Changes in Streamflow Statistical Structure across United States due to Recent Climate Change

Abhinav Gupta[1,1], Rosemary W H Carroll[1,1], and Sean M Mckenna[1,1]

[1]Desert Research Institute

November 30, 2022

Abstract

A variety of watershed responses to climate change are expected due to non-linear interactions between various hydrologic processes acting at different timescales that are modulated by watershed properties. Changes in statistical structure (spectral properties) of streamflow in the USA due to climate change were studied for water years 1980-2013. The Fractionally differenced Autoregressive Integrated Moving Average (FARIMA) model was fit to the deseasonalized streamflow time-series to model its statistical structure. FARIMA allows the separation of streamflow into low frequency (slowly varying) and high frequency (fast varying) components. Results show that in snow dominated watersheds, the contribution of low frequency components to total streamflow variance has decreased over the study period, and the contribution of high frequency components has increased. The change in snow dominated watersheds was primarily driven by changes in rainfall statistics and changes in snow water equivalent but also by changes in seasonal temperature statistics. Among rain-driven watersheds, the contribution of high frequency components generally increased in arid regions but decreased in humid regions. In both humid and arid rain-driven watersheds, increasing winter temperature was responsible for the change in streamflow regimes. These results have consequences for predictability of streamflow in the presence of climate change. We expect that changes in the high frequency component will result in poorer predictability of streamflow.

Hosted file

`supporting_information.docx` available at https://authorea.com/users/559044/articles/607890-changes-in-streamflow-statistical-structure-across-united-states-due-to-recent-climate-change

Changes in Streamflow Statistical Structure across United States due to Recent Climate Change

Abhinav Gupta*[1], Rosemary W H Carroll[2], Sean M. McKenna[2]

[1]Division of Hydrologic Sciences, Desert Research Institute, 755 E. Flamingo Rd., Las Vegas, NV, 89169, United States of America
[2]Division of Hydrologic Sciences, Desert Research Institute, 2215 Raggio Pkwy, Reno, NV 89512, United States of America

*Corresponding Author: Abhinav Gupta (abhinav.gupta@dri.edu)

Highlights

(1) Change in climatic statistics has resulted in a change in streamflow statistical structure
(2) Landscape characteristics play an important but secondary role in changing streamflow statistical structure
(3) Increase in winter temperature increases (decreases) the high frequency component of streamflow in arid (humid) regions

Abstract:

A variety of watershed responses to climate change are expected due to non-linear interactions between various hydrologic processes acting at different timescales that are modulated by watershed properties. Changes in statistical structure (spectral properties) of streamflow in the USA due to climate change were studied for water years 1980-2013. The Fractionally differenced Autoregressive Integrated Moving Average (FARIMA) model was fit to the deseasonalized streamflow time-series to model its statistical structure. FARIMA allows the separation of streamflow into low frequency (slowly varying) and high frequency (fast varying) components. Results show that in snow dominated watersheds, the contribution of low frequency components to total streamflow variance has decreased over the study period, and the contribution of high frequency components has increased. The change in snow dominated watersheds was primarily driven by changes in rainfall statistics and changes in snow water equivalent but also by changes in seasonal temperature statistics. Among rain-driven watersheds, the contribution of high frequency components generally increased in arid regions but decreased in humid regions. In both humid and arid rain-driven watersheds, increasing winter temperature was responsible for the change in streamflow regimes. These results have consequences for predictability of streamflow in the presence of climate change. We expect that changes in the high frequency component will result in poorer predictability of streamflow.

Keywords: Streamflow, Climate change, FARIMA, Spectral analysis, snow-dominated watersheds, Rain-driven watersheds

1. Introduction

The global hydrologic water balance will be impacted directly by climate change (Milly et al., 2005; Milly & Dunne, 2016; Mote et al., 2018; Manabe & Broccoli, 2020) which will alter streamflows. The extent and nature of hydrologic change depends upon several factors including watershed geomorphological characteristics (Lee & Delluer, 1972; Rodriguez-Iturbe & Rinaldo, 1997, Chap. 7), vegetation characteristics and soil properties (Eagleson, 1978), the dominant mode of streamflow production (snowmelt or rain, quick flow, baseflow etc.), changes in vegetation characteristics (e.g., Milly, 1997), and the pre-existing climate against which changes occur. Thus, a rich variety of watershed responses can be expected due to the change in climate as summarized through climate statistics (Gordon et al., 2022). The hydrologic responses of watersheds to climate change need to be understood to devise an effective adaption strategy.

Because of strong feedbacks between various components of a hydrologic systems, climate change can potentially lead to profound changes in watershed hydrologic regime. Hydrologic regime here refers to the interaction between different components of hydrologic process which produce hydrologic fluxes such as streamflow and evapotranspiration (ET). An example is the feedback between climate, soil, and vegetation properties (Rodriguez-Iturbe et al., 1999, 2001). Soil stores some of the precipitation as soil moisture which is taken up by the vegetation (Porporato et al., 2001). Climate has a strong control over soil moisture dynamics via precipitation frequency and depth (Laio et al., 2001). Also, the intensity of the climatic control on soil moisture dynamics is directly affected by soil properties such as soil texture, soil depth, and water holding capacity. Vegetation provides feedback to the atmospheric properties via transpiration and, at long timescales, soil properties via plant residue decomposition in soils (Eagleson, 1982). Thus, vegetation properties influence climate through the soil zone. These feedbacks operate at different timescales. The feedback between climate and soil moisture dynamics is fastest, followed by the feedback between climate and vegetation (via soil moisture dynamics). The feedback between vegetation and soil properties is slowest. Therefore, effects of climate change are expected to be observable at different timescales.

Streamflow is the integrated response of a watershed's hydrology, which is affected by inherent properties such as soil depth and texture, bedrock permeability, and topography that influence hydrology. Thus, studying changes in streamflow characteristics provides the clues to understanding the changes in watershed hydrologic regime. Hydrologists have employed various mathematical models (simulation approaches) to understand the streamflow response of a watershed at different timescales. These models can be broadly classified as deterministic models (Beven, 2011), stochastic models (Klemes, 1978), and statistical models (Montanari et al., 1997). The model that is used depends upon the spatial scale (watershed scale, regional scale, global scale, etc.) and timescale (daily, monthly, yearly, etc.) at which simulations/predictions are required along with the purpose of simulations/predictions (policy making, scientific hypothesis testing).

For most of the models used, some parameters of the model need to be calibrated against observations. The values that these parameters take depends upon climate statistics (mean annual precipitation depth, precipitation frequency, seasonal mean temperatures etc.) and watershed properties. Temporal non-stationarity introduced by climate change (Milly et al., 2008) makes the calibrated parameters dependent upon observation time-period. In fact, climate change may directly affect the physical characteristics of a watershed via change in vegetation characteristics (Milly, 1997). This introduces additional uncertainty in model projections/predictions in the presence of climate change. For example, Stephens et al. (2020) showed that changes in rainfall statistics along with changes in atmospheric CO_2 can change the soil moisture statistics. It may take a few years for a calibrated hydrologic model to adjust to the new equilibrium conditions. Other examples of climate change impacting watershed hydrologic characteristics include changes in snowpack in the western USA (e.g., Belmecheri et al., 2016), and change in baseflow and stormflow (e.g., Ficklin et al., 2016). In summary, the problem is that climate non-stationarities may make a hydrologic model calibrated and validated against historical observations unreliable for prediction/simulation in changed conditions.

Some strategies have been proposed to address this problem. Klemes (1986) proposed differential split-sample testing to test the robustness of a model under change, but such strategies may not be useful in case of large changes, especially if the change in a watershed is toward a drier hydrologic regime (Stephens et al., 2020). Singh et al., (2011) proposed a space-time symmetry approach under an uncertainty framework to estimate streamflows in a watershed in the presence of regime change. The idea behind space-time symmetry is to use available hydrologic information across different watersheds to predict future streamflow in another watershed. The assumption is that the spatial variability in hydro-climatological characteristics across watersheds is a good representation of the temporal variability that can be expected due to climate change. The idea of space-time symmetry has been demonstrated to be useful at yearly timescale using the Budyko framework (e.g., Sivapalan et al., 2011). Success of machine learning (ML) methods in estimating streamflows at gauged and ungauged locations at a daily timescale (Kratzert et al., 2018) suggests that there is a considerable amount of hydrologic information shared between different watersheds. However, there is limited evidence of successful application of space-time symmetry at a daily timescale (see, Singh et al., 2011), especially under a changing climate. Therefore, there is a need to further test this idea at daily timescale. Such a testing procedure would require identifying watersheds that have undergone hydrologic regime change. This is the main motivation for this work.

In this study, change in the statistical structure of streamflow time-series was studied. We assume that a significant change in a watershed's hydrologic regime will result in a significant change in the statistical structure of streamflow. Recently, it has been shown that streamflow statistical structure is also indicative of streamflow dynamics to some extent (Betterle, et al., 2019) which further justifies studying the changes in streamflow statistical structure to understand the effect of climate change on hydrologic regime.

The statistical structure of streamflow time-series exhibits long-term persistence (Hurst, 1951) meaning that autocorrelations in streamflow decrease very slowly with time-lag. Studying the statistical structure of a stationary time-series is equivalent to studying its spectral properties. Previous work has shown that the power spectral density (PSD) of streamflow scales linearly on log-log graph (Tessier et al., 1996), that is, $h(\omega) \propto \omega^{-\alpha_{\mathrm{h}}}$, where $h(\omega)$ denotes PSD at angular frequency $\omega[T^{-1}]$ and α_{h} denotes the slope of the scaling relationship. Also, a typical streamflow time-series exhibits two scaling regimes (two different values of α_{h}) with scale break occurring between 1-20 days (Hirpa et al., 2010). Kim et al., (2016) analyzed the changes in streamflow PSD to study the effects of urbanization on hydrologic regime in South Korean watersheds. Specifically, they studied the changes in the slopes of two scaling regimes and the change in scale break point. Bras & Rodriguez-Iturbe (1993) and Chow et al. (1978) also illustrated the usefulness of spectral analysis in streamflow time-series analysis. Gudmundsson et al. (2011) studied the contribution of low frequency component (greater than 1-year timescale) to total streamflow variance in several European watersheds, but did not examine the change in the low frequency component over time. A systematic analysis of hydrologic regime change over time driven by climate change has not been reported to the best of authors' knowledge.

The objectives of this study are as follows:

(1) To conduct a spectral analysis of streamflow time-series in watersheds across USA,
(2) To identify temporal changes in those spectral signatures
(3) To identify the spatial patterns of changes in streamflow regimes, and
(4) To investigate the cause of streamflow regime change.

Other researchers have studied the changes in hydrologic regime due to climate change, but their focus has been toward a few of the hydrologic processes or fluxes such as baseflow, soil moisture, annual streamflow etc. Studying the change in spectral properties of streamflow time-series across a large number of watersheds can provide more holistic insight into changes in hydrologic regime.

2. Modeling Description

2.1 FARIMA model

The Fractionally differenced Auto-Regressive Integrated Moving Average (FARIMA; Montanari et al., 1997) model was used to capture the statistical properties of streamflow time-series. FARIMA is a statistical time-series model which is known to capture streamflow structure very well (Montanari et al., 1997 and 2000). The general form of the FARIMA model is

$$\Phi_p(B)(1-B)^d X_t = \Psi_q(B)\epsilon_t, \qquad (1)$$

where X_t denotes streamflow at time-step t, B denotes the backward shift operator such that $BX_t = X_{t-1}$, d denotes a parameter of the model that takes a value between 0 and 0.5 for streamflow time-series, and ϵ_t denotes uncorrelated white-noise. $\Phi_p(B)$ and $\Psi_q(B)$ denote p^{th} order autoregressive and q^{th} order moving average polynomials, respectively,

$$\Phi_p(B) = \sum_{i=0}^{p} \phi_i B^i\,,\ \phi_0 = 1, \qquad (2)$$

$$\Psi_q(B) = \sum_{i=0}^{q} \psi_i B^i\,,\ \psi_0 = 1, \qquad (3)$$

where ϕ_i and ψ_i are AR and MA parameters. Specifically, the terms AR1, AR2, … are reserved to refer to parameters ϕ_1, ϕ_2,…, respectively. Similarly, the terms MA1, MA2, … are reserved to refer to parameters ψ_1, ψ_2 ,…, respectively. When $d = 0$, the FARIMA model degenerates to an ARMA model. When d takes a positive integer value, it becomes classic ARIMA model promoted by Box and Jenkins (1970).

In the case of positive integer d values, the operator $(1-B)^d$ is the differencing operator as can be seen by setting $d = 1$: $(1-B)X_t = X_t - X_{t-1}$. Also, in this case, the process X_t is non-stationary. The interpretation of the model for the fractional d value is not intuitive. But its effect can be understood via the PSD of the process X_t. The PSD of the FARIMA model has the analytical form (Granger and Joyeux, 1980):

$$h(\omega) = |1-z|^{-2d} \frac{|\Psi_q(z)|^2}{|\Phi_p(z)|^2} \frac{\sigma_\epsilon^2}{2\pi},\ z = \mathrm{e}^{-\iota\omega}, \qquad (4)$$

where $|\cdot|$ denotes absolute value and $\iota = \sqrt{-1}$. For very small values of ω,

$$h(\omega) \propto \omega^{-2d}. \qquad (5)$$

The PSD approaches ∞ as ω approaches 0. Also, Eq. (5) tells us that as d increases, $h(\omega)$ increases(Granger and Joyeux, 1980) . In the time-series domain, it means that an increase in the parameter d results in an increase in the amplitude of low-frequency (long timescales) fluctuations.

The effect of different parameters of the FARIMA model on time-series characteristics has been illustrated in Figure 1 with some synthetic time series. In this illustration, the number of AR (p) and the number of MA parameters (q) were fixed to 1. The value of the MA parameter was fixed at 0.5; the values of AR parameter and d were varied. Figure1a shows the time-series generated by setting FARIMA parameters to different values at a daily timescale. Figure 1b and 1c show the moving average of time-series shown in Figure 1a with moving window lengths of 1 month and 1 year, respectively. When the value of d is increased from 0 to 0.25 keeping the AR1 parameter fixed, the two time-series show similar qualitative behavior at daily timescale (Figure 1a). But at the monthly and yearly timescales, the amplitudes of fluctuations are larger when $d = 0.25$. It shows that the parameter d affects the long timescale (low frequency) behavior of the time-series. The short timescale (high frequency) behavior is unaffected by the parameter d. When the AR1 parameter is increased from 0.25 to 0.75 keeping the parameter d fixed, the amplitude of fluctuations becomes larger at all the timescales. Change in AR1 parameter has more profound impact on the daily timescale fluctuations than the change in parameter d. At long timescales, the change in parameter d has more profound impact on time-series fluctuations than the change in AR1 parameter has.

Area under the PSD of a stationary process is equal to the variance of the process (Priestley, 1982). PSD divided by the variance is referred to as normalized power spectral density (NPSD). Also, the NPSD of a stationary process and its autocorrelation function form a Fourier transform pair (Priestley, 1982). Therefore, analyzing the NPSD of a stationary process is equivalent to analyzing its correlation structure. Also, NPSD provides a clean way of separating the contribution of different frequency components to the correlation structure. Therefore, in this study, the NPSD of the fitted FARIMA models was analyzed to detect streamflow regime changes.

Figure 2 shows the NPSD for different values of FARIMA parameters on a log-log graph. In all three cases, increasing the parameter value increases the NPSD values at smaller frequencies, and decreases the NPSD values at higher frequencies. However, the differences are more profound when the value of d is changed. Also, NPSD of only the extremely high frequency components (>0.3 cycles per day) decreases by increasing the MA1 parameter value.

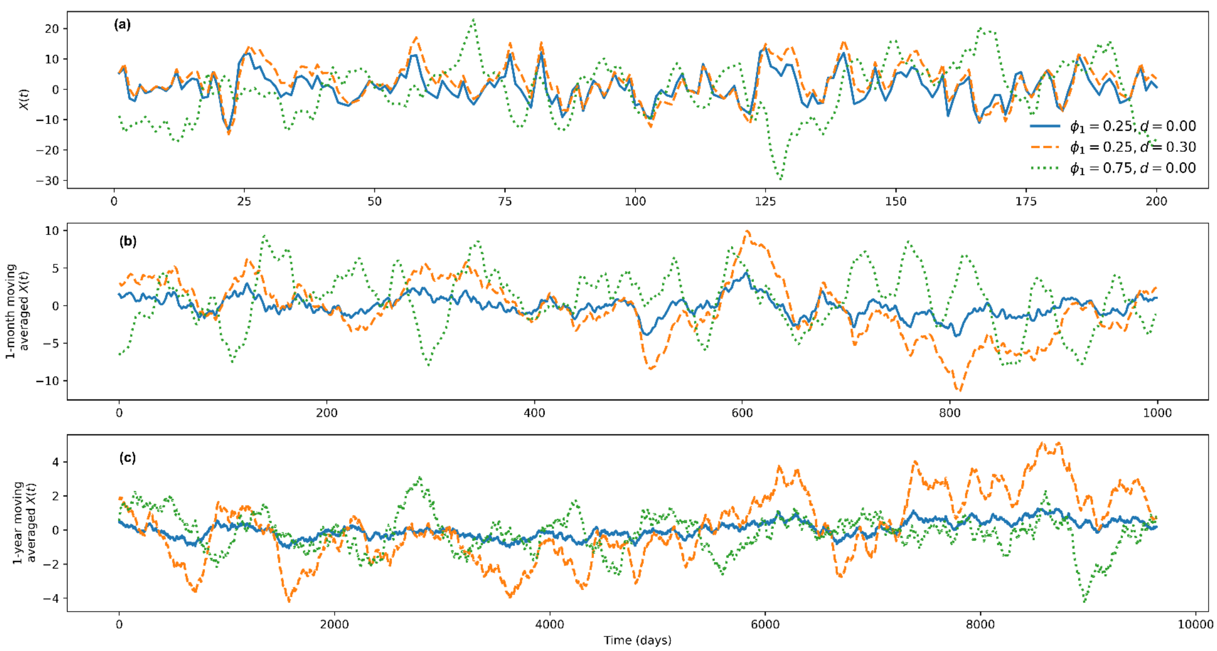

Figure 1. (a) Time-series generated by FARIMA model for different value of AR1 parameter and d parameter at daily timescale; (b) 1-month and (c) 1-year moving average time-series of time-series shown in (a). Time-series was generated for 10000 different timesteps. In subplots (a) and (b), first 200 and 1000 timesteps are shown, respectively, for the sake of clarity.

2.2 Parameter estimation of FARIMA models

Parameters of the FARIMA models were estimated using the same method as that of Monatanari et al. (1997). Details of the parameter estimation method have been provided in Supporting Information (SI). Briefly, a two-step procedure was used to the estimate the parameters. In the first step, a preliminary estimate of the parameter d was obtained using two heuristic methods. The average of the two values obtained using these methods was considered as a preliminary estimate of d. Then the AR and MA model orders p_{opt} and q_{opt} were determined. In the second step, a statistical procedure (see SI) was followed to estimate the parameter d, AR parameters, and MA parameters. In this step, number of AR parameters were fixed to p_{opt} and q_{opt} as obtained in the previous step.

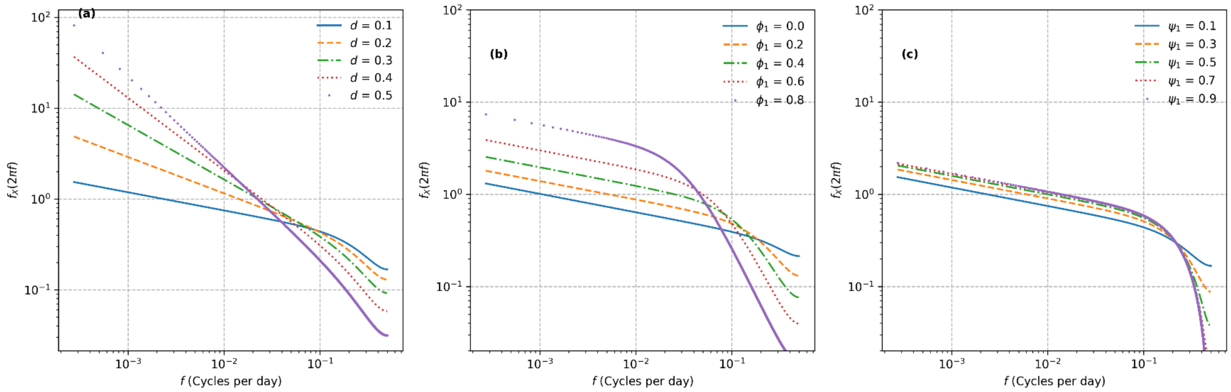

Figure 2. Normalized power spectral density of FARIMA processes for different value of the parameters. The base model has the parameter values $d = 0.1$, $\phi_1 = 0.1$, $\psi_1 = 0.1$. In the subplot (a), (b), and (c), the values of parameter d, ϕ_1 and ψ_1 are changed from their base values, respectively.

To validate the FARIMA models, the autocorrelations of the obtained residual time-series were analyzed. The results are shown in SI. For most of the models, the autocorrelations at any lag were statistically indistinguishable from zero. For a few models, however, the autocorrelation was greater than 0.15 at a few time-steps. These models and corresponding watersheds were removed from the subsequent analysis. The conditions imposed in this study is typically appropriate for model validation (see Montanari et al., 1997). The residuals, however, did not follow the Gaussian distribution for most of the models. But, as pointed out by Montanari et al. (1997) (and the references therein), deviation from Normality does not affect the parameter estimation of FARIMA models.

2.3 Measurement of change in power spectral density

To analyze the changes in hydrologic regime, a moving window approach was taken with the window length of 10 years and with moving step of 3 years (Table. 1). Thus, the study period (1980-2013 water years) was broken up into 9 overlapping windows of 10 years each. The FARIMA model was fit to deseasonalized time-series for different moving average windows as illustrated in Table 1. Thus, as many sets of FARIMA parameters were obtained as the number of moving windows. Each set of parameters results in an NPSD ($f(\omega)$ vs. ω) computed by Equation (4). To detect the changes in streamflow regime, the trend in area under $f(\omega)$ for different ranges of ω was computed (Figure 3). The frequency range was split into five different regions (units in cycles per day – c.p.d.): (1) less than $1/365$ c.p.d. (greater than 1-year timescales), (2) $1/365$ to $1/120$ c.p.d. (4-months to 1-year timescales), (3) $1/120$ to $1/30$ c.p.d. (1-month to 4-months timescales), (4) $1/30$ to $1/15$ c.p.d. (2-weeks to 1-month timescales), and (5) greater than 1/15 c.p.d (less than 2-weeks timescale). For the ease of discussion, two more frequency regions were used: 1/365 to 1/30 c.p.d. (1-month to 1-year timescales) and greater than 1/30 c.p.d. (less than 1-month timescales). The area under NPSD in a given frequency region (ω_1, ω_2) is

$$F(\omega_1, \omega_2) = \int_{\omega_1}^{\omega_2} f(\omega)\,\mathrm{d}\omega, \qquad (6)$$

which is equal to the contribution of the components with frequency between ω_1 and ω_2 to the total variance. Since the area under NPSD is equal to 1, an increase in the contribution of high frequency contribution implies a decrease in low frequency components as is also illustrated in Figure 3.

Let $F_i^j(\omega_i, \omega_{i+1})$ be the area under $f(\omega)$ for i^{th} frequency region and j^{th} time-window. The trend in $F_i^j(\omega_i, \omega_{i+1})$ across time periods can be estimated with a linear fit: $F_i^j(\omega_i, \omega_{i+1}) = \gamma j + c$, where γ is the trend, and c is the intercept. The sign of γ indicates whether the contribution of a frequency region to total streamflow variance is increasing (positive γ) or decreasing (negative γ) over time. The magnitude of γ indicates the extent of change: larger (smaller) magnitude of γ implies larger (smaller) change. A trend was considered statistically significant if the p value of the slope γ was less than or equal to 0.05. We refer to this test as first significance test.

Table 1. An example of moving windows used for analysis.

Window Number	Time-period (years)
1	1980-1989
2	1983-1992
3	1986-1995
4	1989-1998
5	1992-2001
6	1995-2004
7	1998-2007
8	2001-2010
9	2004-2013

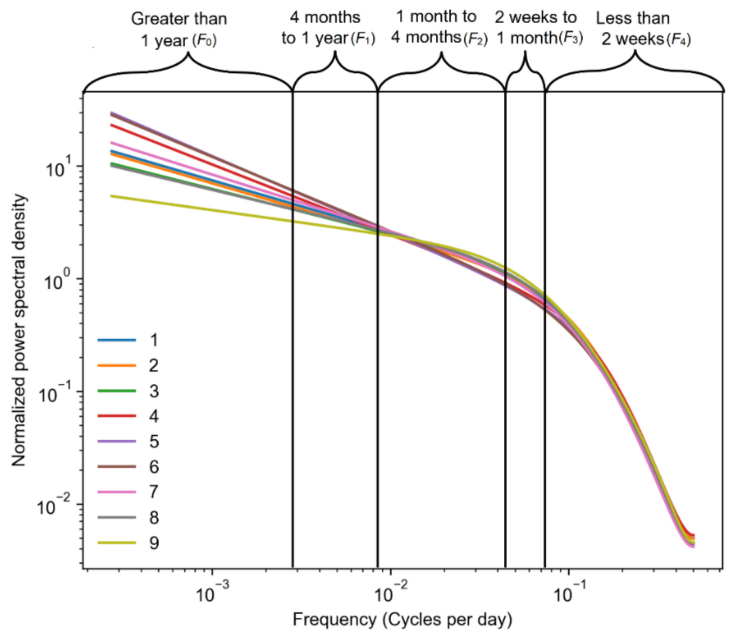

Figure 3. Normalized power spectral density over 9 different time-windows (see Table 1). The frequency range is divided into 5 different regions as labels at the top of the plot.

In addition, statistical significance of each trend was computed by another method. Using the posterior probability distribution of the FARIMA parameters, the posterior probability distribution of NPSD was obtained. This, in turn, was used to compute probability distribution over area under NPSD in each frequency region across the time periods. Thus, for each frequency region, we had probability distribution of $F_i^j(\omega_i, \omega_{i+1})$ for the first and last time-windows. Let these probability distributions be denoted by $P_1(F)$ and $P_2(F)$ with respective mean values m_1 and m_2. For the trend to be significant, we imposed the condition that m_1 and m_2 should belong to different statistical populations. Toward this end, a probability p_s was computed:

$$p_s = \begin{cases} \frac{P_1(F \geq m_2) + P_2(F \leq m_1)}{2}, & m_1 < m_2; \\ \frac{P_1(F \leq m_2) + P_2(F \geq m_1)}{2}, & m_1 \geq m_2. \end{cases} \quad (7)$$

For the trend to be significant, p_s should be less than 0.05. We refer to this test as the second significance test. In summary, a trend was deemed statistically significant only if it came out to be significant using both first and second statistical significance tests. This means that the change in streamflow regime should be consistent in time and the streamflow regime in the first and last time-windows should be significantly different.

We note that Gudmundsson et al. (2011) studied contribution of low frequency components (greater than 1-year timescale) to total streamflow variance in several European watersheds. They estimated this quantity by using the LOWESS method directly instead of using spectral decomposition as discussed above. They did compare their results with those obtained by using the spectral method and concluded that both the methods yield similar estimates. But they only studied the spatial variation of this quantity, not the change in time.

In what follows, area under NPSD in the frequency region greater than 1-year timescale will be denoted by F_0. Similarly, area under NPSD in the frequency region 4-months to 1-year timescales, 1-month to 4-months timescales, 2-weeks to 1-month timescales, less than 2-weeks timescales, 1-month to 1-year timescales, and less than 1-month timescales will be denoted by F_1, F_2, F_3, F_4, F_5, and F_6, respectively.

2.4 Methodology for finding causes of changes in statistical structure of streamflows

To understand the changes in statistical structure of streamflows, statistical methods were used. First, the variables related to the change in $F_i, i = 0,1, \ldots 6$ were identified. Second, possible mechanisms via which each variable might have affected the F_i values were hypothesized. To carry out this analysis watersheds were divided into two groups: snow-dominated and rain-dominated watersheds. The analysis was carried out separately for these two groups.

The variables explored include static catchment attributes including soil properties, geological properties, topography, and climate. Change in climatic statistics were also explored as possible causes of change in F_is. These include change in precipitation related variables and change in temperature related variables. For example, change in total annual precipitation depth, change in OND (Oct-Nov-Dec) total precipitation depth, and change in mean annual temperature. Change in climatic variables was computed using the same moving windows as for the case of change in streamflow statistical structure (Table 1). Additionally, variables capturing snowmelt dynamics in snow-dominated watersheds and rainfall-runoff dynamics in rain-dominated watersheds were also used. The details of these variables are given in section 6 and 7 and in SI. A list of all the variables used in this study is included in Table A1.

Among all the variables, important variables explaining the change in F_i were identified using the random forest algorithm (Brieman, 2002) and simple linear regression. A variable was considered important using simple linear regression if the regression coefficient was statistically significantly different from 0 at 5% significance level. Two linear fits were made for each combination of ΔF_i and predictor variable: (1) using all the watersheds, and (2) using only the watershed for which ΔF_i was significant according to both first and second significance test. All the variables for which the slope of either of the two linear fits was significant at the 5% significance level were considered important. Random forest has the advantage that it can identify non-linear correlations between two variables. However, we found that both the random forest and linear regression yielded the same variables as important.

Though linear regression yields the important predictor variables it can be misleading because of large scatter in the relationship between ΔF_i and other variables. Essentially, the linear fit may have a statistically significant slope, but it is possible that not all the watersheds satisfy the relationship suggested by the line. Therefore, probability densities of important variables conditioned upon the event that ΔF_i was positive or negative were plotted to understand the effect of a variable on ΔF_i. This procedure is similar to computing mutual information between ΔF_i and a variable, but more transparent as shown in section 7.

3. Study area and data

To achieve the objectives of this study, Catchments Attributes and Meteorology for Large Sample studies (CAMELS) dataset (Addor et al., 2017a and 2017b) was used. The CAMELS dataset was chosen because it contains hydro-meteorological dataset for a large number of watersheds (671) across the contiguous USA. Also, the CAMELS watersheds are unregulated and free of anthropogenic land-use changes. The time-period of the data is water years 1980-2013. In this study, we included watersheds that had at least 30 years of complete data; there were a total of 614 such watersheds.

Exploratory analysis shows that significant warming has occurred in CAMELS watersheds across USA. Figure 4 shows the trends in several climatic variables over the study period. These trends were computed as slope of the linear fit on the plot of climatic variable vs. year. A trend was considered statistically significant if the p value of the slope was less than 0.05. Mean minimum daily temperature has increased (positive trend) for most of the watersheds with largest increases across the western US. There exist a few watersheds where the mean minimum daily temperature has decreased (though the trend is statistically insignificant in most of these watersheds). The majority of these cooling watersheds lie in the Great Plains region and Florida (a reference to different hydro-climatological regions is given in Appendix). There exists considerable variation in the trend of mean maximum daily temperatures. Snow-dominated watersheds located in the Rocky Mountains and High Plains have experienced a large increase in mean maximum daily temperatures. Several rain-dominated watersheds located in the Pacific Northwest and Pacific Coast have experienced a decreasing trend in mean maximum daily temperatures. Many of the watersheds located in the eastern USA experienced a negative trend in mean maximum daily temperatures (though statistically insignificant), especially those in the Great Plains. Further, Figures 4c and 4d show trend in OND (Oct-Nov-Dec) and AMJ (April-May-Jun) maximum daily temperature. Maximum daily temperatures in OND months increased across USA with large increases in the arid Great Plains, High Plains, Mississippi Valley, humid Atlantic Coast, and Great Lakes region. The OND maximum daily temperature trends are moderate in the Gulf and Pacific Coast, and the Pacific Northwestern watersheds. Maximum daily temperature in AMJ months has decreased across USA except in western Gulf Coast. Most significant decreases were noted in the Pacific Northwest, Pacific Coast, and Atlantic Coast. As will be discussed below, changes in OND and AMJ maximum temperatures have significant control over changes in streamflow regime.

Figures 4e-4h shows changes in rainfall statistics. There is a strong north-south gradient in the trend in number of rain days: In northern (southern) watersheds, number of rain days have increased (decreased). The trend in number of storms has a weak north-south gradient. In many regions, the number of rainstorms has decreased but number of rain days have increased. This implies that more rain is falling in fewer storms of longer duration in these regions. These regions include the Pacific Northwest and north-eastern part of Atlantic Coast. In the north-eastern part of Atlantic coast, total rainfall depth and mean storm depth has increased. The trend in total rainfall depth has a strong north-south gradient, especially in eastern USA: total rainfall increased in northern watersheds and decreased in southern watersheds. Mean storm depth - the average rainfall depth on rainy days - has more spatial variability compared to the other three rainfall statistics. The only clear patterns are that mean storm depth has increased in the Atlantic Coast region and decreased in the High Plains region.

In summary, Figure 4 convincingly shows that both temperature and rainfall statistics have changed across the USA. Since temperature and precipitation have strong control over hydrologic regime, at least some of the CAMELS watersheds are likely to have undergone a hydrologic regime change. Increase in atmospheric CO_2 can also result in changes in vegetation characteristics such as water use efficiency (Donohue et al., 2013) which, in turn, may affect the hydrologic regime. Significant increases in temperatures along with the fact that global average CO_2 has increased over the period 1980 to 2014 (from 338.91 ppm in 1980 to 397.34 ppm in 2014; Dlugokencky and Tans, gml.noaa.gov/ccgg/trends/, accessed on 17 Mar 2022) indicates significant change in climate has occurred between this period beyond the natural climate variability.

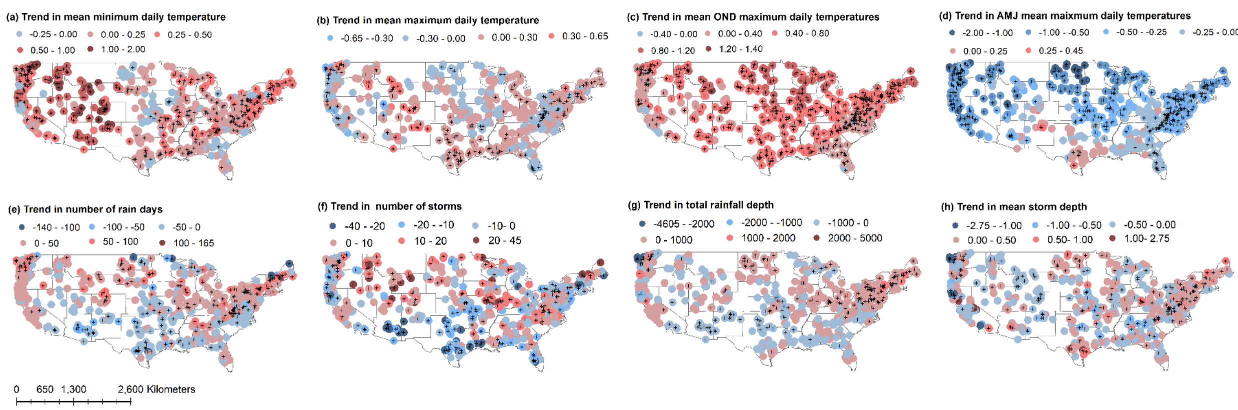

Figure 4. Trends in climatic variables (a) daily minimum temperature, (b) daily maximum temperature, (c) and (d) OND and AMJ daily maximum temperatures, respectively, (e) number of rain days (in days decade^{-1}), (f) number of storms (in decade^{-1}), (g) total rainfall depth (in mm decade^{-1}), and (h) mean storm depth (in mm day^{-1} decade^{-1}). The units of all the temperature statistics are °C decade^{-1}. The red colored symbols indicate positive trend and blue colored symbols indicate negative trend. The '+' sign indicates that trend is statistically significant at 5% level. One time-window refers to 10 years period as indicated in Table 1.

4. Spatial distribution of streamflow regime in USA as measured by NPSD

Figure 5 (a, b, c, d) shows the contribution of different frequency regions to streamflow variance in CAMELS watersheds during the first time-window (1980-1989 water years). Contribution of greater than 1-year timescales components to total streamflow (F_0) was less than 10% in most of the rain dominated watersheds of eastern USA and Pacific Northwest (Figure 5c). Conversely, large contributions from this frequency region were found in snow dominated watersheds in the Rocky Mountains region, the High Plains, the Sierra Mountains in California, and the Pacific Coast.

The contribution of 1-month to 1-year timescale component (F_5; Figure 5b) is very small in the Great Plains and the Mississippi Valley compared to that in other regions. The highest value of F_5 (>50%) was found in snow dominated watersheds of the Rocky Mountains and High Plains. In the Pacific Northwest and the Atlantic Coastal region, F_5 values range from 25 to 50%. The values of F_5 follow the broadscale pattern of baseflow index (BFI; see Figure 4 in Addor et al., 2017). The BFI values are below 0.5 in Great Plains and Mississippi Valley, greater than 0.6 in Rocky Mountains and High Plains, and between 0.40 and 0.60 in Pacific Northwest and Atlantic Coastal region. Moreover, the scatter plot (not shown) of the BFI and F_5 shows that as the BFI increases from 0 to 0.4, the contribution of this frequency region also increases. Beyond, a BFI value of 0.4, however, there exist a few watersheds where F_5 values are low. Overall, the contribution of baseflow to total streamflow appears to be an important factor determining the values of F_5. Interflow might also be responsible for the contribution of 1-month to 1-year frequency region.

The contribution of less than 1-month timescales component, F_6, Figure (5a) to total streamflow variance is small (<25%) in cold snow dominated watersheds of the western USA. In the Pacific Northwest and Pacific Coast, F_6 values are between 25% and 75%, but mostly greater than 50%. In most of the eastern USA watersheds, the contribution of this frequency component is greater than 50%. In the Great Plains and the Mississippi valley, the contribution of this component is greater than 75% in many watersheds. These are dry watersheds where most of the rainwater evaporates back to the atmosphere, and only the intense storms reach the river network. Therefore, the contribution of low (high) frequency components is very low (high) in these watersheds. Since the contributions of low and high frequency components are one-to-one related (an increase in one implies a decrease in other), BFI explains some of the spatial variation in F_6: lower BFI means higher F_6. It is noteworthy that in snow dominated watersheds with the fraction of snow > 0.40 (fraction of precipitation falling as snow), the value of F_6 increases with an increase in mean rainfall.

In rain driven watersheds, a linear relationship (slope $= -0.054$, p-value $= 0.0045, R^2 = 0.033$) between the slope of the flow duration curve (FDC; Addor et al., 2017) and F_6 was found. Smaller slopes of FDC imply smaller variability in streamflow. Thus, the negative correlation between FDC slope and contribution of high frequency region indicates that watersheds with less variability in streamflow values exhibit more contributions from high frequency components. For example, in ephemeral streams, streamflow variability is low as it stays dry during most of the water year; therefore, the low (high) frequency component is very small (large).

The contribution of 2-weeks to 1-month timescale component to total streamflow variance (F_3) is very small for most of the watersheds. But there exist a cluster of watersheds in the Pacific Northwest where F_3 values are greater than 20%. In fact, in most of the Pacific Northwestern watersheds, F_3 values are greater than 15%. The F_3 values are also greater than 15% in several eastern snow dominated watersheds.

It was observed that F_3 was positively correlated with mean precipitation ($R^2 = 0.206$, p-value $= 1.70 \times 10^{-28}$), negatively correlated with potential evapotranspiration (PET; $R^2 = 0.115$, p-value $= 1.62 \times 10^{-15}$). This indicates that F_3 values are high in watersheds with high total precipitation and low ET, i.e., F_3 values are high in humid watersheds. Further, F_3 was negatively correlated with low rainfall frequency ($R^2 = 0.157$, p-value $= 6.15 \times 10^{-21}$) and negatively correlated with high rainfall frequency ($R^2 = 0.093$, p-value $= 1.25 \times 10^{-12}$). It indicates that watersheds where rainfall event characteristics are such that it allows the water to stay in the soils for a long time compared to the timescale of quick flow and percolation, the F_3 values are high. These results indicate that interflow may be responsible for creating 2-weeks to 1-month timescales component. Wu et al., (2021) showed that lateral preferential flows are important streamflow generation mechanism in Pacific Northwestern watersheds.

Figure 5e shows the spatial variation of the parameter d in CAMELS watersheds. There is a large spatial variation in the values of d, but some general patterns can be observed. Very high value of d (>0.30) are typically observed in western snow-dominated watersheds where contribution of low frequency components was significant. In most of the eastern rain-driven watersheds, the d values were less than 0.30. There was strong linear relationship between BFI and d value (slope $= 0.22$, $p \approx 10^{-31}$, $R^2 = 0.23$). Also, the linear relationship was stronger when BFI increased from 0 to 0.25 - at very low value of BFI the d value was close to 0. This indicates that the baseflow is the essential factor for the existence of long-persistence in streamflow time-series. Many of the watersheds in the Pacific Northwest, Great Plains, Great Lakes and Atlantic Coast region had d values less than 0.10, despite having moderately high values of BFI (>0.40) except in the Great Plains. The reason for such small value of d is not clear and further exploration is out of the scope of this paper.

The long-term persistence (high d value) in a time-series may result from aggregation of short-memory processes (Granger, 1980). Muldesee (2007) argued that long-term persistence in streamflow time-series may also be a result of aggregation of several short-memory processes in a watershed. They showed that the value of d increases with increasing drainage area as one moves downstream in a river network. Therefore, it is reasonable to expect that watersheds with large drainage areas may show higher d value in their corresponding streamflow time-series. Such a relation between drainage area and d, however, was not observed in this study.

It can be concluded that long-time scale fluctuations and long-term persistence even in a deseasonalized streamflow time-series are determined by low frequency processes such contribution of baseflow, fraction of snow, and possibly interflow. High frequency components are determined by quick flow, interflow, and ET. Also note that other researchers have reported higher contribution of low frequency component to streamflow (e.g., Gudmundsson et al., 2011) compared to those reported in this study. This is due to the seasonal component of the hydrologic cycle. In this study, the seasonal component had been removed from the streamflow time-series; therefore, F_0 values came out to be smaller.

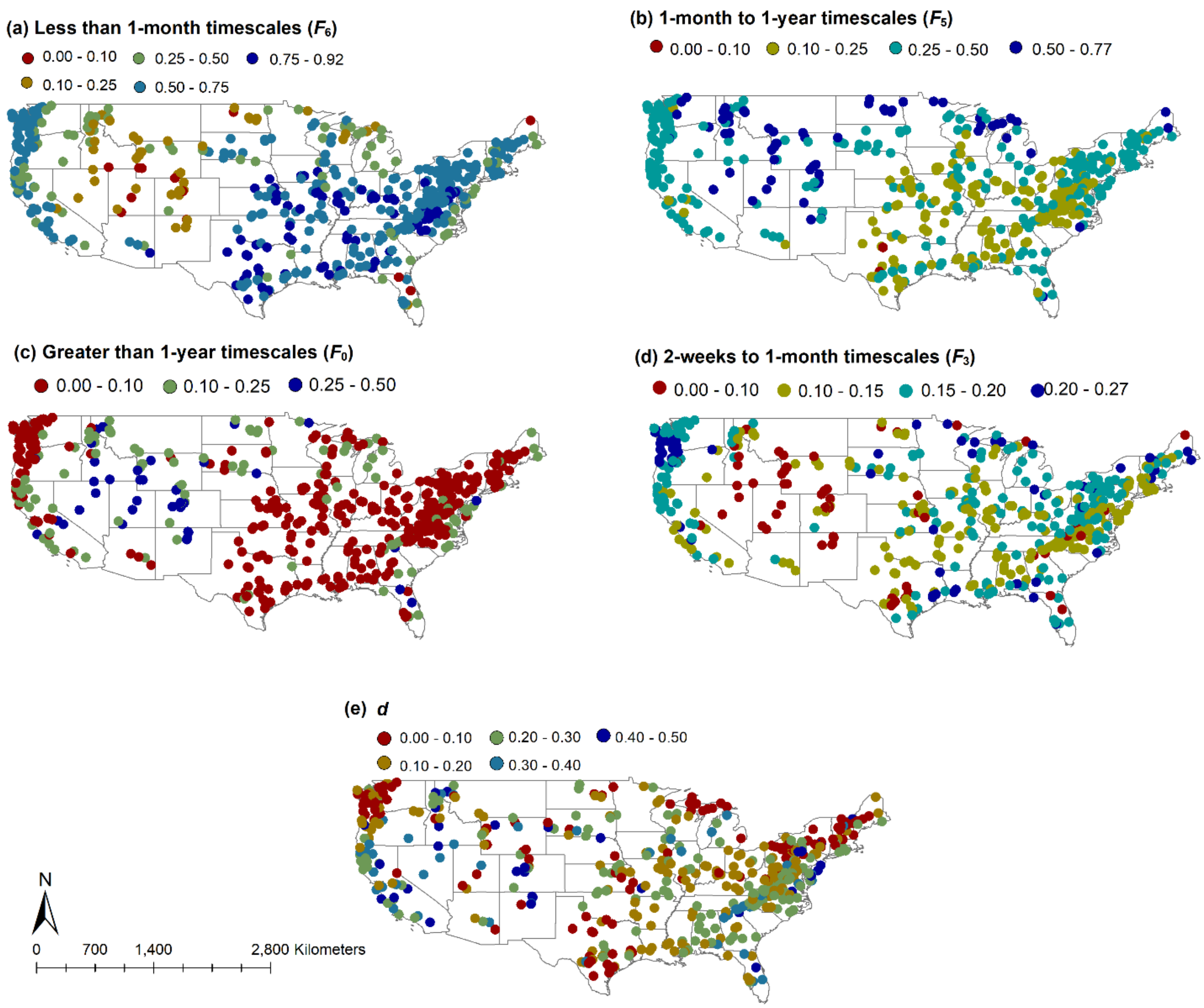

Figure 5. (a), (b), (c) Area under NPSD in different frequency regions, and (d) value of the parameters d across USA. These results correspond to first 10-year moving window.

5. Change in streamflow regime as measured by change in NPSD

Figure 6 shows the spatial distribution of trends in $F(\omega_i, \omega_{i+1})$ for short timescales: Less than 1-month (F_6), 2-weeks to 1-month timescales (F_3), and less than 2-weeks (F_4). Overall, the spatial distribution of trends is patchy. But a spatial structure, albeit weak, is still visible such that watersheds with positive (negative) changes tend to be clustered together in small groups. This is especially true for the watersheds located in the Pacific Northwest, Gulf coast, Atlantic coast, and Great Lakes Region. It indicates that the process that has caused these changes is spatially correlated: change in climate seems to be one of the causes. But climate change alone cannot explain these changes since the correlation length of these trends is significantly smaller than the correlation length of trends in climatic variables such as temperature and rainfall (Figure 4).

Further, it implies that the effect of climate change on streamflow regime is strongly modulated by watershed characteristics such as soil properties, and geomorphological characteristics. This will be explored in subsequent sections.

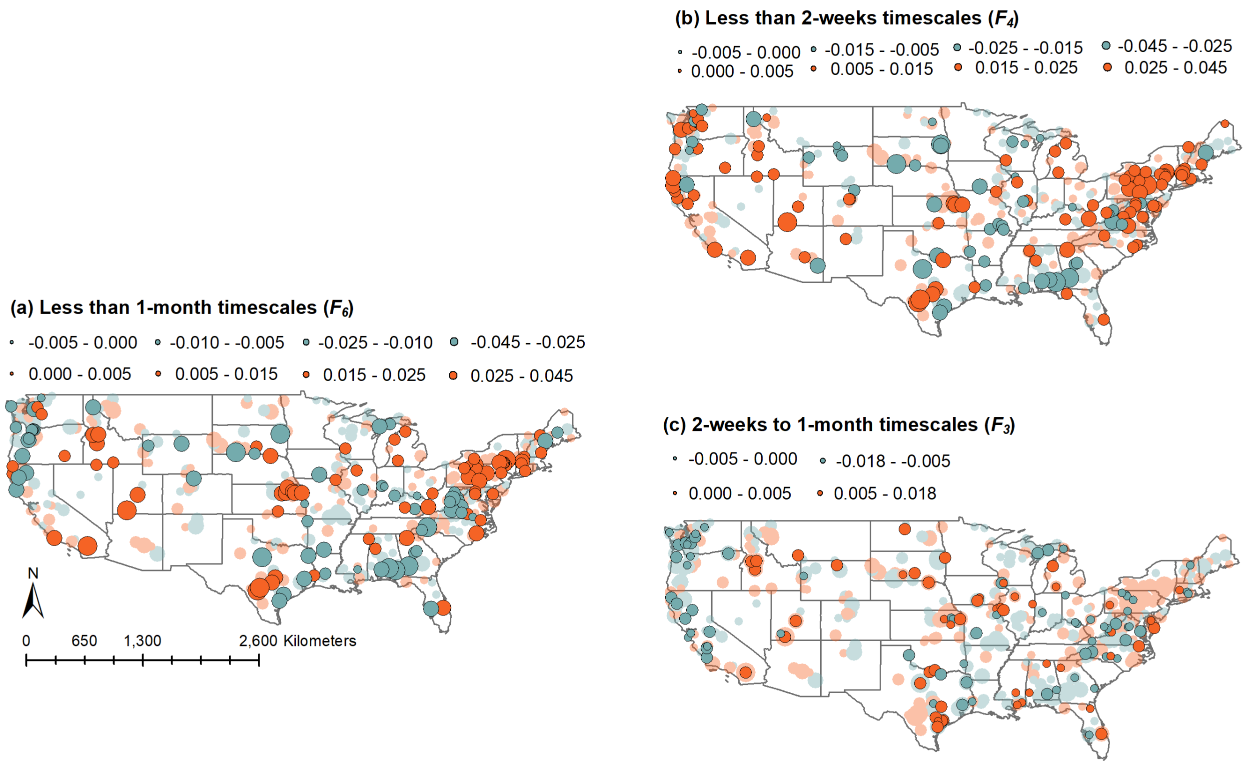

Figure 6. Trend in area under NPSD for high frequency regions (a) less than 1-month timescale, (b) less than 2-weaks timescale, and (c) 2-weeks to 1-month timescale. The watersheds with transparent symbols indicate that the trend is statistically insignificant according to the first significance test. Larger (smaller) sized circles represent larger (smaller) magnitude of change.

Most of the snow dominated watersheds in eastern USA (located in the northern Atlantic Coastal region and Michigan) exhibited positive trends in F_6 and F_4. In western snow dominated watersheds, both negative and positive trends in F_6 and F_4 were observed but most of the statistically significant trends were positive. Watersheds with negative trends were mostly in the eastern Rocky Mountains. The trends in F_3 were positive in most of the Rocky Mountain watersheds and negative in the eastern snow dominated watersheds, but the magnitude of trend was very small compared to that in F_4. Overall, it can be concluded that in majority of the snow-dominated watersheds the contribution of high frequency components to total variance has increased over the study period, with the exception of eastern Rocky Mountains. Several different mechanisms are plausible that could affect this change: (1) Increase in runoff-producing rainfall events, (2) change in temperature snow relationship (Horner et al., 2020), (3) change in snow storage (including spatial distribution), and (4) change in temperature regime. It is likely that the combination of these mechanisms rather than one individual mechanism is responsible for the changes.

In rain driven watersheds, other than spatial clustering of positive trends with positive trends and that of negative trends with negative trends, a few other patterns are visible. Most of the humid watersheds located in the Pacific Northwest region the Gulf Coast region showed a negative trend in F_6. But the trend in F_4 was positive in many of the watersheds in the Pacific Northwest, while in the Gulf Coast the trend in F_4 was also negative. Overall, it appears that humid watersheds are becoming drier which is possible due to change in rainfall statistics in these watersheds. Another possibility is that change in evapotranspiration statistics in these watersheds is caused by change in temperature which, in turn, will change the soil moisture dynamics. A decrease in mean soil moisture in humid watersheds will result in a decrease in the contribution of high frequency components to streamflow. This will be discussed in subsequent sections. In the Great Plains, both increasing and decreasing trends in F_4 and F_6 were observed.

The trend in F_3 showed two clear patterns: (1) Most of the statistically significant trends were negative in the watersheds in the Pacific and Atlantic coastal regions, and (2) Most of the statistically significant trends in the Rocky Mountains, Great Plains, Mississippi Valley, and Gulf Coast were positive. The trends in F_3 were of small magnitude compared to those in F_4 and F_5. This is because the contribution of F_3 (one month to one-year time scales) is very small in most of the watersheds to begin with. A remarkable result is that the F_3 values have decreased in almost all the Pacific region watersheds.

Figure 7 shows the spatial distribution of trends in long timescales fluctuations: Greater than 1-year (F_0), 4-months to 1-year (F_1), and 1-month to 4-months (F_2) timescales. Similar to short-timescale trends, a weak spatial clustering of positive trends with positive trends and negative trends with negative trends is observed for long timescale trends. The magnitude of trends in F_0 is larger in the watersheds located in Western USA. In most of the western snow-dominated watersheds, the value of F_0 decreased, and the magnitude of decrease is relatively large. But the trend was statistically significant only in three watersheds, which might be due to the small magnitude of F_0 value. There is some spatial variability in the F_0 in eastern USA snow-dominated watersheds. This is explained by the fact that in eastern snow dominated watersheds, the contribution of components at greater than 1-year timescales is smaller.

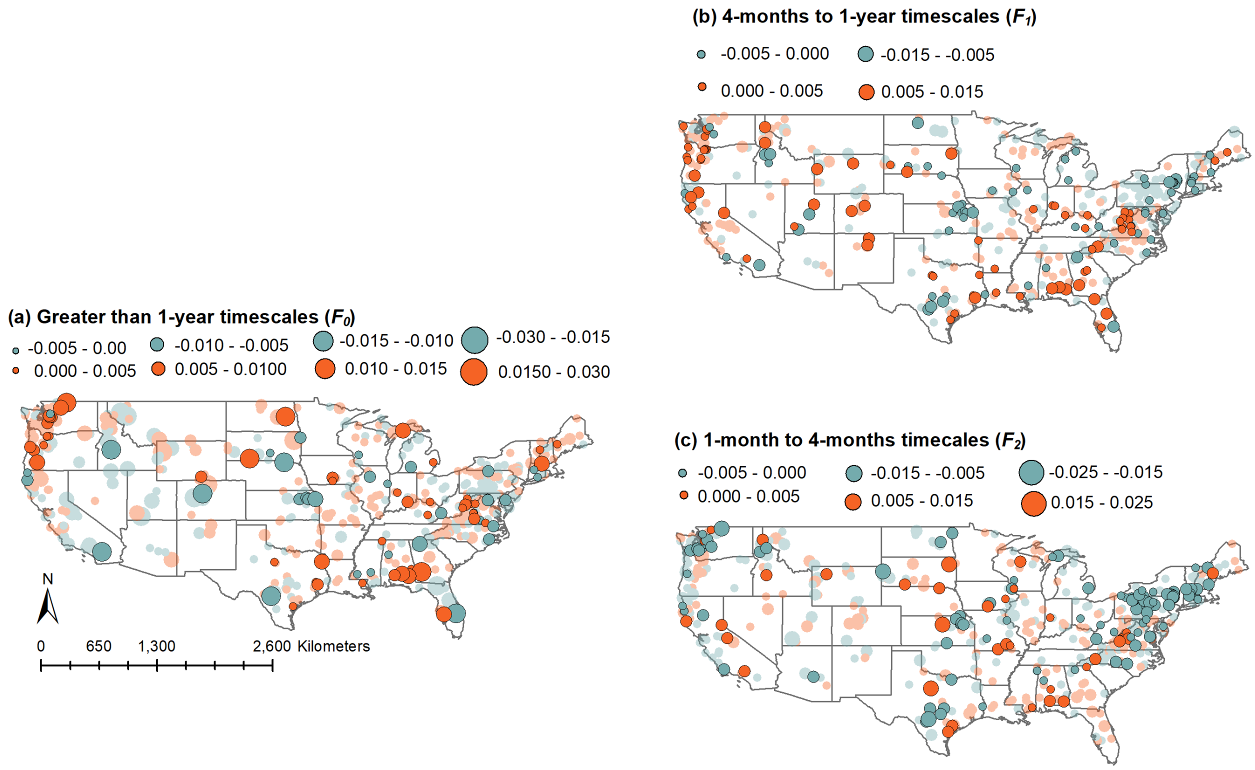

Figure 7. Trend in area under NPSD for low frequency regions (a) greater than 1-year timescale, (b) 4-months to 1-year timescale, and (c) less than 4-months timescale. The watersheds with transparent symbols indicate that the trend is statistically insignificant according to the first significance test. Larger (smaller) sized circles represent larger (smaller) magnitude of change.

The values of F_1 and F_2 decreased in most of the eastern snow-dominated watersheds. The value of F_1 increased in all the snow dominated watersheds in the eastern Rocky Mountains while it decreased in many of the western Rocky Mountains. The reason for difference in trends of eastern and western snow dominated watersheds is discussed below.

Most of the rain-dominated watersheds in the Pacific Northwest exhibited positive trends in F_0 and F_1, and negative trends in F_2. Similarly, most of the watersheds in the Pacific Coast exhibited negative trends in F_0 though trend was statistically significant only for one watershed. The trends in F_0, F_1, and F_2 were positive in most of the Gulf Coast watersheds. Most of rain dominated watersheds in the Great Plains exhibited a decrease in F_0, F_1, and F_2. But there were several watersheds in this region where F_0, F_1, and F_2 increased.

In summary, streamflow statistical structure has changed in many of the watersheds across USA. There is some spatial structure in the regime change: watersheds close to each other show similar types of changes. The spatial structure of change in snow dominated watersheds is stronger than in rain-dominated watersheds. Also, the western and eastern snow dominated watersheds showed some difference in trends in long timescale components. In the western watersheds, the negative trends were observed in F_0 values. In the eastern watersheds, the negative trends were observed in F_1 and F_2. Also, positive trends in F_1 were observed in western snow dominated watersheds. In the humid watersheds of the Pacific Northwest and Gulf Coast, contribution of high frequency components decreased. The next two sections focus on the causes of regime change in snow and rain-dominated watersheds, respectively. The discussion of causes of change in high frequency and low frequency effects is generally limited to the F6 and F5, respectively.

6. Causes of streamflow regime change in snow-dominated watersheds

In this section, we explore the causes of streamflow regime changes in snow-dominated watersheds. Most of these watersheds are in the Rocky Mountains, High Plains, and the Atlantic region. There are other watersheds where snowmelt contributes to streamflow, but rainfall is the primary driver in those watersheds. In snow-dominated watersheds, snowmelt is the primary driver of streamflow. Snow accumulates during the winter season during low temperatures and melts during spring and early summer due to rising temperatures. The process of snowmelt is largely controlled by the amount and spatial distribution of snowpack, measured as snow water equivalent (SWE), and dynamics of temperature. The changes in streamflow regime in snow-dominated watersheds may occur due to change in the SWE and/or temperature dynamics. Change in either of the two will result in the change in temperature-snowmelt relationship. Note that precipitation falls as liquid also in these watersheds but that is the secondary determinant of streamflow regime.

In this study, snow signatures proposed by Horner et al., (2020) were used to identify the changes in temperature snow relationship. They defined streamflow, temperature, and SWE regimes as a 30-day moving average of seasonal component. Let us denote streamflow, temperature, and SWE regimes by Q_{reg}, T_{reg}, and SWE_{reg}, respectively. Figure 8 shows the relationship between temperature and streamflow regimes for a hypothetical snow dominated watershed. The segment AB is the snowmelt period where both streamflow and temperature rises. Streamflow reaches its peak at point B. After point B, temperature continues to rise but streamflow decreases because of the lack of snow availability. During segment CD, temperature decreases without significant change in streamflow. During the segment DA, snow accumulates. The segments AB and CD capture the snowmelt dynamics. Horner et al. (2020) fitted linear relationships between temperature and streamflow regimes to model segments AB and CD and defined the slopes of these segments as snow signatures. In the study, we found that the linear relation was a good model for the segment AB but not for the segment BC. Therefore, we focused only on segment AB which we refer to as the rising limb of temperature-streamflow relationship. Let this relationship be modeled as

$$\hat{Q}_{\mathrm{reg},i} = \delta_{\mathrm{snow}} T_{\mathrm{reg},i} + \beta_{\mathrm{snow}}, \tag{8}$$

where $T_{\mathrm{reg},i}$ and $\hat{Q}_{\mathrm{reg},i}$ denote the temperature and estimated streamflow regime value on i^{th} day of the water year during the first phase of snowmelt (limb AB), δ_{snow} and β_{snow} denote the slope and intercept of the relationship. We used both δ_{snow} and β_{snow} as the snow signatures.

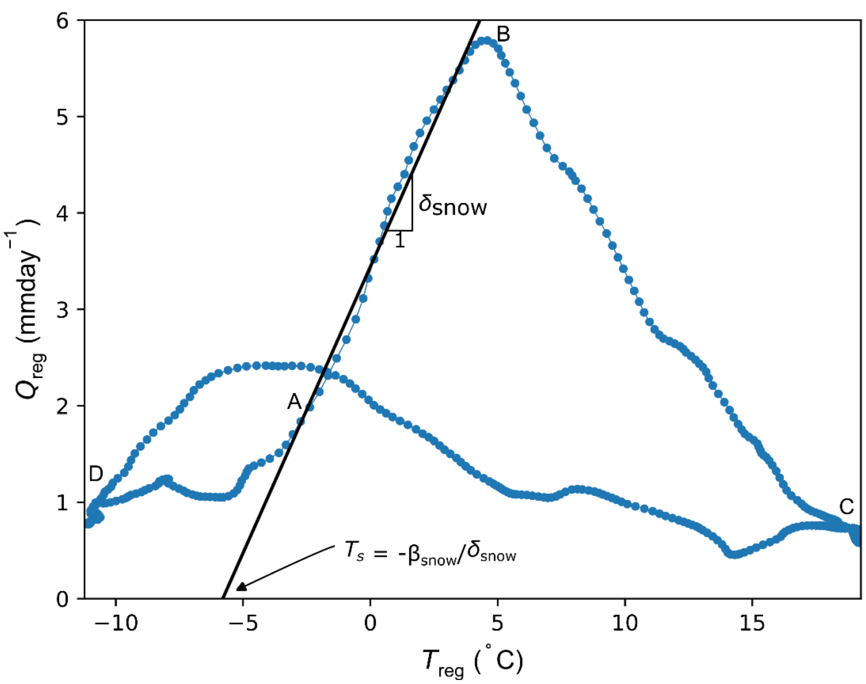

Figure 8. Relation between the temperature and streamflow regimes. T_{reg} is the temperature regime of the mean watershed temperature. T_{s} denotes the threshold mean watershed temperature at which snowmelt starts. The locations of the points A, B, C, and D is approximate.

The slope, δ_{snow}, is a measure of rate of increase of snowmelt per unit increase in temperature. The intercept β_{snow} is the streamflow when the mean temperature is zero and snowmelt has not started. An intuitive way of thinking about β_{snow} is as follows. For a given value of δ_{snow}, the value of β_{snow} determines the point where line AB intersects with the x-axis ($Q_{\mathrm{reg}} = 0$). By making Q_{reg} equal to 0 in Eq. (8), one gets $T_{\mathrm{reg}} = \beta_{\mathrm{snow}}/\delta_{\mathrm{snow}}$. Thus, given δ_{snow}, the intercept β_{snow} is the *measure* of threshold mean watershed temperature required to start the snowmelt. Keeping the δ_{snow} fixed, higher β_{snow} implies smaller values of threshold temperature and smaller values of β_{snow} implies larger values of threshold temperature. But note that β_{snow} is *not equal to* the threshold temperature required to start the snowmelt. Along with δ_{snow} and β_{snow}, time to peak – number of days since the start of the water year after which streamflow regime peaks – was also computed as a snow signature. We computed the snow signatures for the moving time windows of 10 years each as illustrated in Table 1. Subsequently, trends in these signatures were computed over the time-windows. The trend values provide an estimate of change in snow signatures. The trends in these snow signatures are discussed in SI. In the context of this paper, trends in snow signature are related to the change in snowmelt dynamics.

Next, we look at how the change in snowmelt dynamics along with other watershed properties have affected the streamflow regime as obtained by the FARIMA model. Figure 9 shows the important predictor variables that determine the change in F_6, the high frequency ($<$ 1 month) components. Blue and orange solid are the probability densities of variables conditioned upon the positive and negative trends for all the watersheds, respectively. Green and red dash curves are the probability densities of variables conditioned upon the positive and negative trend for all the watersheds where trend was statistically significant. Several important variables were related to the change in rainfall statistics: trend in mean storm depth, trend in JAS (July-August-September) average rainfall depth, trend in average high rainfall duration and depth, and trend in total storm depth. Increase in all these statistics is associated with an increase in F_6. For example, watersheds where mean storm depth increased, positive change in F_6 was more likely. This is expected because an increase in high rainfall duration, and depth would result in an increase in high frequency fluctuations. The same argument applies for increase in mean and total storm depth. The mean storm depth increased in most of the eastern snow dominated watersheds (Figure 4). It tells us that increase in F_6 in eastern snow dominated watersheds is related to increase in the precipitation.

Mean watershed temperature is another important variable. Watersheds with warmer temperatures were more likely to result in an increase in F_6 than those with colder temperatures. It might be related to the fact that, in western USA, SWE is decreasing at a higher rate in warmer watersheds than that in colder watersheds (Mote, 2006). Disappearance of snow would reduce the contribution of low frequency component of streamflow and, by implication increase the contribution of high frequency component.

Another temperature related important variable is the trend in AMJ (Apr-May-Jun) maximum daily temperature. This quantity has decreased in most of the watersheds. In the watersheds with moderate (large) decrease, the F_6 was likely to increase (decrease). To investigate the effect of changes in AMJ maximum daily temperature on the change in F_6, the probability density plots of all the predictor variables were plotted conditioned upon AMJ maximum daily temperature being less and greater than -0.20. It was observed that the significant decrease in AMJ maximum daily temperature occurred in humid watersheds and in watersheds with aridity index less than 1.5. About 65% of the watershed with the moderate decrease in this quantity were arid. The snow dominated arid watersheds are primarily located in western USA. The snow dominated humid watersheds are primarily located in eastern USA, Pacific northwest, and Northern Rocky Mountains. Thus, change in AMJ maximum daily temperature has different effects in wet/moderate-dry and dry watersheds. The mechanism behind the effect of AMJ temperature was unclear.

Soil properties that were important in determining the trends in F_6 were sand fraction, silt fraction, soil conductivity, soil depth, and depth to bedrock. Watersheds with sandy and high conductivity soils were more likely to exhibit a decrease in F_6. Watersheds with clayey and low conductivity soils were more likely to exhibit an increase in F_6. One of the differences between the watershed with clayey and sandy soils was that in the former the average high rainfall depth increased more significantly. In $\approx 20\%$ of the watersheds with sandy soils, average high rainfall depth decreased. In the watersheds with clayey soils, the OND (Oct-Nov-Dec) temperatures increased moderately, whereas in the watersheds with sandy soils, the OND temperatures increased significantly. Also note that in most snow dominated watersheds, the high rainfall occurs mainly in winter season. These observations lead to the following hypothesis. In the watersheds with clayey soils, increase in high rainfall depth together with only moderate increase in winter maximum daily temperature is responsible for the increase in F_6: moderate increase in winter maximum daily temperature ensures that soil moisture does not decrease significantly. In the watershed with sandy soils, decrease or only a moderate increase in high rainfall depth with large increase in winter maximum daily temperature is responsible for significant decrease in soil moistures. This decrease in soil moisture is responsible for decrease in F_6.

Finally, trend in δ_{snow} and trend in time-to-peak are important variables for determining the change in F_6. Higher the increase in δ_{snow}, higher the increase in F_6; higher the decrease in time-to-peak, higher the increase in F_6. Both, the increase in δ_{snow} and the decrease in time-to-peak suggests an increase in snowmelt rate. This, in turn, implies that water is reaching the river network faster which decreases the contribution of low frequency component and increases F_6 values. In summary, in snow-dominated watersheds change in rainfall depth and duration, increase in winter (OND) and decrease in spring (AMJ) temperatures, and change in streamflow-temperature relationship is responsible for change in F_6.

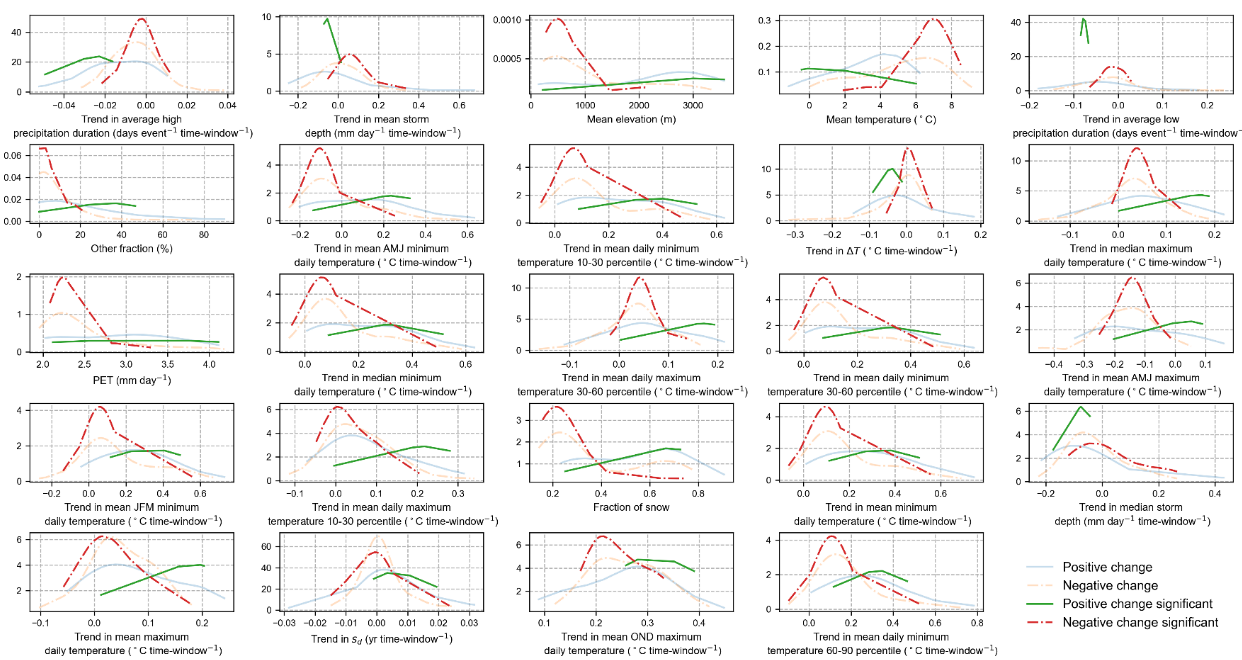

Figure 9. Snow-dominated watersheds. Probability distribution of important predictor variables at less than 1-month timescale

Figure 10 shows the probability distribution of important variables that determine the change in the contribution of 1-month to 1-year timescale components (F_5) – only the 24 most important variables are shown in the figure. Rainfall related important variables were the trend in high rainfall duration, trend in mean and median storm depth, and trend in total storm depth. Increase in mean, median, and total storm depth was associated with a decrease in F_5. High rainfall duration decreased in most of the watersheds. If the decrease in average rainfall duration was large, then the watershed was more likely to exhibit an increase in F_5; if the moderate decrease or increase in average rainfall duration was observed, watershed was likely to exhibit a decrease in F_5. As discussed above, changes in rainfall statistics also explained changes in F_6. Basically, increase in storm depth and increase in high rainfall duration are related to increase in high frequency components and decrease in low frequency components.

Mean elevation, mean temperature, and fraction of snow were also important variables. Watersheds with lower (higher) mean elevation, higher (lower) mean temperature, and smaller (higher) value of fraction of snow were more likely to exhibit a decrease (increase) in F_5. The threshold value of fraction of snow at which the sign of change in F_5 transitions from negative to positive is 0.4. The fraction of snow is less than 0.4 in eastern US snow dominated watersheds and greater than 0.4 for most of the western snow dominated watersheds (Figure 3 in Addor et al., 2017). This indicates that the change in F_5 is different in eastern and western US watersheds which was also observed in Figure 7. Moreover, Figure 7 clearly shows that F_5 $(= F_1 + F_2)$ decreased in most of the eastern snow dominated watersheds while it increased in western snow dominated watersheds.

Further investigation revealed that in the majority of the eastern snow dominated watersheds the following quantities have increased: number of rain days, total storm depth, and mean storm depth (Figure 4). As discussed above, increase in these quantities is related to increase in F_6, thus, almost by implication decrease in F_5. Figure S10 shows that in eastern snow-dominated watersheds SWE increased over the study period. In general, increase in SWE is expected to result in increase in F_5. Therefore, it may be concluded that in eastern snow-dominated watersheds change in rainfall statistics is the dominant control over change in streamflow regime. We caution here that this statement is applicable to deseasonalized streamflow time-series only. The seasonal component of streamflow may have been profoundly impacted by the change in SWE.

In western US snow dominated watersheds, the change in F_5 had large spatial variability. The SWE decreased in most of these watersheds (Figure S10). Change in rainfall statistics has some spatial variability but the following general observations can be made: (1) total storm depth has decreased or has only slightly increased, (2) mean storm depth has decreased in most watersheds but there exist some watersheds in the Southwest region with significant increase, and (3) number of storms and number of rain days have increased (decreased) in most of the northern (southern) watersheds. Therefore, it can be concluded that change in rainfall statistics have at least some control over change in streamflow regime in western snow dominated watersheds also. In summary, the differences in change in rainfall statistics explain the differences in changes in F_5 in eastern and western snow-dominated watersheds.

Another observation was that several temperature related variables were important for determining the change in F_5. Some of these variables include trend in AMJ minimum and maximum daily temperatures, trend in mean daily minimum and maximum temperatures, trend in mean JFM minimum daily temperature, and trend in mean OND maximum daily temperature. Both mean minimum and maximum daily temperatures increased in most of the snow dominated watersheds. A moderate increase was associated with a decrease in F_5 and a significant increase was associated with an increase in F_5. As discussed above, increase in temperature affects soil moisture regime which, in turn, affects the streamflow regime. However, change in temperature can also directly affect the low frequency components of streamflow, for example, via change in baseflow characteristics, and change in snowpack storage. These mechanisms have been discussed above.

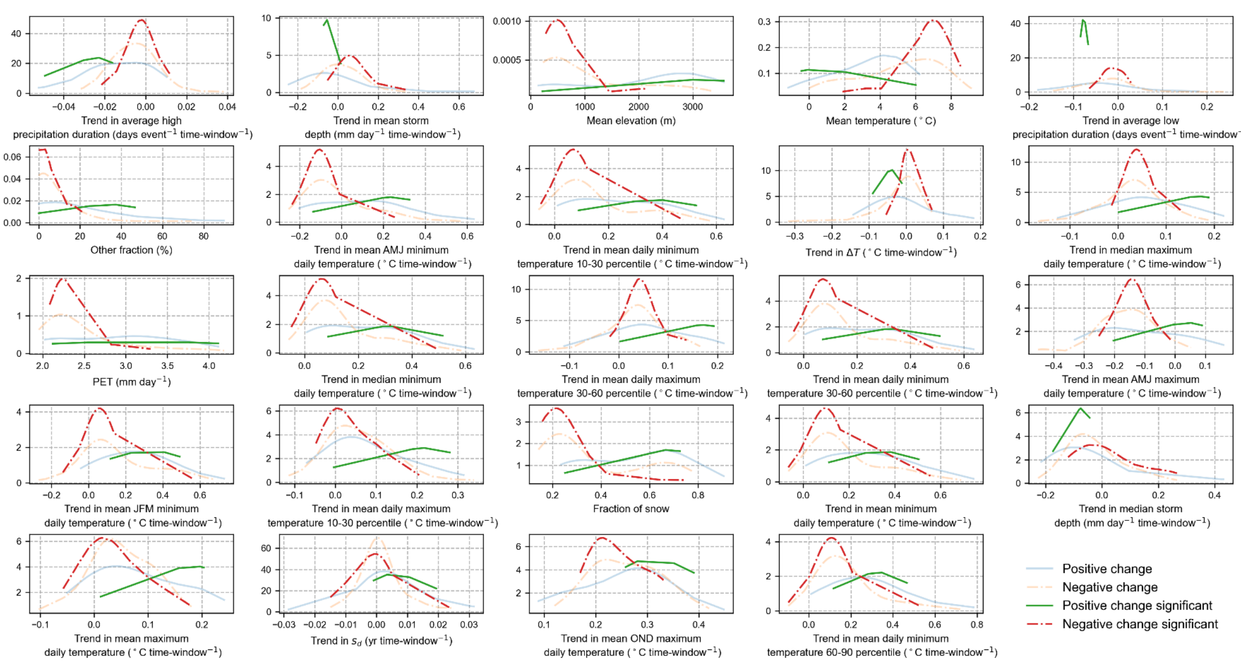

Figure 10. Snow-dominated watersheds. Probability distribution of important predictor variables at 1-month to 1-year timescales

7. Causes of streamflow regime changes in rain dominated watersheds

In rain dominated watersheds rainfall is the primary driver of streamflow. Some of the rainwater is intercepted by the plant canopy and other structures, some of the rainwater infiltrates into the soil, and the rest of the rainwater runs off and eventually reaches the rivers. Most of the intercepted rainwater evaporates back to the atmosphere. Some of the infiltrated water goes to groundwater through percolation, some of the infiltrated water goes back to atmosphere in the form of soil evaporation and plant transpiration, and rest of the infiltrated soil water flows below the earth surface to nearby streams which is referred to as interflow. Groundwater also flows to the river, which is referred to baseflow. These processes occur at vastly different timescales and are affected strongly by several watershed properties including their spatial distribution. It is possible that change in the rainfall-runoff response of a watershed is responsible for change in streamflow regime in rain-driven watersheds. In this study, we used a conceptual event-based model to simulate rainfall-runoff response of rain-driven CAMELS watersheds.

The details of the modeling are discussed in SI. In summary, hydrograph separation was carried out using streamflow and rainfall data in each of the watersheds (Lamb and Beven, 1997; see Collischonn and Fan et al., 2013 for hydrograph separation). Each rainfall-runoff event was modeled using the SCS-CN method (Ponce and Hawkins, 1996; Mishra and Singh, 1999; Geetha et al., 2007; Soulis and Valiantzas, 2012; Soulis and Valiantzas, 2013) and 2-parameter gamma distribution as unit hydrograph (Botter et al., 2013). There were a total of four model parameters $\lambda, CN, \alpha,$ and β. The first two parameters belong to the SCS-CN model and the last two parameters belong to unit hydrograph. The mean and variance of the unit hydrograph is α/β and α/β^2, respectively. These parameters were estimated for each of the rainfall-runoff event using the Dynamic Dimension Search (DDS) algorithm (Tolson and Shoemaker, 2007) with the objective of minimizing mean-square-error between observed and simulated direct runoff. Once these parameters are obtained for each of the rainfall-runoff events, then the change in these parameters over time can be used as a measure of the change in the rainfall-runoff response of a watershed. One difficulty is that these parameters have high variability from event to event. Therefore, the change in probability distributions of these parameters had to be measured. This was achieved using the moving windows as illustrated in Table 1. All the events contained in a moving window were used to create a probability distribution of the four parameters. The change in probability distribution was measured by estimating the trend in several statistics of the probability distributions which includes mean, mean of 0-10 percentiles, mean of 10-30 percentiles, mean of 30-60 percentiles, mean of 60-90 percentiles, and mean of 90-100 percentiles. The important variables were recognized using the same method as in snow dominated watersheds.

Figure 11 shows the conditional probability density of important variables for the classification of positive and negative trends at less than 1-month timescale (F_6) in rain dominated watersheds. Some of the important variables are OND mean maximum daily temperature, trend in median minimum daily temperature, and aridity. The value of F_6 increased in many of the arid watersheds while it decreased in most of the humid watersheds. Further, F_6 increased in the watersheds in which OND maximum daily temperature increased significantly. It was observed that arid rain-driven watersheds had higher increase in OND maximum daily temperature (Figure 4), higher increase in number of dry days, higher increase in JAS maximum and minimum daily temperature, and decrease in monthly rainfall variation. Also, changes in average rainfall depth in arid watersheds during OND and JAS months were small (not shown). All these factors indicate that the increase in evaporation is more than the increase in rainfall in the arid watersheds which has resulted in the decrease of low frequency components of streamflow in these watersheds. And the decrease in low frequency components is responsible for increase in high frequency components. Figure 11 also shows that increase median minimum daily temperature is associated with increase in F_6. This further supports the hypothesis that decrease in contribution of low frequency components in arid watersheds is due to increase in evaporation, and subsequent decrease in low frequency component.

Many of the humid watersheds where F_6 decreased are located in the Pacific Northwest and the Gulf Coast region where rainfall is more frequent in winter months. It was observed that OND rainfall depth decreased in most of the humid watersheds and OND temperature increased moderately in these watersheds. These two factors can explain the decrease in F_6 in these watersheds. Increase in temperature implies higher potential evaporation and higher actual evaporation (because humid watersheds are energy limited), and lesser soil moisture. Thus, more rainwater is absorbed by the soils and lesser rainwater reaches the river network in the form of direct runoff. Decrease in rainfall further amplifies this process. Other observations that support this hypothesis are decrease in median storm depth and decrease in high rainfall duration in most of the watersheds. Ficklin et al. (2016) also reported decrease in quick runoff in several watersheds located in the Pacific Northwest and the Gulf Coast which supports this hypothesis.

The values of F_3 have decreased in almost all the Pacific Northwest watersheds. As discussed above, the value of F_3 is partially determined by ET: increase in ET results in decrease in F_3. Therefore, the decrease in F_3 and F_6 in these watersheds suggest the role of temperature in changing the streamflow regime. The value of F_4 increased in some of the watersheds in Pacific Northwest (Figure 6). The reason for this is unclear.

Some of the rainfall related variables such as trend in low rainfall frequency, trend in low rainfall duration and frequency, trend in number of rain days, low rainfall frequency and mean rainfall were also important. These variables are also related to aridity and humidity of the watersheds. Watersheds with low mean rainfall and larger number of dry days are typically arid. Most of the watersheds where number of rain days decreased, number of dry days increased, and low rainfall duration increased, F_6 also increased. This is expected because these trends indicate an increase in aridity of the watershed and arid watershed are known to exhibit high values of F_6. Figure 11 also shows that in most of the watersheds where F_6 has increased, number of rain days have decreased.

Some of the soil properties such as sand fraction and porosity including fraction of forests are also important variables. Most of the watersheds with sandy, smaller porosity soils and large fraction of forest cover exhibited a decrease in F_6. These three variables are correlated since sandy soils are known to be porous and ideal to support forests given the water availability (Eagleson, 1982). It was observed that most of the CAMELS watersheds with sandy soils are located in humid regions with high mean annual rainfall. Thus, the decrease in F_6 in watersheds with sandy soils can be explained as in humid watersheds as discussed above. Another difference between watersheds with sandy and fine soils was that in the former the phase difference between monthly rainfall and evaporation decreased which might have resulted in more rainwater evaporating back to atmosphere, drying of soils, and muted response of watershed to rainstorms. Many of the watersheds in the Pacific Northwest have sandy soil.

One notable point in above discussion is that OND maximum temperature has increased in most of the watersheds, located in both humid and arid climates. In humid watersheds increase is moderate and in arid watersheds increase is large. But this increase has opposite effect on streamflow regimes in humid and arid watersheds. In humid watersheds, increase in OND temperature resulted in an increase in ET, decrease in soil moisture, and a muted response of the watershed to rainfall which resulted in a decrease in high frequency component. In arid watersheds, increase in OND temperature resulted in an increase in ET and a decrease in low frequency component which, in turn, resulted in an increase in high frequency component. Thus, change in OND temperature directly affects the high frequency component in humid watersheds and only indirectly affects it in arid watersheds.

One question remains here: Why the high frequency component is not directly affected by change in OND temperature in arid watersheds? The reason is that in majority of rain driven arid watersheds in USA, rainfall pre-dominantly occurs in spring-summer months (except in California where rain occurs in winter months) (Addor et al., 2017, Fig 3). Thus, an increase in ET in winter months directly affects only the low frequency component, not the high frequency component. High frequency component is formed by the summer rainfall which appears to be unchanged during the study period. This conclusion is further supported by the fact that AMJ (Apr-May-Jun) and JAS (Jul-Aug-Sep) maximum daily temperatures have not increased significantly in these watersheds. AMJ minimum daily temperature also did not increase in most of the watersheds. JAS minimum daily temperature increased significantly only in a few of the arid watersheds (<40%). In contrast to arid watersheds, rainfall occurs in winter months in many of the humid watersheds, especially the ones located in Pacific Northwest. Therefore, change in temperature directly affects the high frequency component in humid watersheds.

Finally, two of the parameters of the rainfall-runoff model came out to be important for determining the streamflow regime change: CN and λ. Decrease in CN and increase in λ seems to be associated with an increase in F_6. This association, however, is weak because several of the watersheds where CN decreased also reported a decrease in F_6. Also, the change in CN and λ is relatively small is most of the watersheds. Therefore, we conclude that change in streamflow regime in rain driven watershed is a direct result of change in climate statistics rather the change in rainfall-runoff response of the watershed.

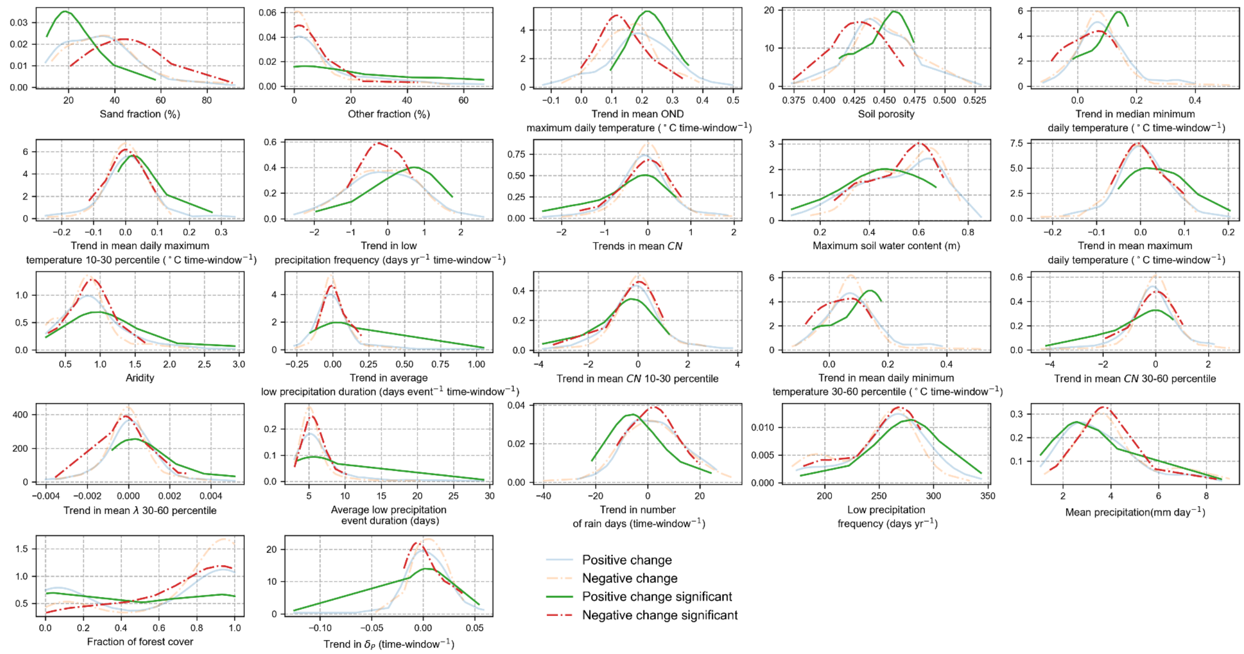

Figure 11. Rain dominated watersheds. Probability distribution of important predictor variables for classification of positive and negative trends at less than 1-month timescales

The causes for change in low frequency components is not discussed because fluctuation at greater than 1-year timescales had very small contribution to total streamflow variance in rain dominated watersheds. And, therefore, the contribution of 1-month to 1-year timescale components is almost one-to-one related to less than 1-month timescale contribution.

8. Summary and Conclusions

The main conclusions of this study are summarized in Table 2. It was found that the effect of climate change on streamflow regime change was strongly modulated by watershed static attributes. The contribution of greater than 1-year timescales fluctuations to total streamflow variance is typically very small in rain-driven watersheds, but it is substantial in western snow dominated watersheds where the fraction of snow is greater than 0.4. The contribution of 1-month to 1-year timescale fluctuations strongly depends upon the contribution of baseflow to total streamflow. Also, long-term persistence (value of d) in deseasonalized streamflow time-series depends upon the contribution of baseflow: low values of BFI are associated with weaker long-term persistence. The contribution of 2-weeks to 1-month timescale fluctuations to total streamflow variance appears to be determined by interflow and rainfall. Contribution of high frequency components are mainly determined by quick flow. Thus, spectral analysis of deseasonalized streamflow time-series can be very useful in detecting hydrologic regime changes in a watershed through analysis of streamflow time-series.

In snow-dominated watersheds across the USA, a clear east-west divide was found in terms of change in streamflow regime. F_1 and F_2 decreased (increased) in most of the eastern (western) watersheds. F_0 decreased in most of the western watersheds. The high frequency components increased in most of the snow dominated watersheds. Increases of high frequency components and decreases in low frequency components in snow dominated watersheds were related to increases in rainfall in these watersheds but also to increase in OND temperatures. It could be concluded that trends in rainfall have significant control over streamflow regime change in snow dominated watersheds. Changes in snowmelt-temperature relationships also played a role in changing the streamflow regime in snow-dominated watersheds.

In most rain-driven watersheds and in eastern snow dominated watersheds, the contribution of high frequency (less than one-month) components was greater than 50%. This was particularly the case in the watersheds in the Great Plains and the Mississippi Valley where the contribution of low frequency component is very small due to high ET. In most of the arid watersheds, the values of F_4 and F_6 increased. These increases are related to increases in ET in these watersheds in winter months which decreased contributions from low frequency components and, in turn, increased the contribution of the high frequency components.

The high frequency fluctuations, F_6, decreased in the Gulf Coast watersheds and the Pacific Northwestern watersheds. The reason for this was also the increase in winter ET and decrease in winter rainfall depth in these watersheds. In these watersheds, the dominant rainfall season is winter; therefore, an increase in ET possibly resulted in decrease in antecedent soil moisture and, overall, muted response of rainfall to streamflow. There was a difference in the Pacific Northwest and Gulf Coast watersheds: the values of F_4 increased in majority of the Pacific Northwest region while it decreased in the latter.

The trends in the contribution of fluctuations at different timescales were also related to soil properties such as soil texture, porosity, and fraction of forest. Further analyses revealed that soil properties were an indicator of change in climatic statistics. In snow dominated watersheds with fine soils, high rainfall depth increased, and winter maximum daily temperature increased only moderately. This is hypothesized to have resulted in an increase in F_6 in these watersheds. In the snow dominated watershed with sandy soils, decrease or only a moderate increase in high rainfall depth with large increase in winter maximum daily temperature is hypothesized to result in significant decrease in soil moistures and decrease in F_6.

In the rain dominated watersheds with sandy soil F_6 decreased. Most of the watersheds with sandy soils are in humid region with high mean annual rainfall. Another difference between watersheds with sandy and fine soils was that in the former the phase difference between monthly rainfall and evaporation decreased which might have resulted in more rainwater evaporating back to atmosphere, drying of soils, and muted response of watersheds to rainstorms.

In snow dominated watersheds change in temperature-snowmelt relationship is responsible at least to some extent for streamflow regime change. The change in temperature-snowmelt relationship is likely due to change in spatiotemporal snow statistics and temperature statistics rather than any physical changes in the watersheds. Although, change in vegetation density might also be responsible for the changes. In rain dominated watersheds, the change in rainfall-runoff relationship appears to be negligible.

We note that conclusions reported in this study apply only to deseasonalized streamflow time-series. Changes in seasonal components are not studied in this paper. Nevertheless, the results presented in this study convincingly show that changes in streamflow regime have occurred across USA. Although the pattern of changes is patchy, there is substantial spatial structure. These changes have consequences for accurate simulation of streamflow time-series in the presence of climate change. Decreasing influence of low frequency components can result in decrease in accuracy of simulations. This is evident in arid watersheds of the Great Plains where the contribution of low frequency components has always been small, and all the models (conceptual, process-based, and ML models) of streamflow have been reported to perform poorly in these watersheds (e.g., Konapala et al., 2020).

In this study, only the effect of climatic statistics change on streamflow regime change has been explored. But streamflow regime can also change due to change in natural changes in land-use such as due to forest disturbance (e.g., Goeking & Tarboton, 2022). The effects of such changes on streamflow statistical structure should be the topic of future study. Moreover, we believe that it would be worthwhile to simulate the hydrologic response of CAMELS watersheds using a detailed process-based model to understand the changes in various hydrologic quantities in these watersheds.

Finally, the analysis carried out in this study identifies only the variables that play a role in determining the changes in streamflow regime. The specific mechanisms creating the changes could not be identified using this analysis. Nevertheless, a few hypotheses regarding changes in the hydrologic mechanisms that might have led to streamflow regime change have been proposed. Data between water years 1980-2013 was used to achieve the objectives. Though 30-35 years of data are not enough to identify all the changes in streamflow regime due to climate change because natural climate oscillation occurs at 30-year timescale, such data can still reveal useful pattern of hydrologic change (e.g., Ficklin et al., 2016). Besides, it is well known that systematic changes in global temperatures and rainfall patterns have occurred over the study period (Manabe & Broccoli, 2020). Therefore, we believe that it is prudent to look for streamflow regime changes across the USA due to climate change over the period used in this study.

Table 2. A summary of streamflow statistical structure and change in streamflow statistical structure in different regions of USA

Geographic region	**Climate**	**Streamflow statistical structure**	**Change in streamflow statistical structure**	**Cause of change**
Pacific Northwest	Humid	High values of F_3, F_5, F_6, low values of F_0	Decrease in F_3 and F_6, increase in F_4 in some of the watersheds	Increase in winter temperature and decrease in winter rainfall depth, resulting in decrease in the strength of interflow seems to be the main cause. Winter is the high rainfall season in these watersheds.
Gulf Coast	Humid	High values of F_6, moderate to high value of F_3 and F_5	Decrease in F_6, F_4, mixed response of change in F_3; Increase in low frequency components F_0, F_1, and F_2	Decrease in winter temperature and decrease in winter rainfall depth, resulting in muted response of these watersheds to rainfall seems to be the main cause. Winter is the high rainfall season in these watersheds.
Great Plains	Arid	Very high values of F_6. Low to	Mixed trends, but majority of the watersheds	Increase in OND temperatures, resulting in increase in ET and

		moderate values of F_0, F_3, and F_5	had increase in high frequency components and decrease in low frequency components	decrease in low frequency components. Spring-summer is the main rainfall season in these watersheds.
Atlantic Coast and eastern most Great Lakes region	Humid	Low value of F_0, high values of F_5 and F_6, low to high values of F_3.	Increase in F_4 and F_6, decrease in F_3 and F_5	Increase in precipitation
Western Rocky Mountains	Arid	Moderate to high values of F_0, high values of F_5, low values of other components	Decrease in F_0, increase in F_4 and F_6; F_1 and F_2 had both positive and negative trends	Increase in temperature, change in rainfall patterns, and decrease in SWE.
Eastern Rocky Mountains	Arid	Moderate to high values of F_0, high values of F_5, low values of other components	Mixed trends, F_1 increased in most of the watersheds; F_0 decreased in some and increased in other watersheds	Increase in temperature, change in rainfall patterns, and decrease in SWE. The cause of differences between eastern and western Rocky Mountains is unclear.

F_0 = Fraction of variance contributed by greater 1-year timescale components; F_1 = Fraction of variance contributed by 4-months to 1-year timescale components; F_2 = Fraction of variance contributed by 1-month to 4-months timescale components; F_3 = Fraction of variance contributed by 2-weeks to 1-month timescale components; F_4 = Fraction of variance contributed by less than 2-weeks timescale components;
$F_5 = F_1 + F_2$; $F_6 = F_3 + F_4$

Appendix:

Table A1. Variables used in the study to interpret the streamflow regime changes

Property	Variables	Remarks
Rainfall	Mean rainfall, rainfall seasonality (see Addor et al., 2017), high rainfall frequency, high rainfall duration, low rainfall duration, trend in mean rainfall depth, trend in total rainfall depth, trend in number of rainstorms, trend in number of rain days, trend in high rainfall frequency, trend in high rainfall duration, trend in high rainfall depth, trend in low rainfall frequency, trend in low rainfall duration, trend in low rainfall depth, trend in OND (Oct Nov-Dec) rainfall depth, trend in JFM (Jan-Feb-Mar) rainfall depth, trend in AMJ (Apr-May-Jun) rainfall depth, trend in JAS (Jul-Aug-Sep) rainfall depth	
Temperature	Mean temperature, trend in mean minimum daily temperature, trend in mean maximum daily temperature, trend in median minimum daily temperature, trend in median minimum daily temperature, trend in median maximum daily temperature, trend in SD (standard deviation) maximum daily temperature, trend in SD minimum daily temperature, trend in OND minimum (maximum) daily temperature, trend in JFM minimum (maximum) daily temperature, trend in AMJ minimum (maximum) daily temperature, trend in JAS minimum (maximum) daily temperature, trend in mean minimum (maximum) daily temperature 0-10 percentiles, trend in mean minimum (maximum) daily temperature 10-30 percentiles, trend in mean minimum (maximum) daily temperature 30-60 percentiles, trend in mean minimum (maximum) daily temperature 60-90 percentiles, trend in mean minimum (maximum) daily temperature 90-100 percentiles,	
Snow statistics	Fraction of snow, trend in snow water equivalent (SWE)	For snow dominated watersheds
Geomorphological characteristics	Mean elevation, mean slope, drainage area	
Climate indices except precipitation	Potential evapotranspiration (PET), aridity, runoff	
Monthly climate statistics	Temperature amplitude (ΔT), mean normalized rainfall amplitude (δ_{P}), temperature phase (s_{T}), rainfall phase (s_{P}), phase difference between rainfall and temperature (s_{d})	Berghuijs and Woods, (2016)
Soil properties	Soil depth, depth to bedrock, soil conductivity, fraction of sand content, fraction of clay content, fraction of silt	Addor et al., (2017)

	content, fraction of organic content, water holding capacity, other fractions	
Land use	Fraction of forest	
Location	Latitude, Longitude	
Rainfall-runoff response	Trend in λ, CN, α/β, α/β^2 and mean of different percentiles on these quantities	Only for rain-driven watersheds (see SI)
Temperature streamflow relationship	Trend in rising limb slope, trend in rising limb intercept, trend in streamflow regime time-to-peak	Only for snow-dominated watersheds (see SI)

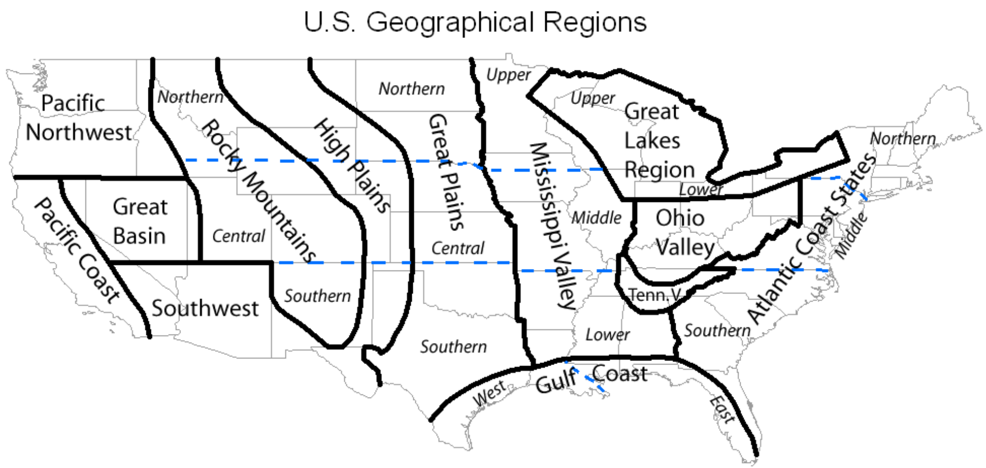

Figure A1. Map of the geographical regions referred to in this study (https://www.ncdc.noaa.gov/temp-and-precip/drought/nadm/geography)

Data Availability Statement:

All the data used in this study are publicly available with relevant references provided in the text.

Acknowledgements:

AG was supported by Maki Postdoctoral Fellowship at DRI to carry out this work. Authors acknowledge Chris Pearson, and Patrick Sawyer for providing feedback on this work. Authors thank Jaideep Ray for suggesting some of the methodology implemented in the paper and providing feedback on a draft of this paper.

References:

Addor, N., Newman, A. J., Mizukami, N., & Clark, M. P. (2017). The CAMELS data set: catchment attributes and meteorology for large-sample studies. *Hydrology and Earth System Sciences*, 21(10), 5293-5313.

Addor, N., Newman, A., Mizukami, M., & Clark, M. P. (2017). Catchment attributes for large-sample studies. Boulder, CO: UCAR/NCAR. https://doi.org/10.5065/D6G73C3Q

Belmecheri, S., Babst, F., Wahl, E. R., Stahle, D. W., & Trouet, V. (2016). Multi-century evaluation of Sierra Nevada snowpack. *Nature Climate Change*, 6(1), 2-3.

Berghuijs, W. R., & Woods, R. A. (2016). A simple framework to quantitatively describe monthly precipitation and temperature climatology. *International Journal of Climatology*, 36(9), 3161-3174.

Betterle, A., Schirmer, M., & Botter, G. (2019). Flow dynamics at the continental scale: Streamflow correlation and hydrological similarity. *Hydrological processes*, 33(4), 627-646.

Beven, K. J. (2011). Rainfall-runoff modelling: the primer. John Wiley and Sons.

Botter, G., Basso, S., Rodriguez-Iturbe, I., & Rinaldo, A. (2013). Resilience of river flow regimes. *Proceedings of the National Academy of Sciences*, 110(32), 12925-12930.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: forecasting and control. John Wiley and Sons.

Bras, R. L., & Rodriguez-Iturbe, I. (1993). Random functions and hydrology. Courier Corporation.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

Chow, V. T. (1978). Stochastic modeling of watershed systems [French Broad River Basin, North Carolina as an example]. *Advances in Hydroscience*.

Collischonn, W., & Fan, F. M. (2013). Defining parameters for Eckhardt's digital baseflow filter. *Hydrological Processes*, 27(18), 2614-2622.

Donohue, R. J., Roderick, M. L., McVicar, T. R., & Farquhar, G. D. (2013). Impact of CO2 fertilization on maximum foliage cover across the globe's warm, arid environments. *Geophysical Research Letters*, 40(12), 3031-3035.

Eagleson, P. S. (1982). Ecological optimality in water-limited natural soil-vegetation systems: 1. Theory and hypothesis. *Water Resources Research*, 18(2), 325-340.

Ed Dlugokencky & Pieter Tans, NOAA/GML (gml.noaa.gov/ccgg/trends/), date accessed: 17 Mar 2022.

Ficklin, D. L., Robeson, S. M., & Knouft, J. H. (2016). Impacts of recent climate change on trends in baseflow and stormflow in United States watersheds. *Geophysical Research Letters*, 43(10), 5079-5088.

Geetha, K., Mishra, S. K., Eldho, T. I., Rastogi, A. K., & Pandey, R. P. (2007). Modifications to SCS-CN method for long-term hydrologic simulation. *Journal of Irrigation and Drainage Engineering*, 133(5), 475-486.

Goeking, S. A., & Tarboton, D. G. (2021). Variable streamflow response to forest disturbance in the western US: A large-sample hydrology approach. *Water Resources Research*, e2021WR031575.

Gordon, B. L., Brooks, P. D., Krogh, S. A., Boisrame, G. F., Carroll, R. W., McNamara, J. P., & Harpold, A. A. (2022). Why does snowmelt-driven streamflow response to warming vary? A data-driven review and predictive framework. *Environmental Research Letters*.

Granger, C. W. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of econometrics*, *14*(2), 227-238.

Granger, C. W., & Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1(1), 15-29.

Gudmundsson, L., Tallaksen, L. M., Stahl, K., & Fleig, A. K. (2011). Low-frequency variability of European runoff. *Hydrology and Earth System Sciences*, 15(9), 2853-2869.

Hirpa, F. A., Gebremichael, M., & Over, T. M. (2010). River flow fluctuation analysis: Effect of watershed area. *Water Resources Research*, 46(12).

Horner, I., Branger, F., McMillan, H., Vannier, O., & Braud, I. (2020). Information content of snow hydrological signatures based on streamflow, precipitation and air temperature. *Hydrological Processes*, 34(12), 2763-2779.

Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, 116(1), 770-799.

Kim, D. H., Rao, P. S. C., Kim, D., & Park, J. (2016). 1/f noise analyses of urbanization effects on streamflow characteristics. *Hydrological Processes*, 30(11), 1651-1664.

Klemeš, V. (1978). Physically based stochastic hydrologic analysis. *In Advances in hydroscience* (Vol. 11, pp. 285-356). Elsevier.

Klemeš, V. (1986). Operational testing of hydrological simulation models. Hydrological sciences journal, 31(1), 13-24.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall–runoff modelling using long short-term memory (LSTM) networks. *Hydrology and Earth System Sciences*, 22(11), 6005-6022.

Laio, F., Porporato, A., Ridolfi, L., & Rodriguez-Iturbe, I. (2001). Plants in water-controlled ecosystems: active role in hydrologic processes and response to water stress: II. Probabilistic soil moisture dynamics. *Advances in Water Resources*, 24(7), 707-723.

Lamb, R., & Beven, K. (1997). Using interactive recession curve analysis to specify a general catchment storage model. *Hydrology and Earth System Sciences*, 1(1), 101-113.

Lee, H. T., & Delleur, J. W. (1972). A program for estimating runoff from indiana watersheds, part iii: analysis of geomorphologic data and a dynamic contributing area model for runoff estimation. https://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1025&context=watertech

Manabe, S., & Broccoli, A. J. (2020). *Beyond global warming: How numerical models revealed the secrets of climate change*. Princeton University Press.

Milly, P. C. D. (1997). Sensitivity of greenhouse summer dryness to changes in plant rooting characteristics. *Geophysical Research Letters*, 24(3), 269-271.

Milly, P. C., & Dunne, K. A. (2016). Potential evapotranspiration and continental drying. *Nature Climate Change*, 6(10), 946-949.

Milly, P. C., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., & Stouffer, R. J. (2008). Stationarity is dead: whither water management?. *Science*, 319(5863), 573-574.

Milly, P. C., Dunne, K. A., & Vecchia, A. V. (2005). Global pattern of trends in streamflow and water availability in a changing climate. *Nature*, 438(7066), 347-350.

Mishra, S. K., & Singh, V. P. (1999). Another look at SCS-CN method. *Journal of Hydrologic Engineering*, 4(3), 257-264.

Montanari, A., Rosso, R., & Taqqu, M. S. (1997). Fractionally differenced ARIMA models applied to hydrologic time series: Identification, estimation, and simulation. *Water Resources Research*, 33(5), 1035-1044.

Montanari, A., Rosso, R., & Taqqu, M. S. (2000). A seasonal fractional ARIMA model applied to the Nile River monthly flows at Aswan. *Water Resources Research*, 36(5), 1249-1259.

Mote, P. W. (2006). Climate-driven variability and trends in mountain snowpack in western North America. *Journal of Climate*, 19(23), 6209-6220.

Mote, P. W., Li, S., Lettenmaier, D. P., Xiao, M., & Engel, R. (2018). Dramatic declines in snowpack in the western US. *Npj Climate and Atmospheric Science*, 1(1), 1-6.

Mudelsee, M. (2007). Long memory of rivers from spatial aggregation. *Water Resources Research*, *43*(1).

Ponce, V. M., & Hawkins, R. H. (1996). Runoff curve number: Has it reached maturity?. *Journal of Hydrologic Engineering*, 1(1), 11-19.

Porporato, A., Laio, F., Ridolfi, L., & Rodriguez-Iturbe, I. (2001). Plants in water-controlled ecosystems: active role in hydrologic processes and response to water stress: III. Vegetation water stress. *Advances in Water Resources*, 24(7), 725-744.

Priestley, M. B. (1982). *Spectral analysis and time series: probability and mathematical statistics* (No. 04; QA280, P7.).

Rodriguez-Iturbe, I., Porporato, A., Laio, F., & Ridolfi, L. (2001). Plants in water-controlled ecosystems: active role in hydrologic processes and response to water stress: I. Scope and general outline. *Advances in Water Resources*, 24(7), 695-705.

Rodriguez-Iturbe, I., Porporato, A., Ridolfi, L., Isham, V., & Coxi, D. R. (1999). Probabilistic modelling of water balance at a point: the role of climate, soil and vegetation. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 455(1990), 3789-3805.

Singh, R., Wagener, T., Van Werkhoven, K., Mann, M. E., & Crane, R. (2011). A trading-space-for-time approach to probabilistic continuous streamflow predictions in a changing climate–accounting for changing watershed behavior. *Hydrology and Earth System Sciences*, 15(11), 3591-3603.

Sivapalan, M., Yaeger, M. A., Harman, C. J., Xu, X., & Troch, P. A. (2011). Functional model of water balance variability at the catchment scale: 1. Evidence of hydrologic similarity and space-time symmetry. *Water Resources Research*, 47(2).

Soulis, K. X., & Valiantzas, J. D. (2012). SCS-CN parameter determination using rainfall-runoff data in heterogeneous watersheds–the two-CN system approach. *Hydrology and Earth System Sciences*, 16(3), 1001-1015.

Soulis, K. X., & Valiantzas, J. D. (2013). Identification of the SCS-CN parameter spatial distribution using rainfall-runoff data in heterogeneous watersheds. *Water Resources Management*, 27(6), 1737-1749.

Stephens, C. M., Marshall, L. A., Johnson, F. M., Lin, L., Band, L. E., and Ajami, H. (2020). Is past variability a suitable proxy for future change? A virtual catchment experiment. *Water Resources Research*, 56(2), e2019WR026275.

Tessier, Y., Lovejoy, S., Hubert, P., Schertzer, D., & Pecknold, S. (1996). Multifractal analysis and modeling of rainfall and river flows and scaling, causal transfer functions. *Journal of Geophysical Research: Atmospheres*, 101(D21), 26427-26440.

Wu, S., Zhao, J., Wang, H., & Sivapalan, M. (2021). Regional patterns and physical controls of streamflow generation across the conterminous United States. *Water Resources Research*, *57*(6), e2020WR028086.

References from Supporting Information

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74(368), 829-836.

Montanari, A., Rosso, R., & Taqqu, M. S. (1997). Fractionally differenced ARIMA models applied to hydrologic time series: Identification, estimation, and simulation. *Water Resources Research*, 33(5), 1035-1044.

Seabold, S., & Perktold, J. (2010, June). Statsmodels: Econometric and statistical modeling with python. In Proceedings of the 9th Python in Science Conference (Vol. 57, p. 61).

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, in Petrov, B. N.; Csáki, F. (eds.), 2nd International Symposium on Information Theory, Tsahkadsor, Armenia, USSR, September 2-8, 1971, Budapest: Akadémiai Kiadó, pp. 267–281. Republished in Kotz, S.; Johnson, N. L., eds. (1992), Breakthroughs in Statistics, vol. I, Springer-Verlag, pp. 610–624.

Beran, J. (1994). *Statistics for long-memory processes.* Routledge.

Mote, P. W., Hamlet, A. F., Clark, M. P., & Lettenmaier, D. P. (2005). Declining mountain snowpack in western North America. *Bulletin of the American Meteorological Society*, 86(1), 39-50.

Mote, P. W. (2006). Climate-driven variability and trends in mountain snowpack in western North America. *Journal of Climate*, 19(23), 6209-6220.

Knowles, N., Dettinger, M. D., & Cayan, D. R. (2006). Trends in snowfall versus rainfall in the western United States. *Journal of Climate*, 19(18), 4545-4559.

Belmecheri, S., Babst, F., Wahl, E. R., Stahle, D. W., & Trouet, V. (2016). Multi-century evaluation of Sierra Nevada snowpack. *Nature Climate Change*, 6(1), 2-3.

Berg, N., & Hall, A. (2017). Anthropogenic warming impacts on California snowpack during drought. *Geophysical Research Letters*, 44(5), 2511-2518.

Collischonn, W., & Fan, F. M. (2013). Defining parameters for Eckhardt's digital baseflow filter. *Hydrological Processes*, 27(18), 2614-2622.

Lamb, R., & Beven, K. (1997). Using interactive recession curve analysis to specify a general catchment storage model. *Hydrology and Earth System Sciences*, 1(1), 101-113.

Ponce, V. M., & Hawkins, R. H. (1996). Runoff curve number: Has it reached maturity?. *Journal of Hydrologic Engineering*, 1(1), 11-19.

Mishra, S. K., & Singh, V. P. (1999). Another look at SCS-CN method. *Journal of Hydrologic Engineering*, 4(3), 257-264.

Geetha, K., Mishra, S. K., Eldho, T. I., Rastogi, A. K., & Pandey, R. P. (2007). Modifications to SCS-CN method for long-term hydrologic simulation. *Journal of Irrigation and Drainage Engineering*, 133(5), 475-486.

Soulis, K. X., & Valiantzas, J. D. (2012). SCS-CN parameter determination using rainfall-runoff data in heterogeneous watersheds–the two-CN system approach. *Hydrology and Earth System Sciences*, 16(3), 1001-1015.

Soulis, K. X., & Valiantzas, J. D. (2013). Identification of the SCS-CN parameter spatial distribution using rainfall-runoff data in heterogeneous watersheds. *Water Resources Management*, 27(6), 1737-1749.

Brutsaert, W. (2005). *Hydrology: an introduction*. Cambridge University Press.

Tolson, B. A., & Shoemaker, C. A. (2007). Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resources Research*, 43(1).