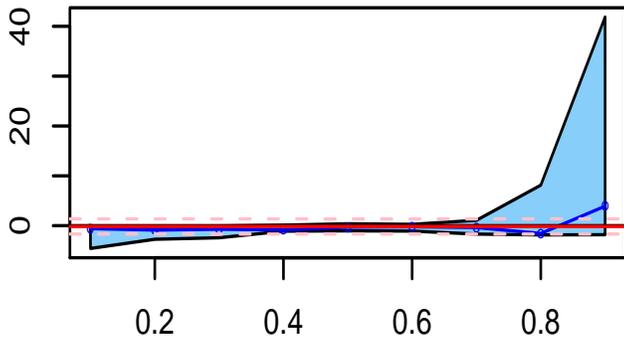
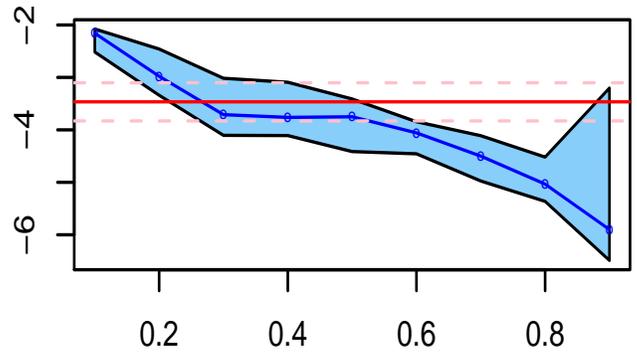There is a positive relationship between max water-level and max change in water-level.

Insufficient data for anything else!
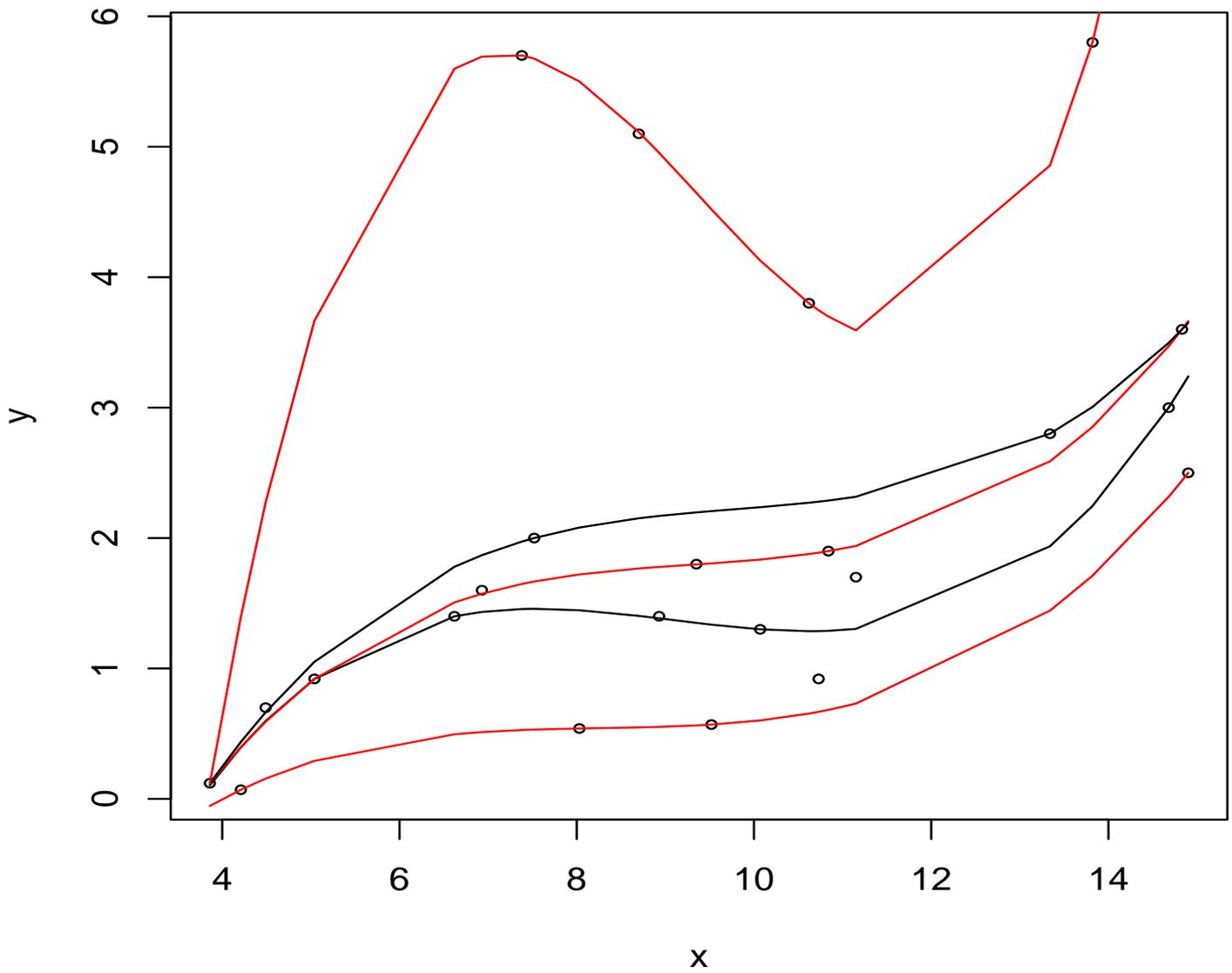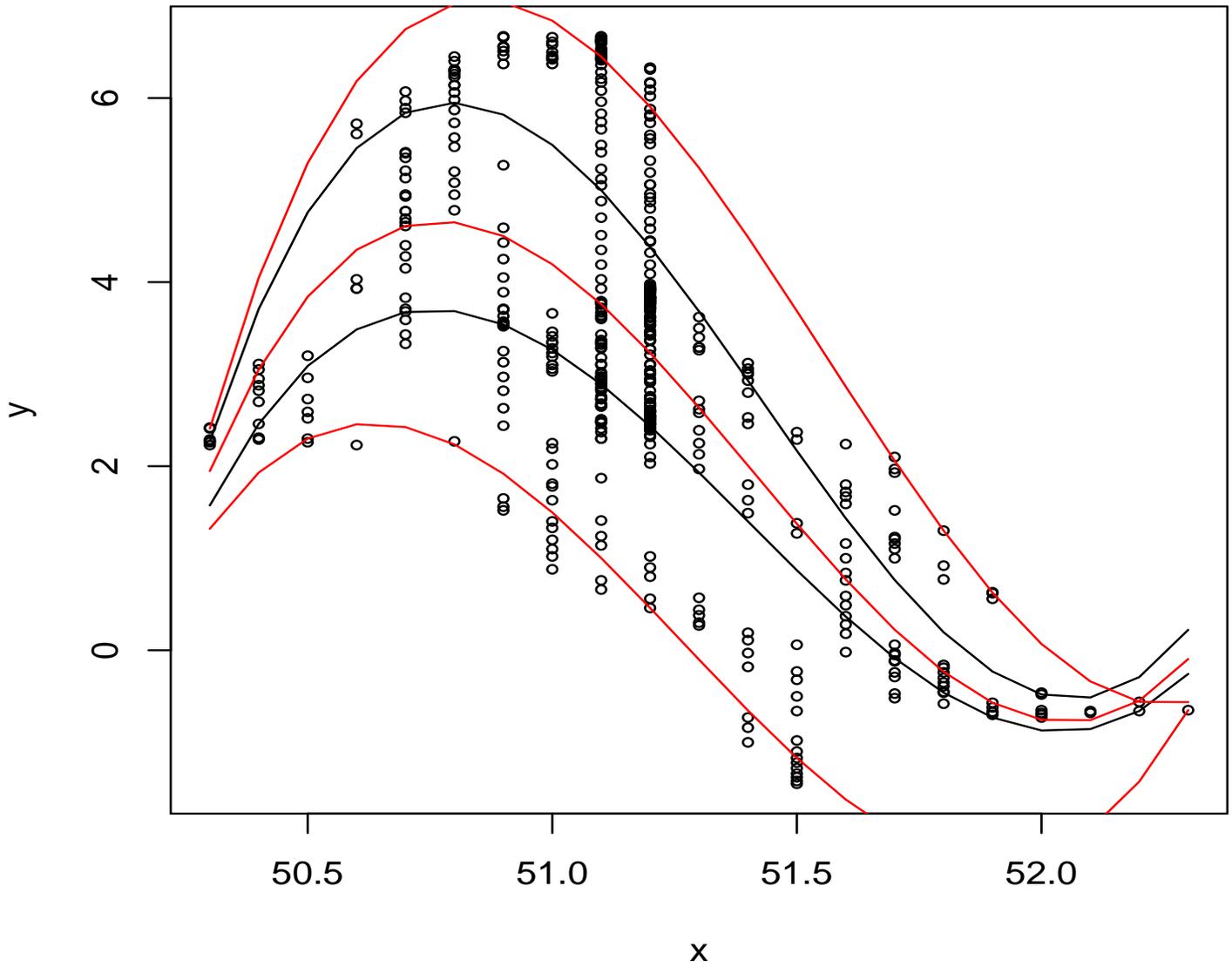
Crossing Quantiles?
Probably overfitting!

Water Level vs. Temperature at San Fransisco
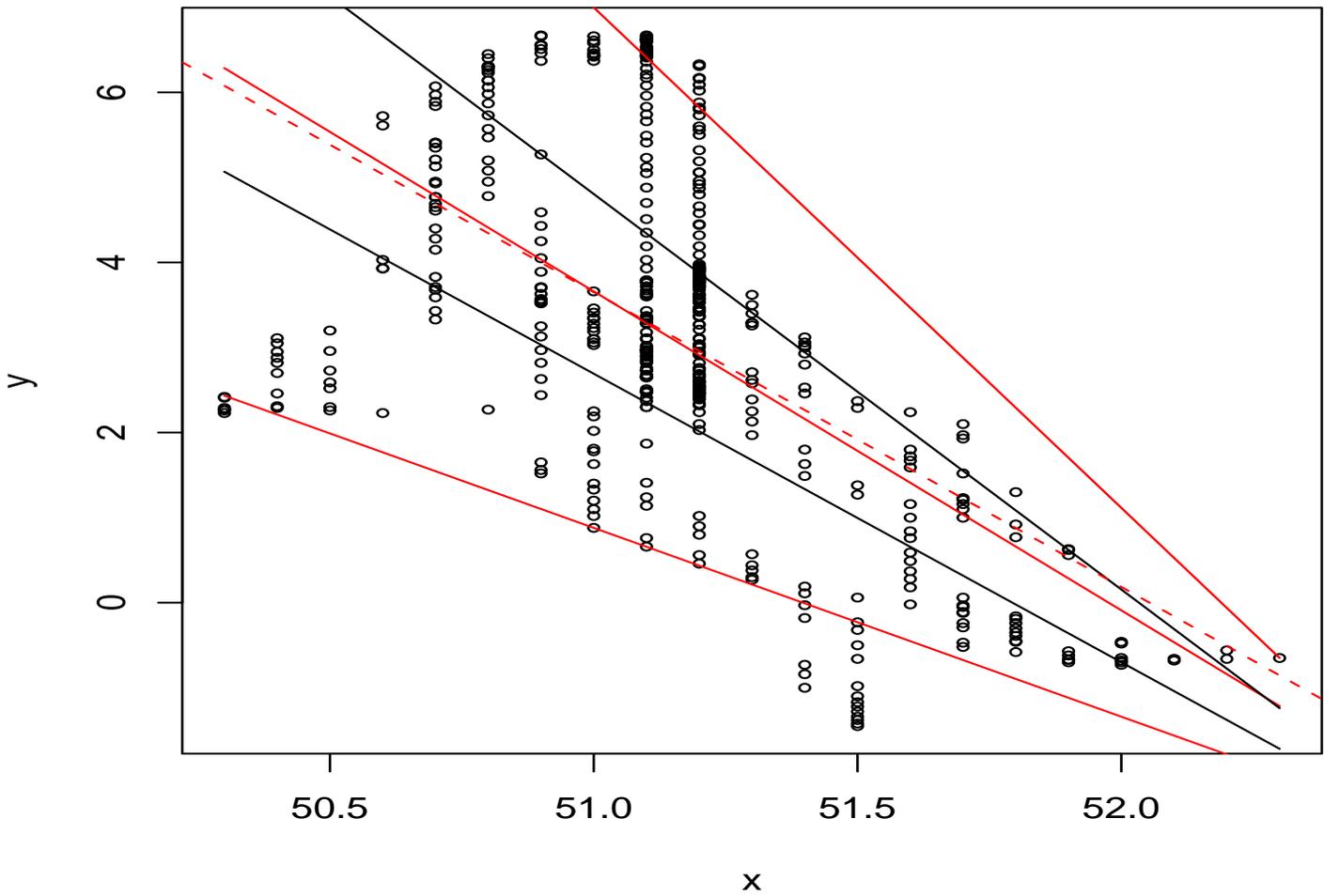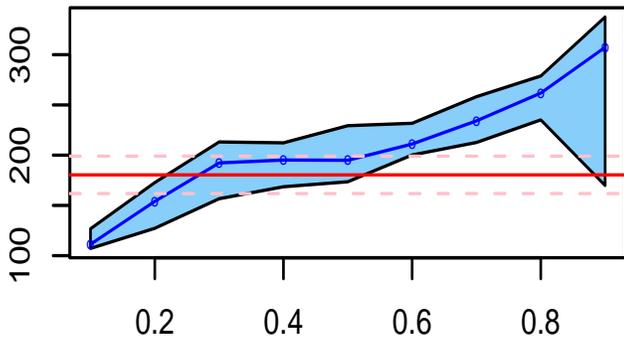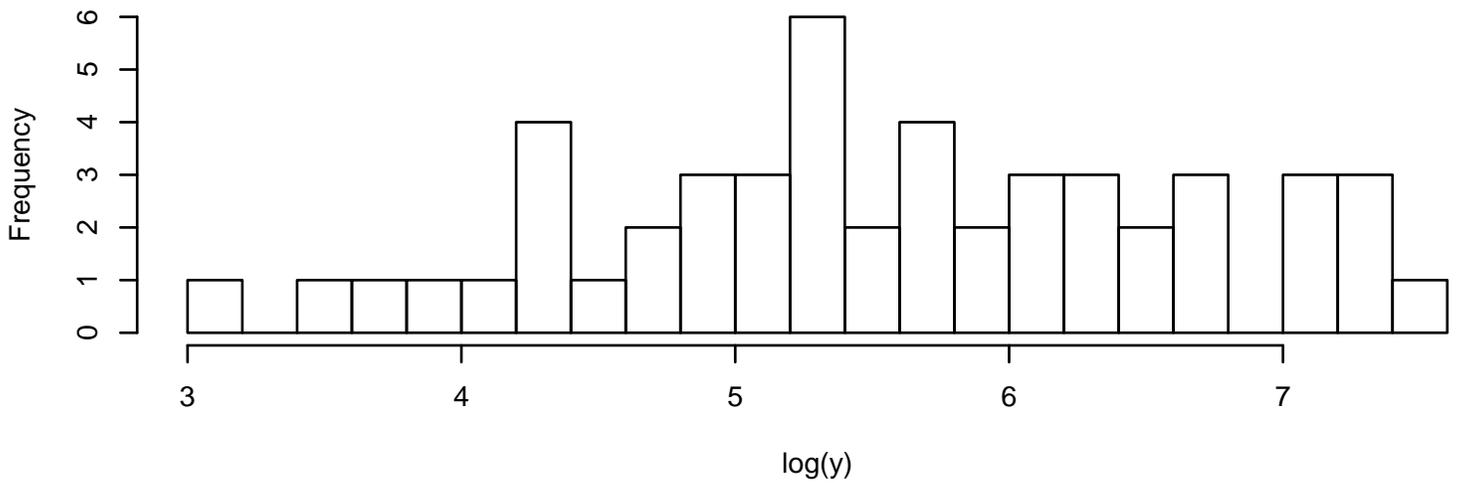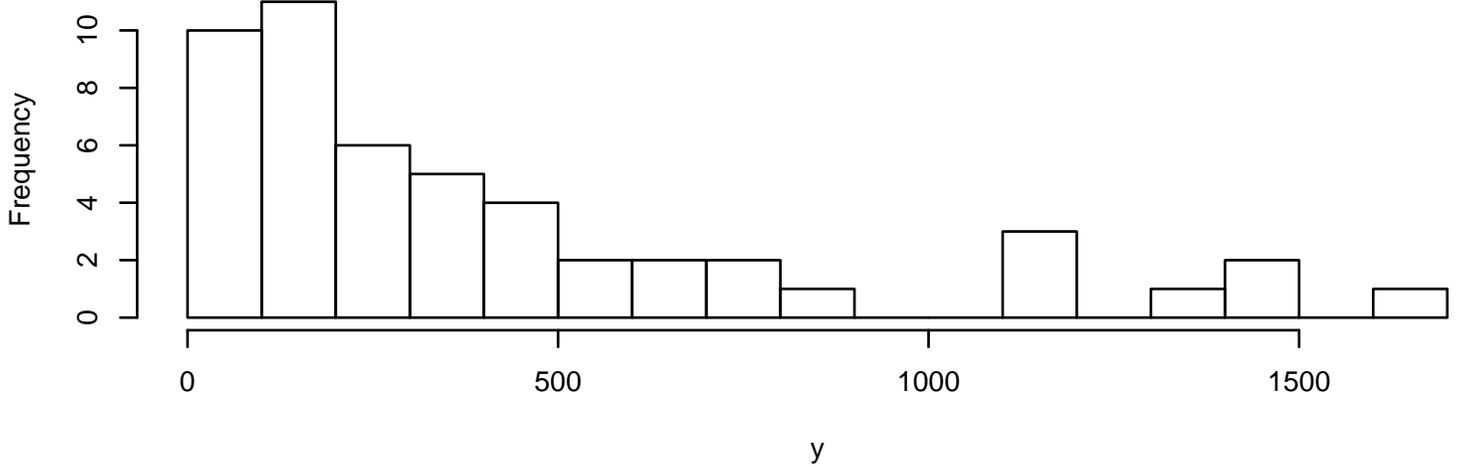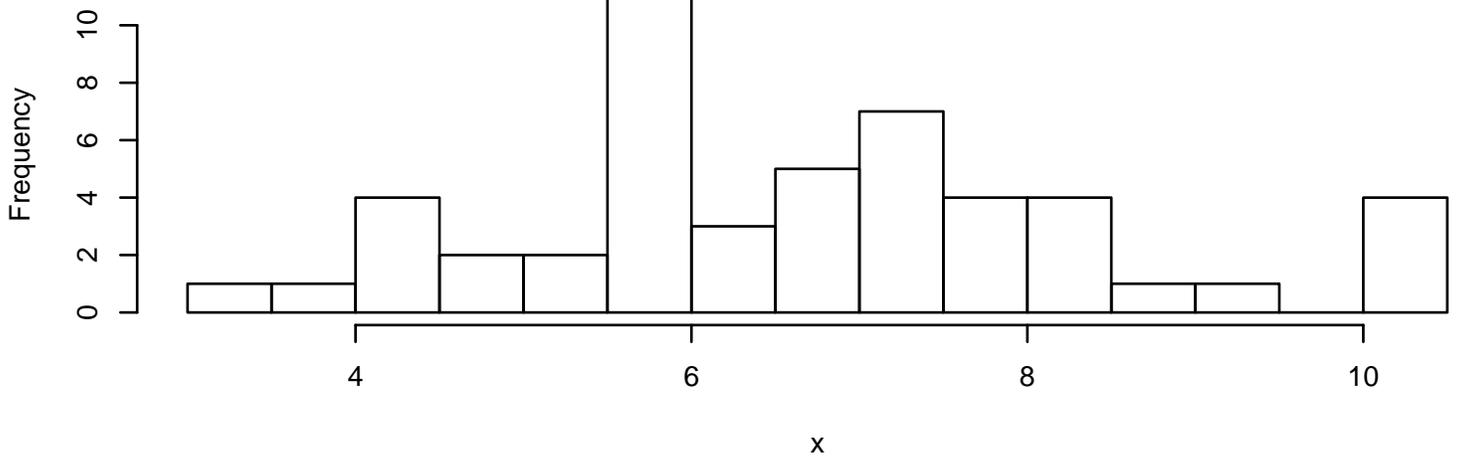
Surge inversely related to water temperature.

Reverse-fanning, i.e., less uncertainty in water-level at high temperatures.

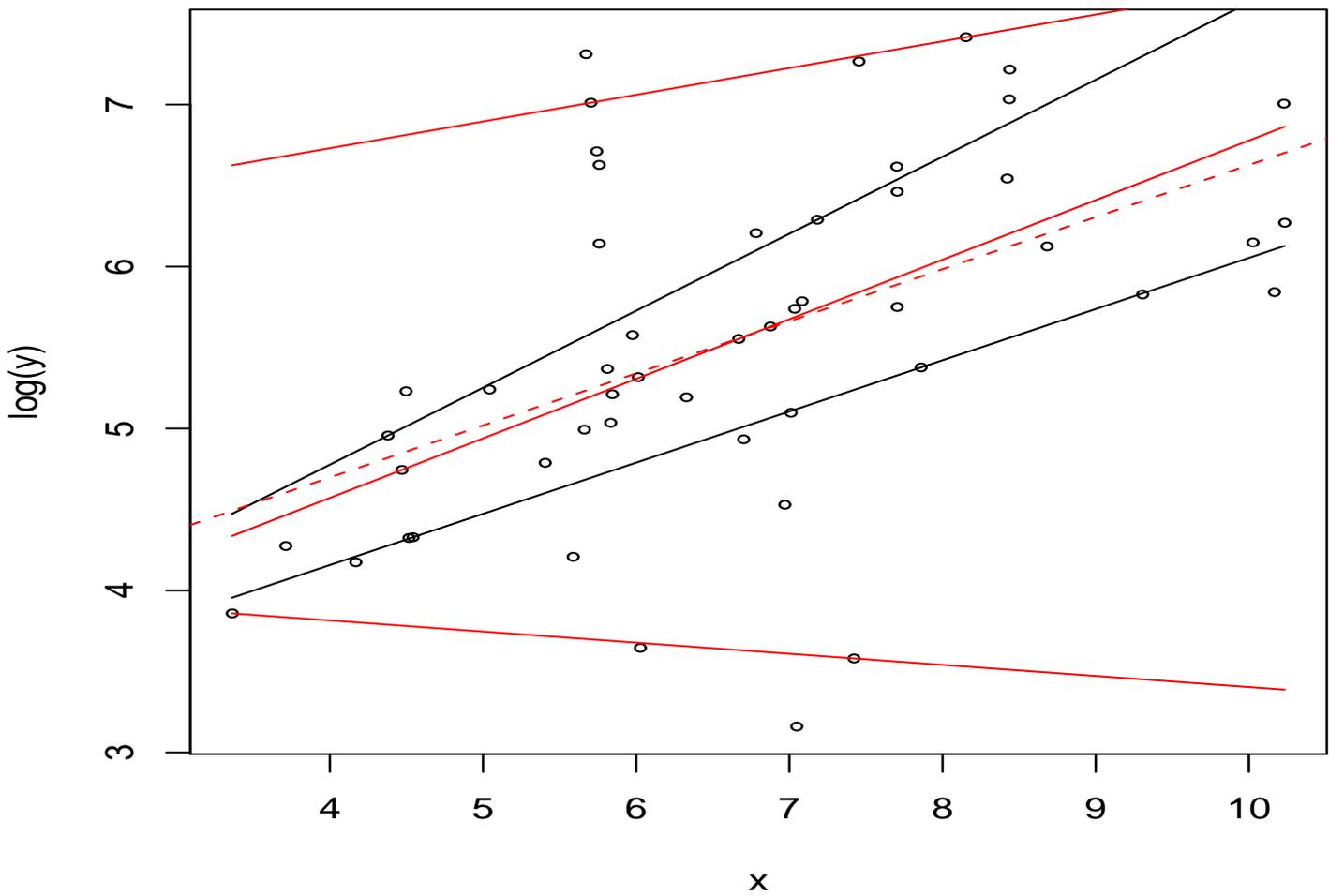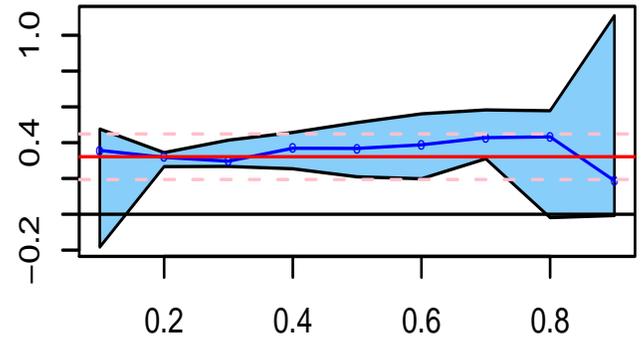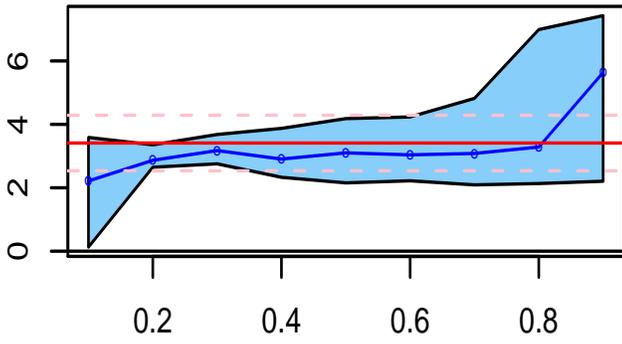Nonlinear, but why?

Shape of distribution of y changes with x.

Stream Flow vs. Precipitation
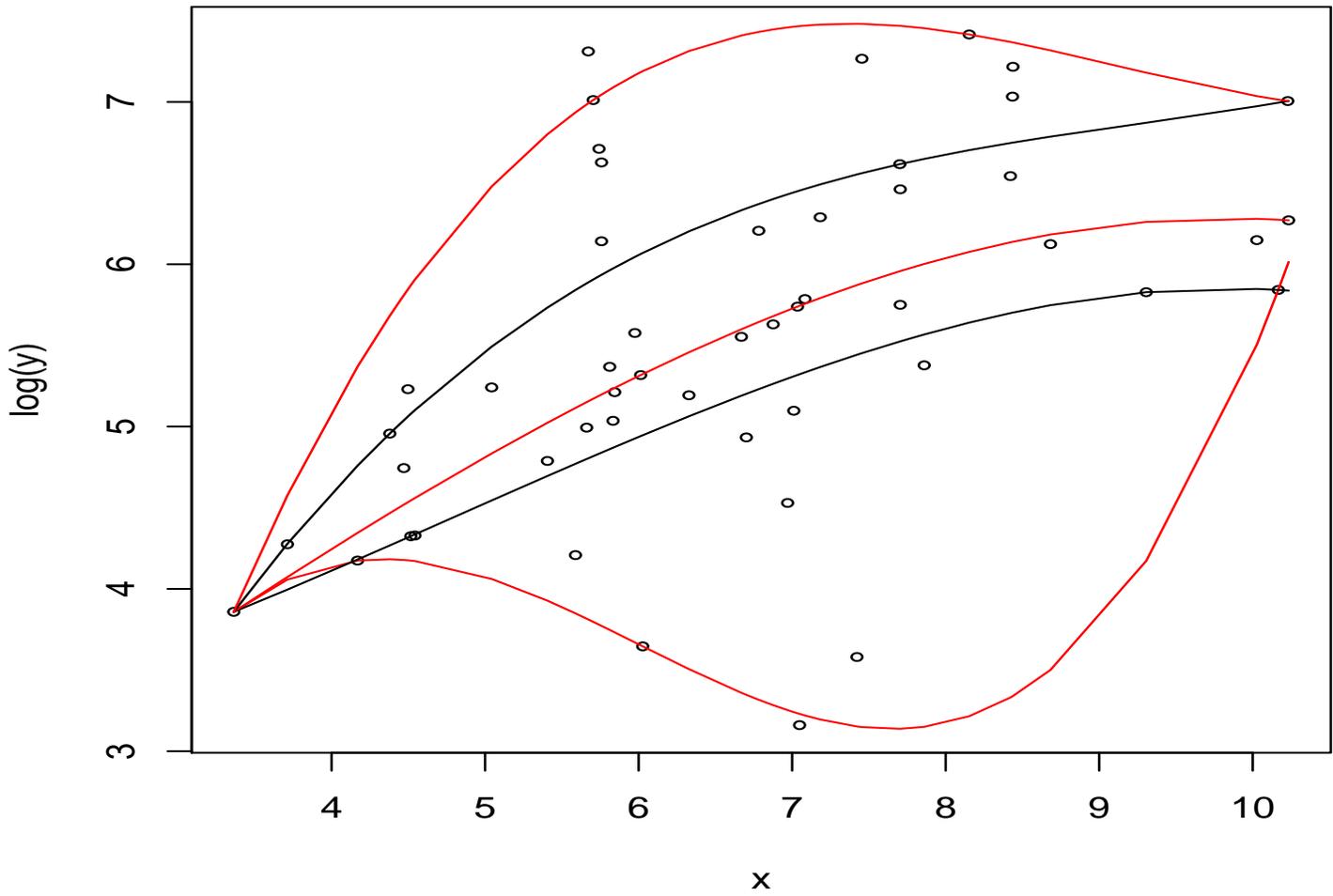
Transform stream flow to logarithm of stream flow.

Increasing prcp leads to increasing flow.
But not for the lowest quantile (0.05)! The upper quartile (0.95) doesn't increase as fast as the middle portion of the distribution.
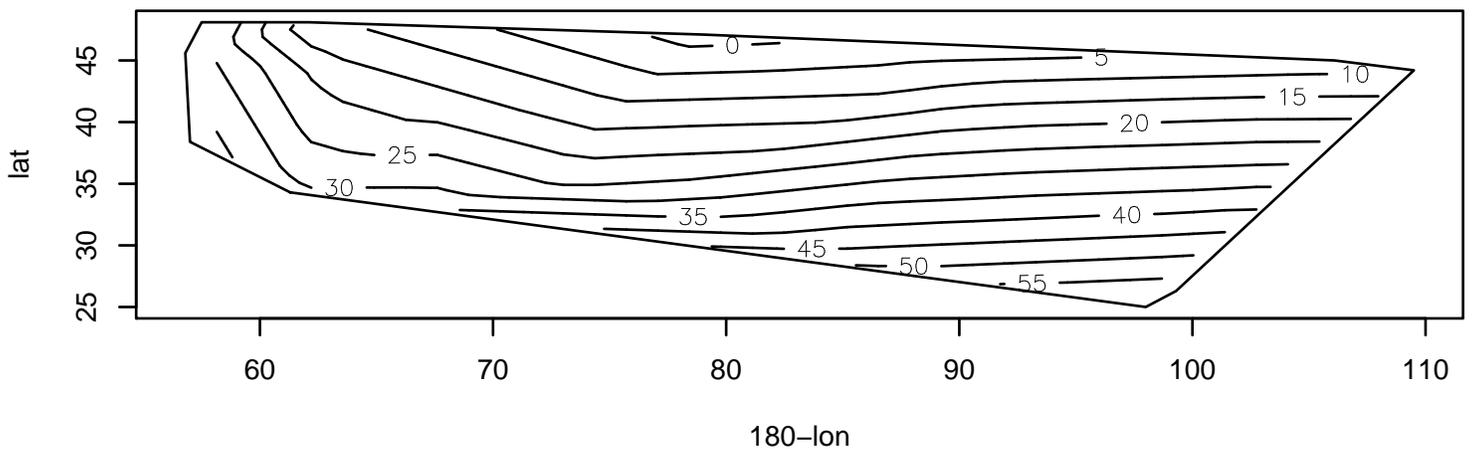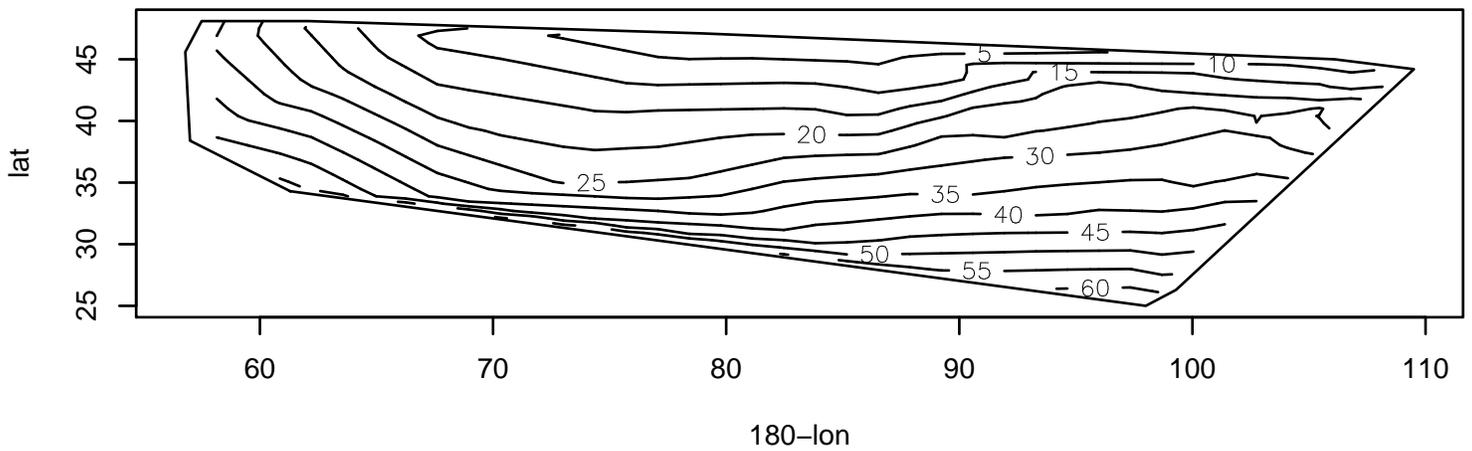
The middle portion of the distribution of flow shifts up with prcp.
Beware of extrapolation (0.05 quantile)!

Flow vs. Prcp with Quantile Random Forrest
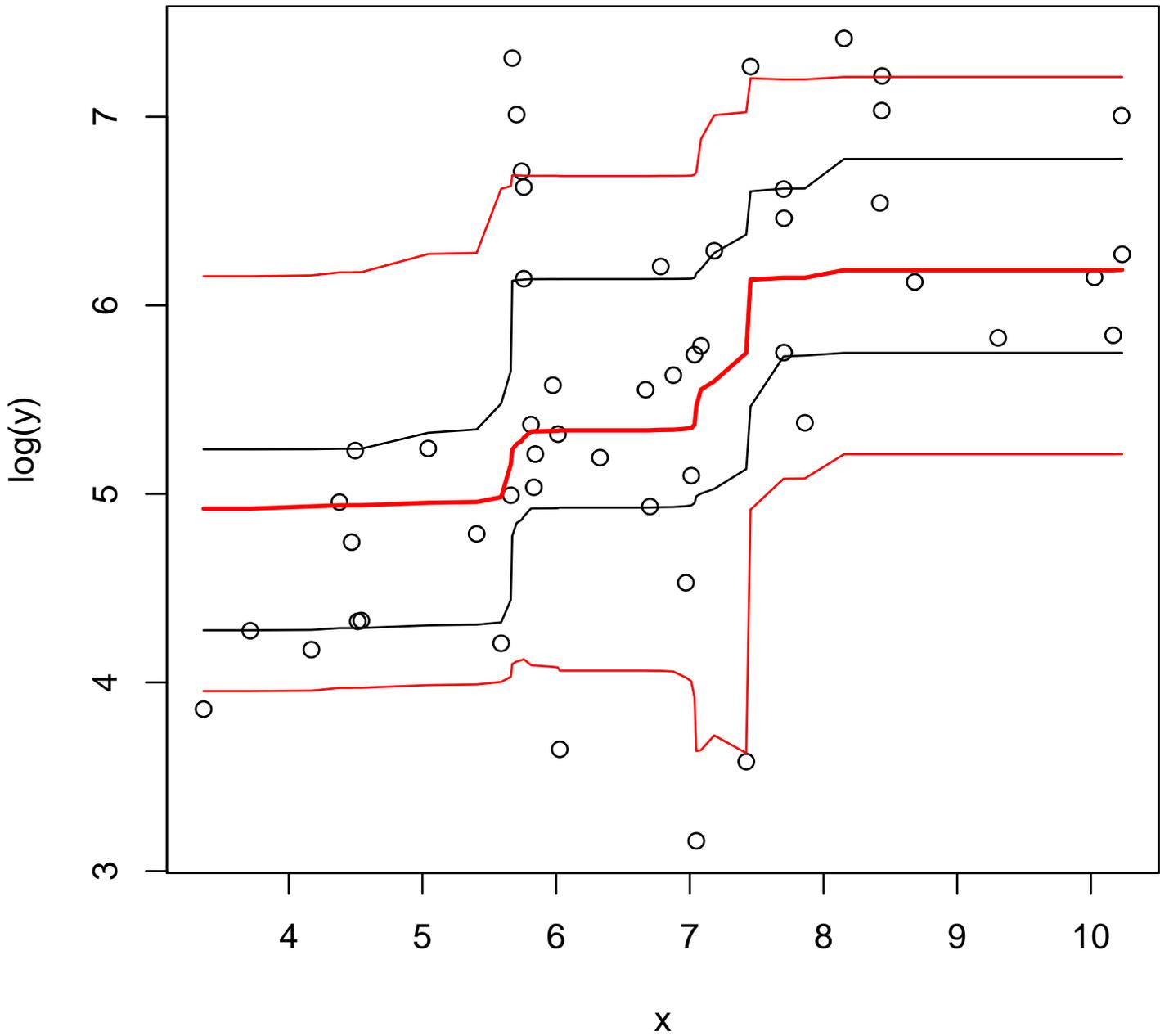
Abrupt shifts in flow when prcp is about 5.5 and 7.5.
Again, lower quantile (0.05) is mostly insensitive to prcp.

Jan. Min. Temperatures vs. Lat. and Lon.
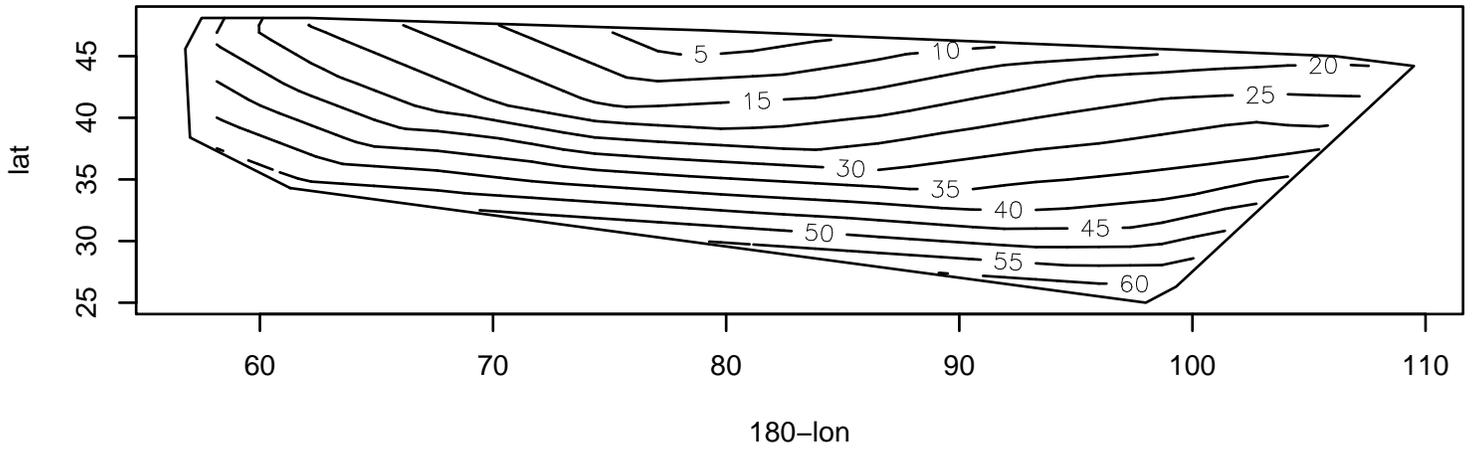
Quantiles: 0.99 (top), 0.5 (middle), 0.01 (bottom).

Top vs. median: SE and W mostly same; SW and NE warmer in top.

Bottom vs. median: SW and East of Mississippi cooler in bottom; No gradient across Mexican border. (Acknowledment: Scott Sandgathe)

Conclusions I

"Ordinary" regression, linear or not, provides only a single summary measure for the conditional distribution of response (predictand) - the mean - given the predictor.

Quantile regression provides a more complete picture of the conditional distribution.

The idea naturally lends itself to all the generalizations that have followed ordinary regression.

Many problems would actually benefit more from the prediction of extreme values (rather than the mean), anyway.

Even when the expected value is of interest, QR assays uncertainy via prediction intervals.

Juvenile Conclusions

The number of tornadoes is on the rise.
Max grows faster than min, the latter being mostly constant.

The whole shape of the conditional distribution of the number of tornadoes today varies as a function of the number of tornadoes yesterday.

Global temperature is on the rise.
The variation (between years) is dropping.

Max water level is positively associated with max change in water level.
Beware of overfitting.

Water level is negatively associated with water temperature.
The shape of the conditional distribution of the former varies with the latter.

Increased prcp leads to increased stream flow (for one station).
But not for the lowest quantiles of stream flow.

The strong temperature gradient across the Mexican border exists only in the median, and not in the tails of the distribution.