



Supporting Information for "Significance of trends toward earlier snowmelt runoff, Columbia and Missouri Basin headwaters, western United States"

Johnnie N. Moore, Joel T. Harper, Mark C. Greenwood

This is auxiliary material for this article contains one text file, four tables, and three figures from the article "Significance of trends toward earlier snowmelt runoff, Columbia and Missouri Basin headwaters, western United States" from *Geophysical Research Letters* on the 28 August 2007. <http://onlinelibrary.wiley.com/doi/10.1029/2007GL031022/full>

Moore, J. N., J. T. Harper, and M. C. Greenwood (2007), Significance of trends toward earlier snowmelt runoff, Columbia and Missouri Basin headwaters, western United States, *Geophys. Res. Lett.*, 34, L16402, doi:10.1029/2007GL031022.

Made available through Montana State University's [ScholarWorks](http://scholarworks.montana.edu)
scholarworks.montana.edu

Online Supplement to Moore et al. [2007] by Mark Greenwood

PART I: Comparing different measures of center of a hydrograph:

We chose to analyze the day of different percentiles of flow in contrast to Stewart et al. [2005] who use the center of mass flow that they call CT. $CT = \sum_i t_i q_i / \sum_i q_i$ where t_i is the day of the water year, q_i is the daily discharge. If we let $Q = \sum_i q_i$ be the total annual discharge, CT can be re-written to show that it is the average of the day of the water year weighted by the proportion of the annual flow on that day (p_i) as follows: $CT = \sum_i t_i q_i / Q = \sum_i t_i \frac{q_i}{Q} = \sum_i t_i p_i$. This provides a day of the average flow for the water year. In this section of the supplement, we compare the results for CT to our day of median flow. In the second section, the detection of linear trends is compared between the CT and our day of 50th DQF.

To illustrate the differences in the measures of timing, we will use the flow records from Gage #14. This gage is USGS Gage 13186000, South Fork of the Boise River, near Featherville, ID, Lat 43 29'45", Long 115 18'29" (NAD83), at an elevation of 4218 ft. Figure 1 contains the daily hydrographs and cumulative discharge functions for 1951 and 2005 for Gage 14. The CT for 1951 was day 218.1 and the day of 50th DQF was 229 and for 2005 the CT was 215 and the day of 50th DQF was 232. The sensitivity of the mean to the length of the tail of the distribution is seen in the difference between the measures of center in these two hydrographs.

The difference in dates between CT and the day of the median flow is not that important if we are trying to assess changes over time, but the CT will be more sensitive to "spikes" in flow at the beginning and end of the record since means are more sensitive to outliers than medians. With a mean, the timing of the spike is of critical importance but for the median, changing the location of any spike in the first 50% would not change the median at all. All that matters in that example is that the "spike" is in the first 50%. The 25th percentile that we also explored might be more sensitive to those early season spikes but also has similar robustness to locations of spikes within the first 25% of the flow.

Put another way, these hydrographs are generally skewed distributions and with skewed distributions, the median is generally preferred to the mean to accurately represent the center of the distribution. The mean is pulled in the direction of the longer tail in a skewed distribution. This also suggests that the CT is very sensitive to the choice of the beginning of the water year. If the beginning of the water year is redefined, then the mean will change dramatically. The median will change but not as dramatically.

PART II: Analysis of measures of center for Gage 14 over 55 years:

To illustrate the methods used and diagnostics to assess the linear regression assumptions, we show a complete analysis of the 50th DQF for Gage #14 and then compare those results to what would have been observed for CT. For this gage the average 50th DQF is 230.7 with a standard deviation of 9.6 days, the

average CT is 216.7 with a standard deviation of 11.8. Total annual discharge is transformed to have mean 0 and standard deviation 1 (standardized discharge for year t denoted Q_t) to be able to compare slope coefficients between gages that have very different centers and spreads. Figure 2 contains the two measures of center, plotted versus year and standardized annual discharge.

The regression models and diagnostics used here can be found in intermediate regression textbooks such as Kutner et al. [2005] and much of this is also available in books such as Davis [2002]. All calculations are carried out using R (www.r-project.org), an open source statistical analysis program, but could be performed in most statistical software packages.

For the 50th DQF for this gage, we will first consider the simple linear regression model using Year as the explanatory variable ($DQF_t = \beta_0 + \beta_1 Year + \epsilon_t$). We chose to perform 2-sided t-tests for $H_0: \beta_1 = 0$ versus $H_A: \beta_1 \neq 0$ at the 10% significance level with an adjustment for multiple testing. Note that this test is exactly the same as testing that the true correlation coefficient, ρ , between 50th DQF and Year is different from 0. The Pearson product-moment correlation coefficient between Year and 50th DQF is -0.251.

The linear regression model results are found in Table 1. Additionally, the R^2 is 0.063 for this model, which says that 6.3% of the variation in the Day of Median Flow is explained by Year, which is much smaller than the R^2 that is observed when total discharge is used as the explanatory variable below.

The unadjusted 90% confidence interval for the time trend slope coefficient is from -0.282 to -0.017 days per year (or -15.6 to -0.95 over 55 years). Upon adjustment, the confidence interval goes from -0.3092 to 0.009 days per year or -17 to 0.5 days of change per 55 years. The p-value of 0.064 is compared to a significance level of $\alpha=0.10$ initially but upon adjustment for multiple testing is compared to $\alpha^*=0.05$, so it is significant without adjustment and not significant upon adjustment. This effect on detection of a significant result was observed 12 times in the 126 tests that we performed. If many more tests were considered in the study, many more of these types of changes would be observed.

To use the standard simple linear regression model, we assume that the errors are independent, normally distributed, have constant variance and show no indication of nonlinearity. Independence is assessed by evaluating potential autocorrelation in the residuals using an autocorrelation plot in Figure 3(i) and a hypothesis test for autocorrelation. We can assess normality using a Q-Q plot as in Figure 3(ii), where we are comparing the distribution of the residuals to a normal distribution, with normality of residuals corresponding to a one-to-one relationship between the distribution of the residuals and the normal distribution. Constant variance is assessed visually in plots of the residuals in Figure 3(iii) and also using a hypothesis test for heteroskedasticity. Nonlinearity can also be assessed visually in this same plot.

We consider the assumption of independent errors through calculating the correlation in the residuals at different lags or years apart which provides the autocorrelation function (ACF). The ACF would indicate a problem if any of the correlations estimated at different lags were outside the dashed bands, suggesting a significant correlation within the residuals at the lag distance apart.

Note that the first autocorrelation displayed in Figure 3(i) is large and that it is for the lag 0 autocorrelation, which corresponds to the correlation of the residuals with themselves and is always 1. None of the other autocorrelations are larger in magnitude than 0.264.

A test for significant lag 1 autocorrelation in the residuals can also be performed using the Durbin-Watson test, which leads to a test statistic of 2.24 and a p-value of 0.468. Here, failing to reject the null hypothesis suggests that there is no evidence of a significant autocorrelation in the residuals and that the independence assumption is likely reasonable.

If these results had suggested a significant autocorrelation, then this could have been incorporated into the parametric model without resorting to nonparametric techniques. Because autocorrelation was not significantly present, these modifications were not required even though they were initially considered.

Next we consider the normality of residuals which includes a partial assessment of outliers since the presence of outliers would violate the normality assumption. While the points do have a modest systematic deviation in the upper right tail, this is only a moderate indication of a violation of the normality assumption. The sample size of 55 is large enough to provide some robustness to violation of this assumption.

A plot of the residuals versus year as in Figure 3(iii) is useful for assessing the constant variance assumption and identifying potential outliers. There is one potential outlier in 1978 but otherwise there is no indication of a violation of the constant variance assumption. We can also test for constant variance using the Breusch-Pagan test for heteroskedasticity. Here, this provides a p-value of 0.2137 which suggests that there is not significant evidence of non-constant variance.

For an outlier to have a large influence on the regression line, it must either be really extreme in magnitude and/or have high leverage. The leverage of an observation is a quantity that can be calculated and corresponds to the distance between the observation and the mean of the explanatory variable. Observations with high leverage have the most potential for influencing the regression line. Plotting the residuals versus their leverage values provides information about whether any potential outliers have high leverage and thus might be highly influential regarding the regression line. In this example, the potential outlier has one of the smallest leverage values so will not have a large impact on the regression line. If the outliers were at the beginning or end of the time series, then this would not be the case. This is typical of the results observed, where unusual years were observed most often in the late 1980s but not at the beginning or end of the series.

We can also test for outliers using the Bonferroni outlier test which results in a p-value of 0.10326. This suggests that there may be an outlier present in the data set but it is by no means conclusive.

For a comparison, the regression results using the standardized discharge as the explanatory variable as displayed in Figure 2(iv) are also considered below. It was often observed that the relationship between total discharge and the different DQFs was nonlinear, which actually implies that potentially even stronger relationships are failing to be detected by forcing linear relationships

between these variables. In other words, the p-values are upper bounds for the p-values that could be achieved if nonlinear models would be considered. The linear methods are a reasonable starting point for inference regarding the linear component of the potentially nonlinear relationship between these variables and provide direct comparison to the significance of the linear changes over time. Further research will be involved in more fully exploring the nonlinear aspects of the relationship between DQFs and total discharge.

The regression results for 50th DQF versus standardized total annual discharge ($DQF_t = \beta_0 + \beta_1 Q_t + \epsilon_t$) are found in Table 2. The R^2 for this model is 0.3653 which suggests that 37% of the variation in the 50th DQF is explained by the linear model using total annual discharge. While this value is not extremely large, it is much better than for the relationship between it with Year.

There are indications of non-constant variance and nonlinearity in the diagnostics but no indication of autocorrelation, outliers or non-normal residuals. This suggests with an appropriate transformation or nonlinear model that these results can be improved further. But this also shows how this gage's 50th DQF is responding to total discharge much more distinctly than it is changing over time. If the gage were clearly responding to the monotonic increase in temperature through earlier snowmelt, the time trend tested for here would be more clearly identified even in the presence of changes due to total discharge.

The final comparison is between the 50th DQF and CT using the same techniques. These two measures of center are highly correlated, with a Pearson correlation of 0.93, so they are measuring similar information. However, there is less evidence of a significant time trend in CT (p-value=0.09, Table 3) and more evidence of a relationship between CT and total annual discharge (p-value=1e-10, Table 4) than for the 50th DQF for this gage. This implies that in working with the CT, it is even more important to consider the relationship with total discharge and that the evidence for a linear change in CT over time may not be as strong as the limited evidence that we detected in our percentile based measures. Note that the results from these two models were not included in the tests that were subject to adjustment for multiple testing and were just presented to illustrate the differences between our measure of the center of a hydrograph and that used in Stewart et al. [2005].

Additional References:

- Kutner, M., Nachstein, C., Neter, J., and Li, W. (2005) *Applied Linear Statistical Models, 5th edition*. McGraw Hill: New York.
- Davis, J. (2002) *Statistics and Data Analysis in Geology, 3rd edition*. Wiley: New York.

Table S1: Regression Results: 50th DQF and Year

Coefficient	Estimate	Stand Err	t	p-value
Intercept	527.72	157.06	3.36	0.002
Slope	-0.1501	0.0794	-1.89	0.064

Table S2. Regression Results: 50th DQF and Standardized Annual Discharge.

Coefficient	Estimate	Stand Err	t	p-value
Intercept	230.74	1.037	222.5	< 2e-16
Slope	5.78	1.047	5.5	1.03e-06

Table S3. Regression Results: CT and Year.

Coefficient	Estimate	Stand Err	t	p-value
Intercept	552.3	195.51	2.83	0.0067
Slope	-0.17	0.099	-1.72	0.092

Table S4. Regression Results: CT and Standardized Annual Discharge.

Coefficient	Estimate	Stand Err	t	p-value
Intercept	216.74	1.083	200.1	< 2e-16
Slope	8.773	1.093	8.027	1.01e-10

Figure S1. Hydrographs of 1955 and 2005 and cumulative discharge functions for gage 14 by day of water year, displayed as a proportion of the total yearly flow for each year.

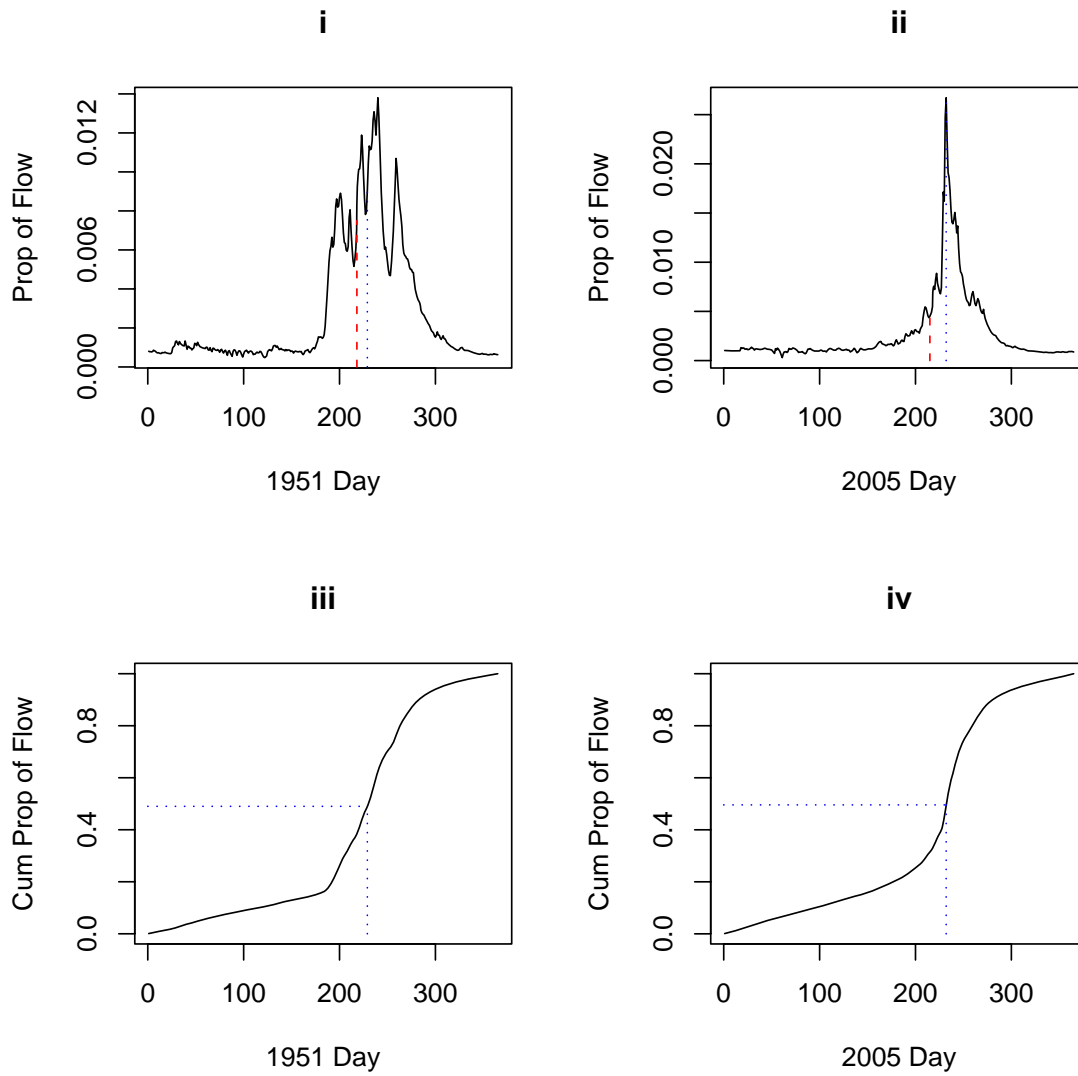


Figure S2. Plot of CT versus time and standardized total discharge and day of 50th DQF versus time and standardized total discharge with linear regression models.

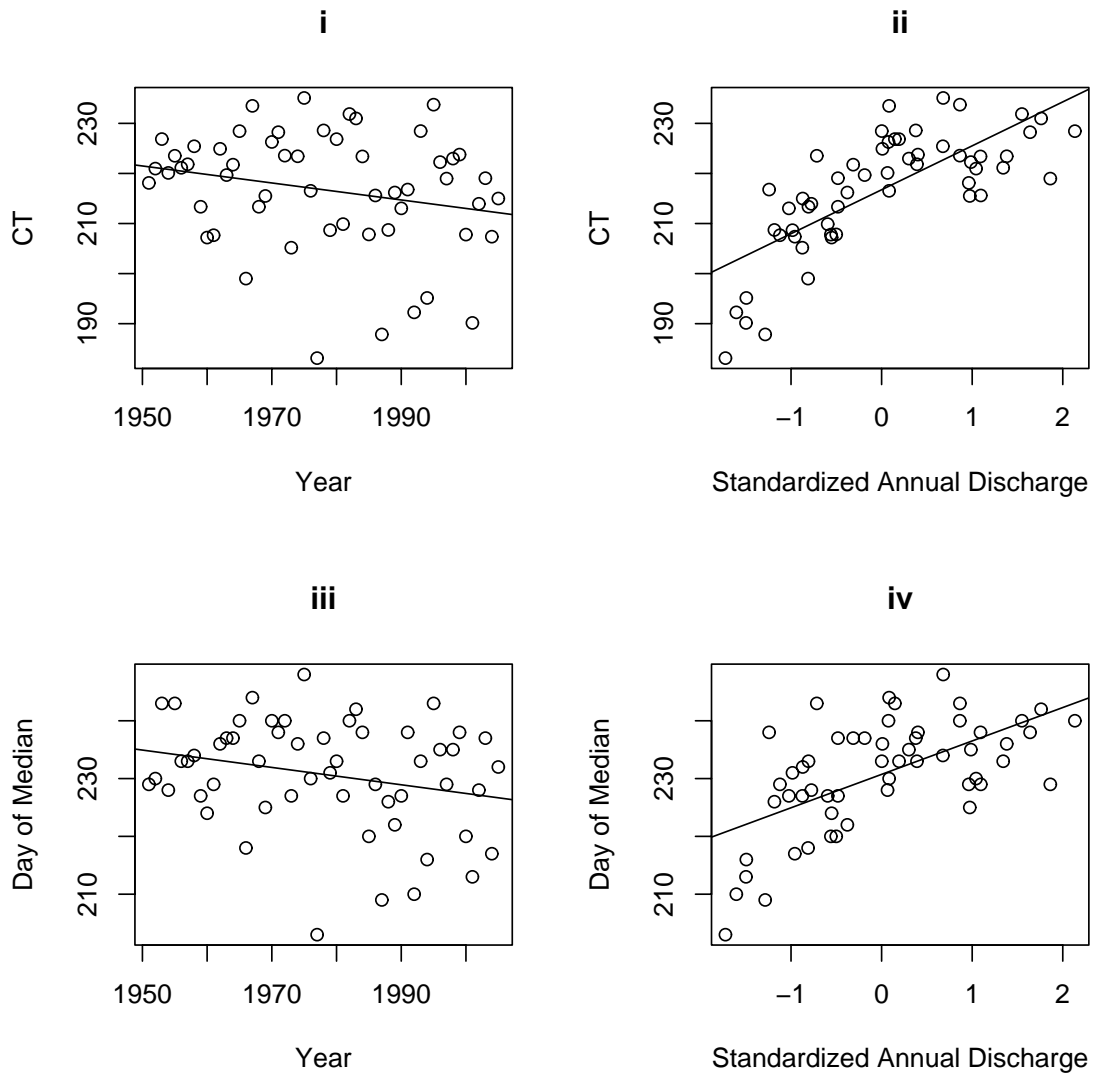


Figure S3. Diagnostic plots for 50th DQF versus time model.

