

Climate Model Assessment of Changes in Winter–Spring Streamflow Timing over North America

JONGHUN KAM

*Department of Civil, Construction, and Environmental Engineering, University of Alabama,
Tuscaloosa, Alabama*

THOMAS R. KNUTSON

NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey

P. C. D. MILLY

U.S. Geological Survey, and NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey

(Manuscript received 30 November 2017, in final form 6 April 2018)

ABSTRACT

Over regions where snowmelt runoff substantially contributes to winter–spring streamflows, warming can accelerate snowmelt and reduce dry-season streamflows. However, conclusive detection of changes and attribution to anthropogenic forcing is hindered by the brevity of observational records, model uncertainty, and uncertainty concerning internal variability. In this study, the detection/attribution of changes in midlatitude North American winter–spring streamflow timing is examined using nine global climate models under multiple forcing scenarios. Robustness across models, start/end dates for trends, and assumptions about internal variability are evaluated. Marginal evidence for an emerging detectable anthropogenic influence (according to four or five of nine models) is found in the north-central United States, where winter–spring streamflows have been starting earlier. Weaker indications of detectable anthropogenic influence (three of nine models) are found in the mountainous western United States/southwestern Canada and in the extreme northeastern United States/Canadian Maritimes. In the former region, a recent shift toward later streamflows has rendered the full-record trend toward earlier streamflows only marginally significant, with possible implications for previously published climate change detection findings for streamflow timing in this region. In the latter region, no forced model shows as large a shift toward earlier streamflow timing as the detectable observed shift. In other (including warm, snow free) regions, observed trends are typically not detectable, although in the U.S. central plains we find detectable delays in streamflow, which are inconsistent with forced model experiments.

1. Introduction

Warming climate can change the hydrologic cycle across temporal and spatial scales (Milly et al. 2008). Over snow-affected regions, warming can result in changes in the magnitude and timing of peak streamflow during the melt season and of low streamflow during the dry season (Kang et al. 2016). Seasonal changes can adversely affect human activities and ecological communities.

For example, winter ice jams increase flood risk (Goulding et al. 2009), a shortage of available water resources during dry seasons increases drought risk (Hodgkins and Dudley 2006), and a shift of the onset of spring (Cayan et al. 2001), or changes in peak streamflow magnitude and timing, can decrease fish survival rates (Hare et al. 2016; Crecco and Savoy 1985) during the spawning season and early life stages.

The sign of the observed changes in timing of winter–spring streamflow over some snow-affected regions of North America is consistent with the sign of the expected hydrologic response to regional warming. Previous climate model studies conclude that some of these regional changes are attributable to human-induced climate

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JCLI-D-17-0813.s1>.

Corresponding author: Dr. Jonghun Kam, jkam@eng.ua.edu

DOI: 10.1175/JCLI-D-17-0813.1

© 2018 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

TABLE 1. List of climate model members used in this study. The grid sizes for the relatively high-resolution climate models (CNRM-CM5 and CSIRO-Mk3.6.0) are highlighted in bold. (Expansions of acronyms are available online at <http://www.ametsoc.org/PubsAcronymList>.)

Name	CMIP5 (9 GCMs)				
	Ensemble number (HIST-ALL and RCP8.5)	Atmospheric grid		Years (CONT)	Ensemble size (HIST-NAT)
		Lat (°)	Lon (°)		
BCC-CSM1.1	3	2.79	2.8125	500	1
CanESM2	5	2.79	2.8125	896	5
CNRM-CM5	10	1.4	1.41	850	6
CSIRO-Mk3.6.0	9	1.8653	1.875	500	5
GISS-E2-R	14	2	2.5	1200	5
GISS-E2-H	8	2	2.5	1841	5
IPSL-CM5A-LR	6	1.895	3.75	1000	3
IPSL-CM5A-MR	3	1.2676	2.5	300	3

change (Hidalgo et al. 2009; Barnett et al. 2008) and that certain seasonal timing benchmarks will arrive a month earlier by the end of the twenty-first century (Stewart et al. 2004). It seems prudent to revisit such detection and attribution conclusions, as more data accumulate, and as additional multimodel and multiforcing climate model simulations become available. Our analysis considers datasets updated through 2012 in many cases, examines a wide selection of North American regions, and has the benefit of the data from phase 5 of the Coupled Model Intercomparison Project (CMIP5) (Taylor et al. 2012). Thus, in comparison to previous studies, we now have longer records available, a larger set of climate model simulations, and in many cases new models using updated physical process treatments. The responses of a global climate model to changes in atmospheric composition due to anthropogenic influence are sensitive to a number of physical processes in the models, such as treatments of clouds and cloud processes, ocean model resolution, and so on, making multimodel tests of robustness important.

To explore the possible effects of anthropogenic climate change on streamflow timing, we assess an index of discharge-weighted streamflow timing, namely the calendar date on which half of the total mass of streamflow measured at a series of gauges (for the January–June season) has been accumulated. This winter–spring center time (WSCT) index of streamflow timing is constructed for simulated monthly runoff from preindustrial control runs (CONT), natural-forcing-only runs (HIST-NAT), and historical forcing runs (HIST-ALL) from the CMIP5 archive (see Table 1). The HIST-ALL runs cover the period 1850–2005 and are forced by time-varying forcings, including anthropogenic changes in atmospheric composition (e.g., greenhouse gas concentrations and aerosol loadings) and land use, as well as natural changes in volcanic influence and incoming solar radiation. The HIST-NAT simulations use only the

subset of natural forcings from the HIST-ALL runs, namely volcanic influences and changes in solar insolation. As a result, the differences between the HIST-ALL and HIST-NAT runs are inferred to be due to anthropogenic influence.

Model bias and scale mismatch are potential issues in hydrologic studies using climate models (Hidalgo et al. 2009). This study focuses on the timing of winter–spring streamflow, rather than the simulated runoff magnitude itself, and so is not directly subject to the large biases (under- and/or overestimations) often found in the amount of precipitation or runoff from climate models. Rather, the timing of winter–spring streamflow is weighted by either streamflow or simulated runoff, and thus is normalized by the accumulated streamflow or simulated runoff from January through June of each year; an overall amplitude bias in streamflow/runoff will not affect WSCT. Mismatch of the horizontal scale is a well-known limitation in applying global climate models to regional hydroclimate studies. However, runoff production generally occurs at scales smaller than either those resolved by the model or those measured by gauges. Furthermore, a benefit in directly using the (nondownscaled) global climate model outputs for regional hydroclimate studies is the focus on physically consistent responses of the climate models to various climate forcings or to unforced internal variability from the control runs. Our analysis includes an evaluation of modeled WSCT climatology to confirm its relative insensitivity to runoff bias and horizontal scale.

The data and methods used in this study are described in section 2. Section 3 presents the results of comparison of CMIP5 WSCT climatology and standard deviations with observations, tests of detection, attribution, and consistency of WSCT trends. In section 4, we discuss uncertainties in our analysis, summarize the findings of this study, and highlight the implications of our findings under future potential climate change.

2. Data and methods

We divide the middle latitudes of North America into 8 regions and compute the regional averages of WSCT using observed monthly streamflow data retrieved from 61 streamflow gauging stations, classified as minimally affected by human activity in the basin, of the U.S. Geological Survey and the Water Survey of Canada over 1933–2012 (Fig. 1a). Regions 1 and 2 cover the western U.S. mountainous and coastal regions, respectively, with the division line based primarily on the climatology of WSCT. East of regions 1 and 2, the regions above 44°N (regions 3 and 6) represent snowmelt-dominant regions for winter–spring streamflow, while the regions between 41° and 44°N (regions 4 and 7) represent mixed influence regions (snowmelt and rainfall) for winter–spring streamflow (Hodgkins and Dudley 2006). Winter–spring streamflows over regions 5 and 8 are dominated by rainfall and thus serve as control regions for the colder regions in our study. These snowmelt-dominant, mixed, and rainfall-dominant regions are consistent with the long-term averaged snow cover areas (1972–2010) in Fig. 1a. We assess the adequacy of the sparse network by trend comparisons (Fig. 1c) with a denser network of stations that began only in 1951 (Figs. 1b,c; 263 stations).

a. Streamflow data

Streamflow from 246 (1951–2000) and 57 (1933–2012) U.S. gauging stations are selected from the Hydro-Climatic Data Network (HCDN) stations; each station used is classified as minimally influenced by artificial diversions, storage, or human activities (Slack and Landwehr 1992). Over Canada, 17 and 4 streamflow gauging stations with no reported regulation are selected from the Hydrometric Database (HYDAT) for the periods 1951–2000 and 1933–2012, respectively. In this study, we select HCDN and HYDAT stations with continuous records over 1933–2012.

b. Simulated runoff from the CMIP5 archive

For the HIST-ALL, HIST-NAT, and CONT runs, we use nine CMIP5 models (CMIP5 home institutions and number of ensemble members for each model/forcing run are listed in Table 1). Before computing regional averages, we regrid all model-simulated runoff data to a $2.5^\circ \times 2.5^\circ$ grid, using the function “area_conserve_remap” from the NCAR Command Language program. We then compute regional averages of simulated runoff for the $2.5^\circ \times 2.5^\circ$ grid cells where observational data are available. For the observations, we weight all stations in a region equally.

The observations are of streamflow, which generally is not available from the models. Therefore, we use model

runoff (monthly means) as a surrogate for streamflow from small basins. The difference between runoff and streamflow is a time lag representing the time of travel through the river basin (Mengelkamp et al. 2001). For the small basins considered here (from 22 to 32 400 km²), assuming a stream velocity on the order of 1 m s⁻¹ and a travel distance on the order of the square root of basin area, this time lag is on the order of from 1 h to 2 days. Any changes in this time lag due to changes in stream velocity will likely be much smaller, and therefore small relative to the changes in WSCT being examined, which are on the order of a few days (Fig. 3, below).

Most CMIP5 groups provide an ensemble of HIST-ALL and HIST-NAT simulations for a given model, with identical climate forcings but starting from different points in the CONT runs. The ensemble average across multiple ensemble members for a given model/experiment provides a more reliable estimate of that model’s response to external forcing than a single ensemble member, because the ensemble averaging helps to average out the influence of internal variability that causes differences between individual ensemble members. Using only a single ensemble member is generally not appropriate, as the externally forced signal can be obscured by internal variability. For the WSCT climatology and interannual standard deviation analyses, we use the HIST-ALL runs from nine CMIP5 models. For detection and attribution assessment, we use paired HIST-NAT/HIST-ALL runs from these same nine CMIP5 models since only these nine models have HIST-NAT runs through 2012. The ensemble means of HIST-ALL forcing runs are extended to 2012, if necessary, using available RCP8.5 scenario ensemble members.

c. Computation for center time of winter–spring streamflow

The center time of winter–spring streamflow is the center of mass of streamflow (where the centering is with respect to time; Stewart et al. 2004), which is an average of timing weighted by total streamflow/runoff. We compute the WSCT, expressed in terms of the day of year (DOY) using monthly mean streamflow q [Eq. (1)]:

$$\text{WSCT (DOY)} = \frac{\sum_{i=1}^6 (t_i - 0.5) q_i}{\sum_{i=1}^6 q_i} \times \left(\frac{181 \text{ days}}{6 \text{ months}} \right), \quad (1)$$

in which timing t_i is given in months. To centralize the time within a month, we use 0.5 and 5.5 ($t_i - 0.5$) to represent the midpoints of the various months from January ($i = 1$) through June ($i = 6$). In this study, only

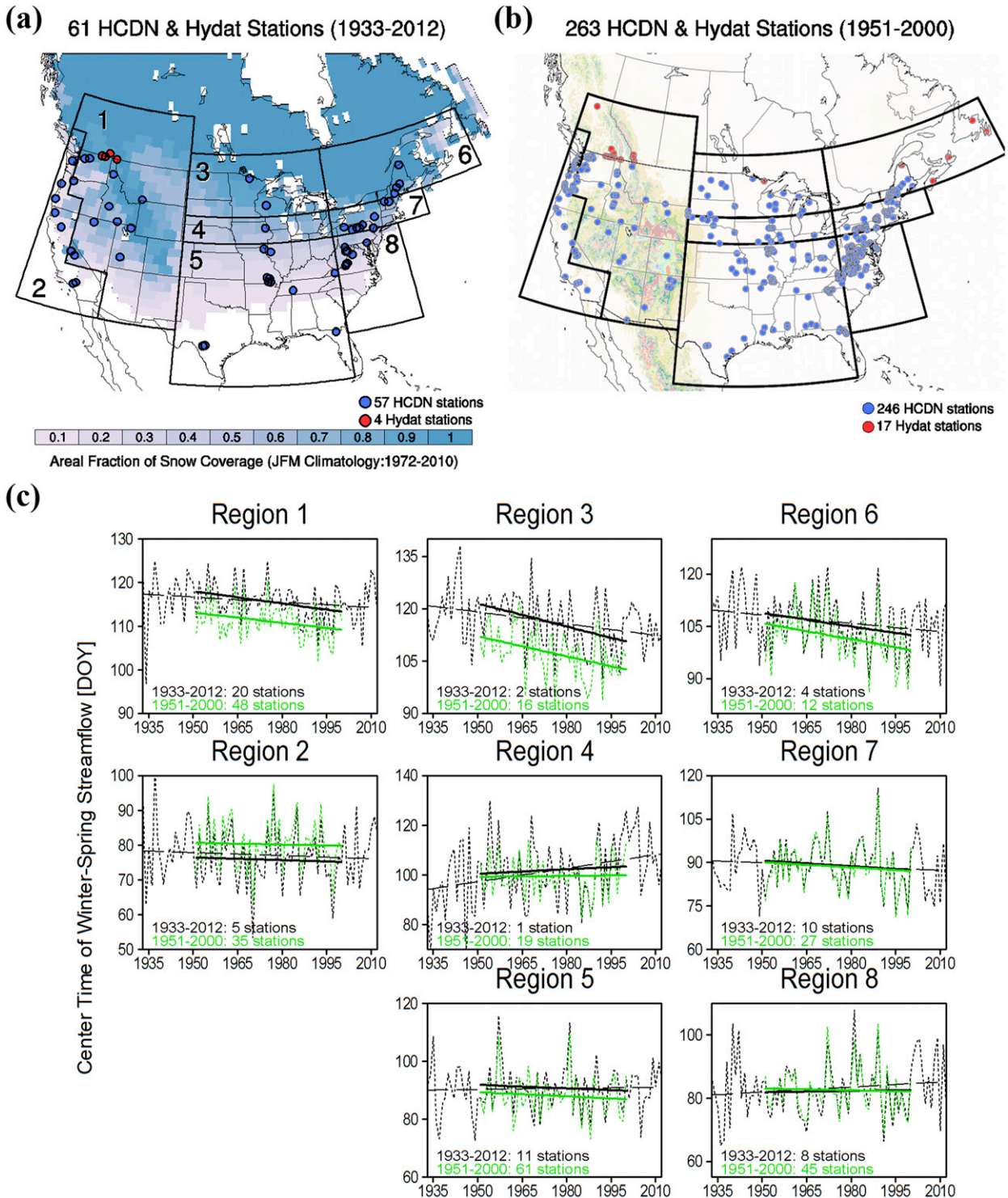


FIG. 1. Station/region locations and regional time series of WSCT. Spatial distributions of the USGS HCDN (blue dot) and Canadian Hydat (red dot) gauging stations for (a) 1933–2012 and (b) 1951–2000. Elevations above 2000 m are color-shaded in (b). The climatological (January–March) fractional area covered by snow for each grid cell is shown in (a) based on 1972–2010 from the National Sea Ice Index Data Center weekly snow cover and sea ice extent data (<http://nsidc.org/data/nsidc-0046>). (c) Annual time series of WSCT over 1933–2012 (black short-dashed line) and 1951–2000 (green short-dashed line). In (c), black solid and long-dashed lines depict linear trends for the full period of record (1933–2012) and the shorter period (1951–2000), respectively, for a subset of stations; the green lines depict linear trends for the period 1951–2000 for a total of 263 stations. The number of stations in each region is listed in the bottom-left corner of each time series plot.

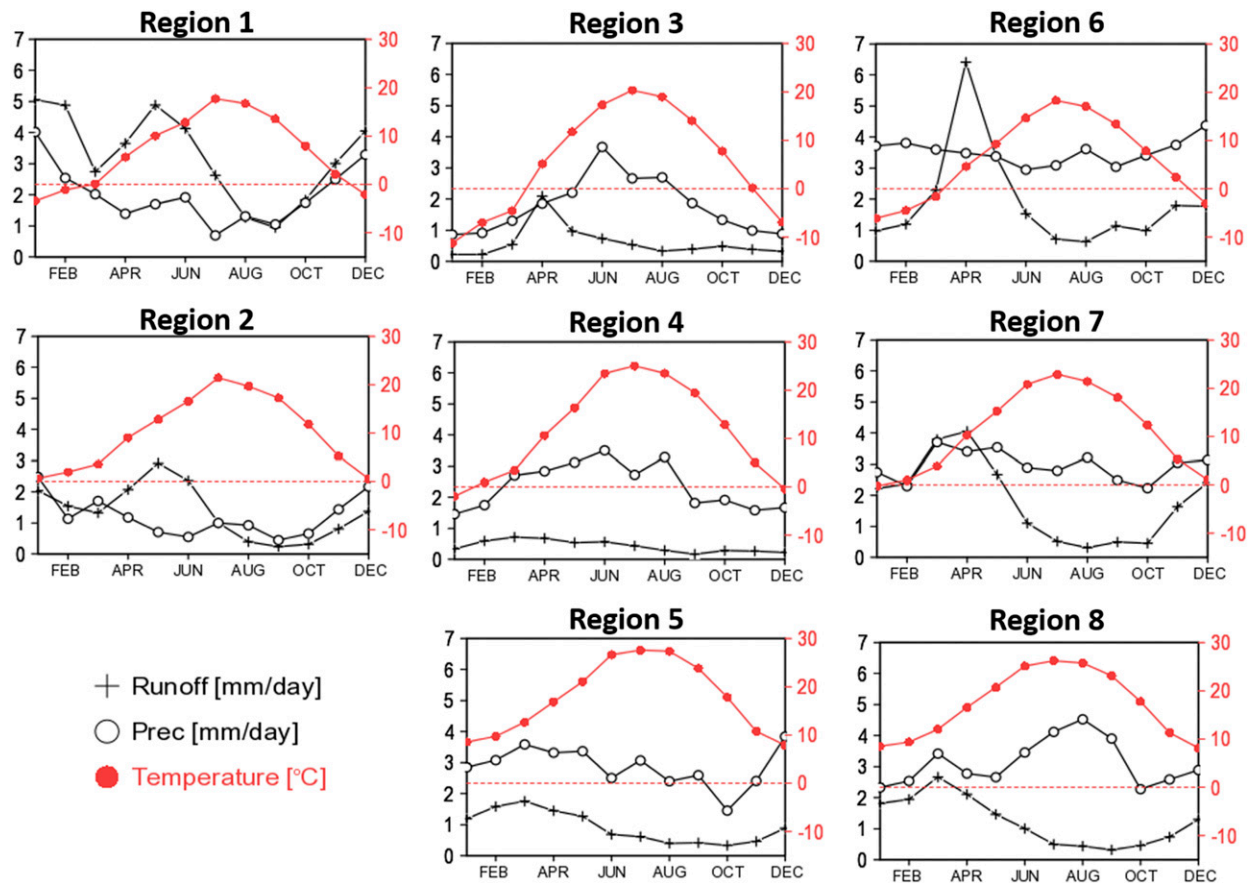


FIG. 2. Seasonality of monthly runoff, precipitation, and temperature averaged over the period 1951–2000. Monthly runoff (plus symbols) is computed by dividing the monthly streamflows at gauges by the corresponding drainage area. Precipitation (black open circles) and temperature (red solid circles) are retrieved from the Climatic Research Unit (CRU) time series (CRU TSv.3.23) data [0.5° (50 km) spatial resolution; <https://crudata.uea.ac.uk/cru/data/hrg/>] and regridded at 2.5° spatial resolution using the function “area_conserve_remap” from the NCAR Command Language program.

the first half of the calendar year (January–June) is used to compute the center of time in order to examine hydrologic changes during snowmelt season and to reduce the influence of rainfall events in early winter (November and December) and midsummer (July and August). Over snow-dominant regions (regions 1, 3, and 6), months when the monthly averaged temperature rises above 0°C occur around the center of the window (March or April; see Fig. 2).

To assess the suitability of modeled monthly runoff for our analysis, we compute the WSCT from both daily and monthly streamflow observations and show a scatterplot comparing the magnitudes of WSCT trends derived using daily versus monthly data, as estimated using the nonparametric Sen slope estimator (Sen 1968). The Sen slope for a given time series is computed from the median of the slopes associated with all possible pairings of points within the time series. Figure 3 shows excellent agreement between the

Sen slope estimates we derive from the daily and monthly data.

d. Construction of the distributions of WSCT from the CMIP5 archive

To estimate the distribution of WSCT trends arising from internal variability, we use the CONT runs from each of the nine CMIP5 models. We randomly sample (with replacement) 1000 segments of length 80 years each from each CONT run. Each of these 80-yr segments corresponds to our 1933–2012 analysis period under the null hypothesis of zero trends. For every 80-yr segment, we compute the nonparametric Sen slopes (Sen 1968) for each of the 1000 segments using only “windows” of the first 21 years of the segment, the first 22 years, and so on up to the full 80 years. These computations parallel those with the observed data in the first sliding trend analysis described below. The Sen slope for a given window is computed from the median

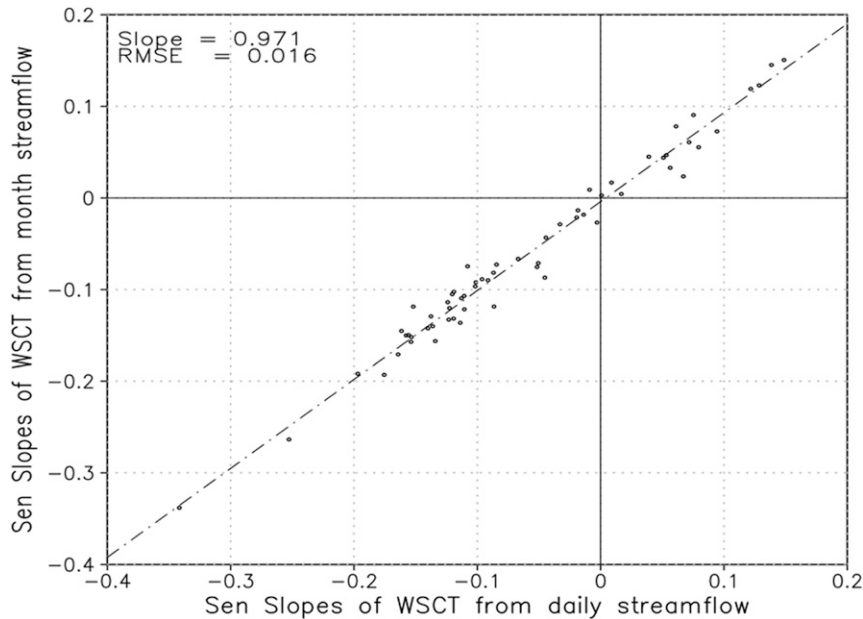


FIG. 3. Comparison of WSCT trends from daily and monthly streamflow records. A scatterplot of the Sen slopes of WSCT from daily (x axis) and monthly (y axis) streamflow records (1951–2000) over 61 HCDN and Hydat stations. The units are days per year.

of the slopes associated with all possible pairings of points within the window. From the distribution of the Sen slopes obtained for each window, we compute the corresponding 5th and 95th percentiles of the Sen slope values for each of the nine models.

e. Detectability, consistency, and attribution tests

Detectability, consistency, and attribution tests are done for each model separately using that model's 5th and 95th percentile ranges of the WSCT slopes. Assuming that internal variability of the climate system can be represented by the CONT runs, the distributions of WSCT trends for the HIST-ALL and HIST-NAT forcing runs are formed using the ensemble means of the HIST-ALL or HIST-NAT forcing runs together with the 5th–95th percentile ranges from the CONT runs to form a 5th–95th percentile range about the ensemble mean trends for each corresponding climate model. From this, we can determine whether the single observed trend is outside or inside the 5th–95th percentile range of the HIST-ALL and HIST-NAT forcing runs for each individual climate model. This leads to 12 possible categories of trend comparison, depending on the location of the observed trend value relative to the HIST-ALL and HIST-NAT distributions (see Fig. S1 in the online supplemental material). Our method assumes that the distributions of WSCT Sen slopes from the HIST-NAT and HIST-ALL runs are similar to those from the control runs except for a shift of the mean, with

no change in the shape of the distribution. Since the CMIP5 project provides only a limited number of ensemble members for the HIST-NAT and HIST-ALL runs of each individual model, it is difficult to assess whether any changes in variability are detectable between the HIST-NAT, HIST-ALL, and control runs. To detect significant changes in the variability and understand their causes, a much larger ensemble set of HIST-NAT and HIST-ALL runs would be required, and these were not generally available for the CMIP5 models.

f. Sliding trend analysis of WSCT

For our sliding trend analysis, we use a fixed start year (1933) and different end years (1953–2012) to examine the robustness of our findings. Through the sliding trend analyses, we can address such aspects as the robustness of detection, attribution, or consistency results to the use of different periods in the full records, or the record length necessary to distinguish observed changes from internal variability. We also conduct a sliding trend analysis using different start years from 1933 through 1992, but with a fixed end year (2012) (see Fig. S2), showing generally consistent detection and attribution results to those we obtain from the sliding trend analysis using different end years but with a fixed start year. Such sliding trend analyses have been used previously in climate change detection studies to assess the robustness of findings (e.g., Knutson et al. 2013) or time scale of emergence of detectable trends (Akhter et al. 2018). As

shown in Knutson et al. (2013), the 5th or 95th percentile range of control run trend magnitudes tends to be larger for shorter trend durations. This makes it more difficult for climate change detection to occur for shorter trend lengths. The important consideration is to use trends from the control run that are consistent in length with the observed trend being considered, which ensures that this effect is accounted for in the trend detection analysis.

g. Computation for decadal internal variability of WSCT

To assess the CMIP5 models' decadal internal variability versus observations, we use a decadal low-pass filter with the half-power point at nine years. We estimate the observed decadal internal variability by subtracting the grand ensemble mean of the HIST-ALL forcing runs from the observed low-pass filtered WSCT. The ensemble mean of HIST-ALL runs, which estimates the multimodel response to external forcing changes, is assumed also to be equal to the observed (OBS) response to these changes. The subtraction (OBS minus HIST-ALL) thus removes from observations an estimate of the expected response to the changes in external forcings (anthropogenic and natural), leaving as a residual our estimate of the internal variability component of Earth's climate system—that being the “observed” variability in a hypothetical world without external forcing changes. To assess the simulated low-frequency (decadal) internal variability, we compute the standard deviation of decadal WSCT from each climate model CONT run and compare this to the observed internal variability, with identical decadal low-pass filtering applied. This comparison between the estimated observed and simulated low-frequency internal variability metrics aids our assessment of how much confidence to place in multidecadal trend detection/attribution or consistency results from the CMIP5 models.

3. Results

a. Comparison of CMIP5 WSCT climatology and standard deviations with observations

Observational data show that the climatological values of WSCT for 1951–2000 range from calendar DOY 70 through 130 across North America (Fig. 4a). Typically, the WSCT along U.S. coastal regions occurs from middle to late March (DOY 70–90), over inland U.S. regions from early to mid-April (DOY 90–110), and in the Rocky Mountain region from late April to early May (DOY 110–130). The grand ensemble mean (the average of the ensemble means of the nine individual climate

models) of WSCT from the models (Fig. 4b) qualitatively captures the observed spatial pattern of climatological WSCT across North America. However, the CMIP5 models' WSCT in the mountainous and northern regions is notably earlier than in observations. Additionally, the CMIP5 climate models overestimate the standard deviation of WSCT in northern regions and over mountainous regions (Figs. 4c,d), whereas they underestimate WSCT variability in the middle of the conterminous United States. These biases in mean and variance of WSCT presumably arise from complex interactions of several factors: 1) biases in climate-model precipitation timing and its variance, 2) biases in the mean and variance of the climate-model delays between precipitation and runoff, 3) biases in grid-scale temperature, and 4) failure of climate models to represent the subgrid variation of elevation (Figs. 4e,f), and hence temperature, precipitation phase, and timing of snowmelt.

b. Initial tests of detection, attribution, and consistency of WSCT trends

We compare observed trends with trend distributions obtained from climate models. We use linear trends over various time periods as a metric to compare models and observations. We are not assuming that the forcing responses are linear trends, but the trend metric provides a convenient basis for comparisons of low-frequency temporal behavior. We evaluate the agreement across nine climate models by counting the number of climate models placed into each of four categories, using trends either for the full period of record (Fig. 5b) or for different end years with a fixed start year (1933) (Fig. 5c). The four categories (schematically illustrated in Fig. 5a) include 1) observed trend consistent with HIST-ALL trend, but not with HIST-NAT trend (A_n), 2) observed trend consistent with both HIST-ALL and HIST-NAT (A_N), 3) observed trend consistent with HIST-NAT but not with HIST-ALL (a_N), and 4) observed trend not consistent with either HIST-ALL or HIST-NAT (a_n). Detectable changes are inferred if observations are inconsistent with HIST-NAT (categories A_n or a_n). We infer a partially attributable human influence where the observed trend is consistent with HIST-ALL, but not HIST-NAT (A_n). That is, anthropogenic influence is inferred if an observed trend is inconsistent with climate model runs simulating natural variability (forced-natural plus internal variability) only, while being consistent with runs that include both anthropogenic and natural forcings. Consistency with a given distribution is defined here as having observations that lie within the 5th–95th percentile range of the modeled distribution. Consistency of observed trends with the HIST-ALL simulations is represented by categories A_N and

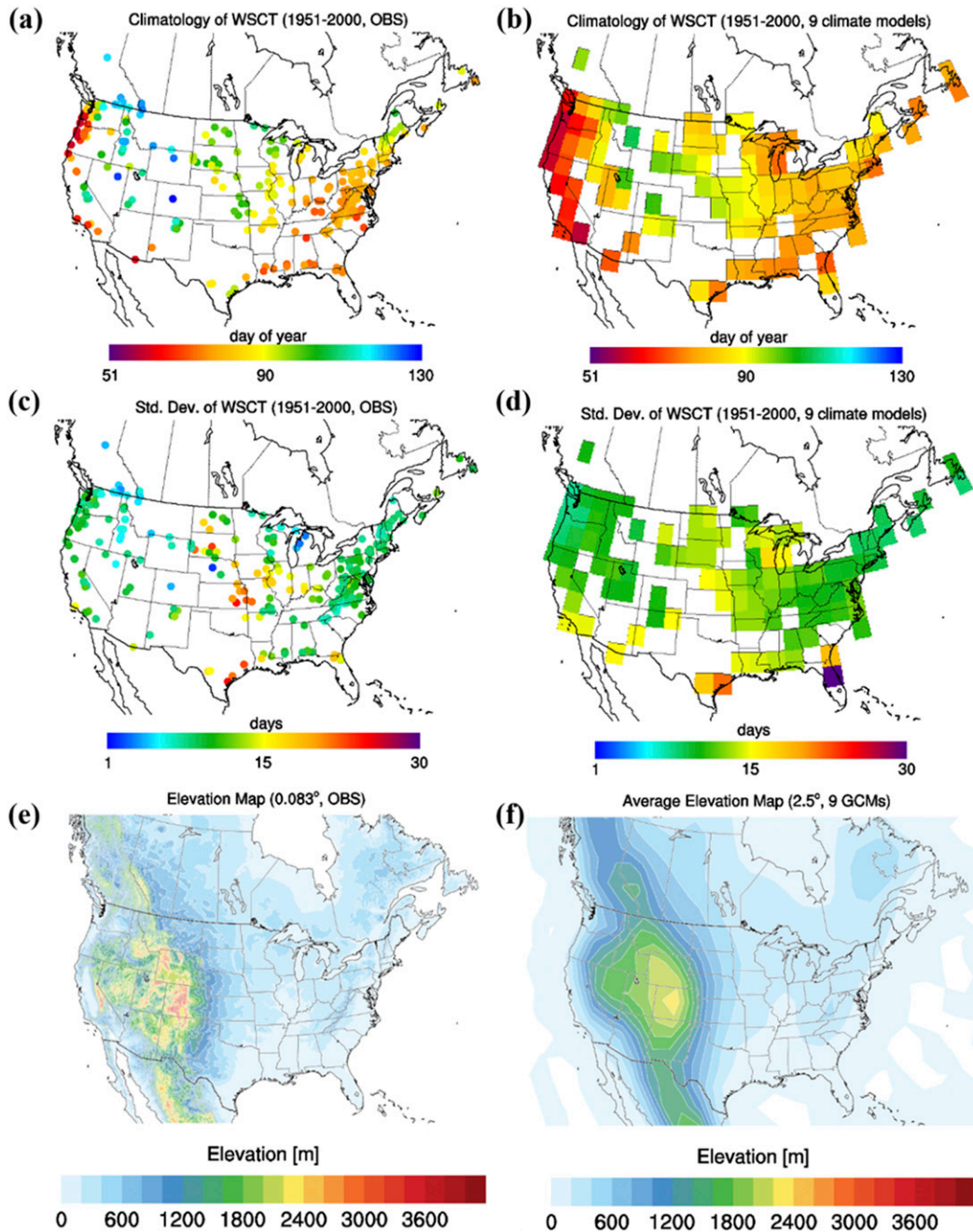


FIG. 4. Comparison of WSCT climatology and standard deviations and elevations between observations and CMIP5. (a),(b) Spatial distributions of the climatology and (c),(d) interannual standard deviation for WSCT are compared between observations and models. The model grand ensemble mean climatology in (b) is based on the average of ensemble means of the nine CMIP5 model HIST-ALL forcing runs. For the models' interannual variability in (d) the standard deviations of individual ensemble members for each of the nine models are computed and averaged to create an ensemble standard deviation estimate for each model. These are then averaged across the nine models to create a grand ensemble standard deviation. The colors depict the calendar DOY in (a) and (b) and number of days in (c) and (d). (e) Digital elevation map at 8-km (0.083°) spatial resolution from observations. (f) The averaged elevation at 250-km (2.5°) spatial resolution from the nine CMIP5 models.

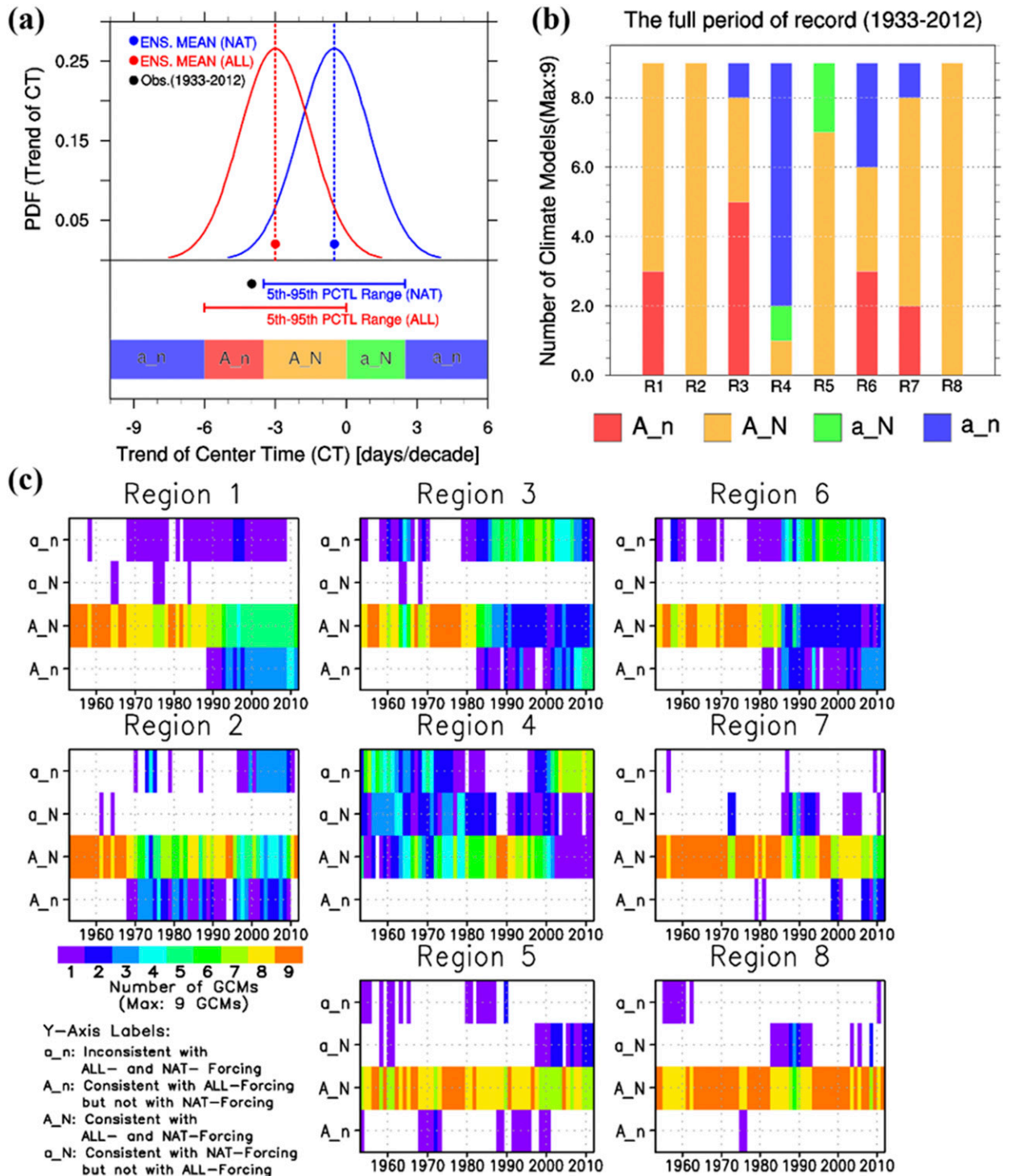


FIG. 5. Detection, attribution, and consistency of WSCT trends. (a) Schematic diagram of the categories for the detection, attribution, and consistency assessments. The four categories are depicted at the bottom: A_n, A_N, a_N and a_n (see text). (b) Summary of the number of climate models assigned to each of the four categories based on the trend analysis for the full period of record (1933–2012) (c) Time series of the number of climate models in each of the four categories, with all the trends starting in 1933. The x axis depicts the end years of the trends.

A_n [i.e., trends can be consistent with HIST-ALL yet be either detectably distinct (A_n) or not distinct (A_N) from the natural-forcing-only simulations]. Categories where the observed trend is inconsistent with the all-forcing simulations indicate cases where models are having difficulty simulating the observed trend behavior, using the current best estimates of historical climate forcing and accounting for the estimated potential influence of internal variability.

Using a criterion of at least five out of the nine individual models in agreement for a robust result, we find that the only tentative case of robust anthropogenic influence is the observed decrease in WSCT over region 3 (north-central United States; 5 models) (Fig. 5b). The most robust detectable signals overall (the sum of A_n and a_n) are in regions 3 and 4 (central plains; 6 and 7 models, respectively), and region 6 (extreme northeastern United States and Canadian Maritimes; 6 models). The HIST-ALL runs are consistent with the observed trends (the sum of A_n and A_N) for the majority of models in all eight study regions except for region 4 (central plains). Region 4, which has only one station with available streamflow records for 1933–2012, has a detectable *increasing* trend in WSCT that is inconsistent with eight of nine HIST-ALL runs, most of which simulate a decreasing trend in WSCT in this region (Fig. S1). Region 6 (extreme northeastern United States and Canadian Maritimes) has a significant decreasing trend, detectable according to six of nine models, but only three of these six models have HIST-ALL runs that are consistent with the observed trend. Of the other three models, one model has too strong a decreasing trend (which one could argue amounts to a possible fourth model indicating attributable anthropogenic influence) while the other two models have positive trends (opposite sign to that observed). Regions 5 (south central United States) and 8 (central Atlantic), which serve here as representative regions for rainfall-dominant runoff regimes, have observed trends that are not detectable but are consistent with HIST-ALL and HIST-NAT runs (A_N) according to most of the climate models and regardless of the end years (Fig. 5c); these regions have little change in the observed WSCT (Fig. 1c). Regions 2 (western coastal) and 7 (New York/southern New England) have only weak, insignificant decreasing trends in WSCT.

Preliminary trend results (Fig. 1c) show that the trend toward earlier WSCT in region 1 (western mountainous region) is much weaker using the full record (1933–2012) than the shorter period (1951–2000) as used in a previous study (Barnett et al. 2008). This difference in trend is owing to a recent upswing in the WSCT time series (Fig. 1c). A recent study (Dudley et al. 2017) finds that

the trends toward earlier WSCT at gauges over the western mountain region are not statistically significant likely due to the sensitivity of trend tests to the start and end dates or difference in the streamflow timing metrics used in previous studies (Hidalgo et al. 2009; Barnett et al. 2008; Stewart et al. 2004). Our results indicate that the evidence for an attributable human influence over region 1 is only marginal (3 of 9 models) using the full period (1933–2012; Figs. 1c and 3b). Thus our analysis of the CMIP5 climate models provides only marginal support for the earlier conclusion (Hidalgo et al. 2009; Barnett et al. 2008) that the streamflow center of timing trends in the Columbia River basin (similar to our region 1) has a detectable change partially attributable to human influence; we note that WSCT trends in our region 1 are much more pronounced over the shorter period (1951–2000) than for the full/extended period analyzed here (1933–2012).

This example shows how some climate change detection results for trends in WSCT can be sensitive to the start and end years of analysis. In comparison to previous studies, there are a number of differences in methodologies (precise region definition, streamflow center time metric, climate models used, regional downscaling use, detection/attribution method, etc.) between our study and that of Barnett et al. (2008), for example, precluding a direct comparison with our study. In any case, our detection results using additional years of data (post-1999 and pre-1950), suggest that it may be prudent to revisit some previous attribution conclusions, since we find that streamflow timing trend magnitudes in the western mountains region are decreased substantially with the additional years of data. Further illustration of the effects of different start dates or end dates on the trend assessments is provided in Fig. 4c and Fig. S2. Our preference is to use the full (longest available) for assessment if possible (e.g., if observations are adequate), as shown in Fig. 5b.

c. Tests accounting for model variance discrepancies versus observations

Realistic simulation of variances of WSCT in the climate models is important for detectability, consistency, and attribution tests, because an overestimation (underestimation) of variance in WSCT makes it too difficult (easy) to detect trends against the background of natural variability, and too easy (difficult) for the HIST-ALL runs to be consistent with the observed values. Figure 6a shows that the estimated observed internal decadal variability in regions 1 and 6 is less than simulated in any of the nine models, while in regions 2, 5, and 8 it is less than simulated in almost all of the models. The tendency toward overestimation of

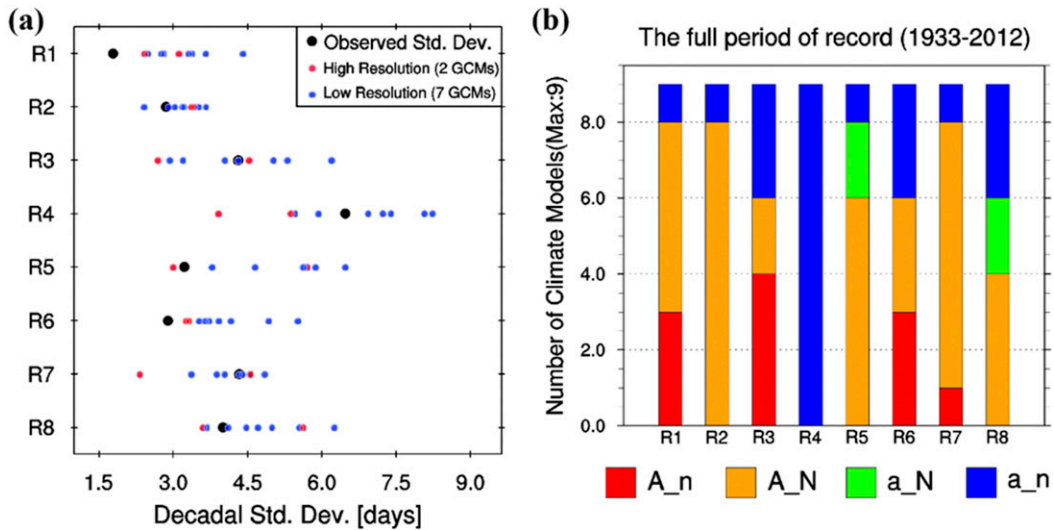


FIG. 6. Comparison of decadal internal variability between observations and CMIP5, and detection/attribution results using bias-adjusted model variability. (a) Decadal standard deviations of WSCT for each study region. The black solid dots depict the observed internal decadal variability estimate, computed from the residual time series defined by the observed time series minus the grand ensemble mean (the mean of the ensemble means of the nine climate models) of the HIST-ALL model simulations. The red and blue dots depict the simulated internal decadal variability of the CONT runs from two high-resolution GCMs (CNRM-CM5 and CSIRO-Mk3.6.0) and seven low-resolution GCMs, respectively. (b) Sensitivity test for the number of climate models assigned to each of the four categories based on the trend analysis for the full period of record (1933–2012) using adjusted 5th–95th percentile envelope ranges obtained by scaling the simulated decadal standard deviations to match observed levels of variability.

decadal variability in the (snow affected) regions 1 and 6 is consistent with results for interannual variability, already noted and discussed. The overestimate of WSCT decadal internal variability implies that the detection results from our analysis using the CMIP5 models are likely too conservative, since the excessively wide range of the modeled trend distributions makes it too difficult to detect a climatic change compared to natural variability.

Recognizing the uncertainties in our control run-based detection and attribution method, we conduct an alternative conventional linear trend analysis to test for detectable changes of WSCT in observations (Table 2). (To address possible temporal persistence/autocorrelation issues, we compute the lag-1 autocorrelation coefficients of residuals from a linear trend and find no significant autocorrelations.) The conventional linear trend analysis indicates detectable changes ($p < 0.05$) in WSCT over regions 3, 4, and 6. This is consistent with the results from our control run-based method, which indicates detectable changes in a majority of models in these same three regions.

While the conventional linear trend tests allow for a check on the robustness of our detection results to the apparent high bias in simulated internal variability, they do not address the related issue that excessive simulated

internal variability will also make it too easy for the HIST-ALL runs to be consistent with observations, nor do they address attribution. Therefore, we also conduct an adjusted variability analysis on the trend envelope range (Fig. 6b) used for detecting and attributing anthropogenic influence on WSCT trends. The adjusted variability results can be compared with the unadjusted variability results in Fig. 5b as a sensitivity test. For this sensitivity test, we scale the 5th–95th percentile envelope ranges of nine individual models by the ratios of the observed to the simulated decadal standard deviations, to adjust for their under- or overestimated decadal standard deviations. Results from this adjusted variability analysis show slightly more tentative anthropogenic attribution results, with only marginal evidence overall for detectable anthropogenic influence for region 3 (four of nine models) or regions 1 and 6 (three of nine models). The most robust detectable changes, regardless of cause, are again found for regions 3, 4, and 6, as with the initial trend assessment. Results for regions 2, 5, 7, and 8 are similar to the initial assessment in terms of detection, with none showing evidence of detectable trends. The HIST-ALL runs are inconsistent with the observed trends in region 4, as with the initial assessment, and also are inconsistent for a majority of models in region 8 in the adjusted analysis.

TABLE 2. Linear WSCT trend slopes and lag-1 WSCT autocorrelation coefficients, with p values, for eight study regions.

Region	1933–2012			
	Slope (days yr ⁻¹)		Lag-1 autocorrelation	
	Value	p value	Value	p value
R1	-0.0497	0.081	-0.0832	0.2359
R2	-0.0209	0.375	-0.0737	0.2621
R3	-0.11027	0.003	0.05826	0.3062
R4	0.182172	0.005	0.08924	0.2186
R5	0.013475	0.734	0.00503	0.4825
R6	-0.07901	0.045	0.08175	0.2384
R7	-0.04378	0.299	-0.02385	0.4179
R8	0.053034	0.230	-0.1850	0.0524

4. Discussion

The main goal of this study is to apply a detection and attribution methodology to winter–spring streamflow timing based on available observational and CMIP5 climate model data. Based on either the conventional linear trend analysis or comparisons of observed trends with the HIST-NAT distributions (with either unadjusted or adjusted modeled variability), the observed changes in WSCT over regions 3, 4, and 6 show the most robust detection signal. Furthermore, marginal evidence for an attributable human influence on streamflow timing is found for regions 1, 3, and 6 for both unadjusted and adjusted variability tests. For the adjusted variability cases an attributable human influence is inferred for three, four, and three of the nine models, respectively for these three regions.

The adjusted versus unadjusted variability tests reinforce the importance of the estimation of variance for the attribution results, as the adjusted results point toward slightly more tentative (i.e., less robust) attribution conclusions than the unadjusted variability cases. Future high-resolution climate model simulations should lead to a more reliable assessment of water resources impacts of climate change in regions with temperature-sensitive (snowmelt) runoff regimes. Along these lines, a recent study (Kapnick and Delworth 2013) has explored the impact of increasing climate model resolution on present-day and future climate simulations of cold-season hydroclimate. They find that enhanced resolution produces both a more realistic present-day climate of snow variables as well as notable differences in their response to warming in some mountain regions, including future projected increases in snowfall in a few high-elevation regions where an increase is not simulated with a lower-resolution model. While having more realistic variability from higher-resolution models could affect our detection results, especially in regions with the strongest positive biases in variability, our alternative

tests using either conventional linear trend analysis or adjusted variability (trend percentile ranges) suggest that the simulated variability biases may not have a very large impact on our overall detection and attribution results.

The findings point toward the need to further investigate several model biases tentatively identified. However, we note that the apparent biases in the means and variability of streamflow timing differ among models, which complicates the physical interpretation of these biases in terms of specific causes. Reduction of the biases we have identified will require the identification of specific physical processes in the CMIP5 climate models that are poorly represented (e.g., subgrid-scale orography effects) and thus need improvement to more realistically capture climate variability. For example, in not being able to resolve complex mountain/valley terrain in a model, the timing of snowmelt aggregated across such varied terrain will be challenging to model adequately (Wang et al. 2016). While reducing such simulation biases is beyond the scope of this study, our study should at least provide motivation for further efforts by the various modeling groups along these lines (e.g., higher-resolution modeling and/or statistical downscaling).

There are a number of different possible approaches for comparing model-simulated variability to observations, including both historical and paleoclimate perspectives. The approach that we use, as a first step for multimodel assessment, is to estimate the real-world internal variability as a residual by subtracting from historical observations the CMIP5 All-Forcing response (an estimate of the real-world response to external forcings). For a regional-scale study, however, alternative methods using paleoclimate proxies are worth exploring in future studies, especially for western hydroclimate (Barnett et al. 2008).

5. Conclusions

In summary, there is marginal evidence for an emerging detectable anthropogenic contribution toward earlier WSCT in parts of North America. The regions with the strongest relative indication of an anthropogenic contribution in our analysis include the north-central United States (region 3), the mountainous western United States/southwestern Canada (region 1), and the extreme northeastern United States and Canadian Maritimes (region 6). However, in none of the regions examined do a majority of the nine CMIP5 models examined robustly support a detectable attribution of an earlier (decreasing) WSCT trend to anthropogenic forcing. At some level, the difficulty in detecting a climate change signal comes down to a low signal-to-noise ratio

(Ziegler et al. 2005). Apparently, for the variable at hand, the climate change influence is not very large compared to interannual/interdecadal variability noise. Over regions 3 and 6, a statistically significant decreasing trend of WSCT was identified in a previous study (Hodgkins and Dudley 2006) but not assessed for attribution to anthropogenic forcing. Our multimodel detection and attribution assessment suggests that earlier findings of anthropogenic influence on the decreasing WSCT trend over some western U.S. mountain regions in previous studies (Hidalgo et al. 2009; Barnett et al. 2008; Stewart et al. 2004), although using different methods and region definitions, may also be sensitive to inclusion of additional (earlier and more recent) years of data in the analysis. However, if anthropogenic climate change is causing an emerging shift toward earlier winter–spring streamflow timing in the above regions, as our analysis tentatively suggests, this would likely have important implications for future climate change–related impacts on human activities and ecosystems in these areas.

Acknowledgments. We thank the WCRP's Working Group on Coupled Modeling, and the modeling groups participating in CMIP5. This study is supported by the NOAA Oceanic and Atmospheric Research (OAR) program (NA14OAR4320106, for J.K.).

REFERENCES

- Akhter, J., L. Das, J. K. Meher, and A. Deb, 2018: Uncertainties and time of emergence of multi-model precipitation projection over homogeneous rainfall zones of India. *Climate Dyn.*, **50**, 3813–3831, <https://doi.org/10.1007/s00382-017-3847-y>.
- Barnett, T. P., and Coauthors, 2008: Human-induced changes in the hydrology of the western United States. *Science*, **319**, 1080–1083, <https://doi.org/10.1126/science.1152538>.
- Cayan, D. R., S. A. Kammerdiener, M. D. Dettinger, J. M. Caprio, and D. H. Peterson, 2001: Changes in the onset of spring in the western United States. *Bull. Amer. Meteor. Soc.*, **82**, 399–416, [https://doi.org/10.1175/1520-0477\(2001\)082<0399:CITOOS>2.3.CO;2](https://doi.org/10.1175/1520-0477(2001)082<0399:CITOOS>2.3.CO;2).
- Crecco, V. A., and T. F. Savoy, 1985: Effects of biotic and abiotic factors on growth and relative survival of young American shad, *Alosa sapidissima*, in the Connecticut River. *Can. J. Fish. Aquat. Sci.*, **42**, 1640–1648, <https://doi.org/10.1139/f85-205>.
- Dudley, R. W., G. A. Hodgkins, M. R. McHale, M. J. Kolian, and B. Renard, 2017: Trends in snowmelt-related streamflow timing in the conterminous United States. *J. Hydrol.*, **547**, 208–221, <https://doi.org/10.1016/j.jhydrol.2017.01.051>.
- Goulding, H. L., T. D. Prowse, and B. Bonsal, 2009: Hydroclimatic controls on the occurrence of break-up and ice-jam flooding in the Mackenzie Delta, NWT, Canada. *J. Hydrol.*, **379**, 251–267, <https://doi.org/10.1016/j.jhydrol.2009.10.006>.
- Hare, J. A., and Coauthors, 2016: A vulnerability assessment of fish and invertebrates to climate change on the Northeast U.S. continental shelf. *PLOS ONE*, **11**, e0146756, <https://doi.org/10.1371/journal.pone.0146756>.
- Hidalgo, H. G., and Coauthors, 2009: Detection and attribution of streamflow timing changes to climate change in the western United States. *J. Climate*, **22**, 3838–3855, <https://doi.org/10.1175/2009JCLI2470.1>.
- Hodgkins, G. A., and R. W. Dudley, 2006: Changes in the timing of winter–spring streamflows in eastern North America, 1913–2002. *Geophys. Res. Lett.*, **33**, L06402, <https://doi.org/10.1029/2005GL025593>.
- Kang, D. H., H. Gao, X. Shi, S. ul Islam, and S. J. Déry, 2016: Impacts of a rapidly declining mountain snowpack on streamflow timing in Canada's Fraser River basin. *Sci. Rep.*, **6**, 19299, <https://doi.org/10.1038/srep19299>.
- Kapnick, S. B., and T. L. Delworth, 2013: Controls of global snow under a changed climate. *J. Climate*, **26**, 5537–5562, <https://doi.org/10.1175/JCLI-D-12-00528.1>.
- Knutson, T. R., F. Zeng, and A. T. Wittenberg, 2013: Multimodel assessment of regional surface temperature trends: CMIP3 and CMIP5 twentieth-century simulations. *J. Climate*, **26**, 8709–8743, <https://doi.org/10.1175/JCLI-D-12-00567.1>.
- Mengelkamp, H.-T., K. Warrach, C. Ruhe, and E. Raschke, 2001: Simulation of runoff and streamflow on local and regional scales. *Meteor. Atmos. Phys.*, **76**, 107–117, <https://doi.org/10.1007/s007030170042>.
- Milly, P. C. D., J. Betancourt, M. Falkenmark, R. M. Hirsch, Z. W. Kundzewicz, D. P. Lettenmaier, and R. J. Stouffer, 2008: Stationarity is dead: Whither water management? *Science*, **319**, 573–574, <https://doi.org/10.1126/science.1151915>.
- Sen, P. K., 1968: Estimates of the regression coefficient based on Kendall's tau. *J. Amer. Stat. Assoc.*, **63**, 1379–1389, <https://doi.org/10.1080/01621459.1968.10480934>.
- Slack, J. R., and J. M. Landwehr, 1992: Hydro-climatic data network (HCDN); a U.S. Geological Survey streamflow data set for the United States for the study of climate variations, 1874–1988. Open-File Rep. 92-129, 193 pp., <https://pubs.er.usgs.gov/publication/ofr92129>.
- Stewart, I. T., D. R. Cayan, and M. D. Dettinger, 2004: Changes in snowmelt runoff timing in western North America under a 'business as usual' climate change scenario. *Climatic Change*, **62**, 217–232, <https://doi.org/10.1023/B:CLIM.0000013702.22656.e8>.
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, <https://doi.org/10.1175/BAMS-D-11-00094.1>.
- Wang, R., M. Kumar, and T. E. Link, 2016: Potential trends in snowmelt-generated peak streamflows in a warming climate. *Geophys. Res. Lett.*, **43**, 5052–5059, <https://doi.org/10.1002/2016GL068935>.
- Ziegler, A. D., E. P. Maurer, J. Sheffield, B. Nijssen, E. F. Wood, and D. P. Lettenmaier, 2005: Detection time for plausible changes in annual precipitation, evapotranspiration, and streamflow in three Mississippi River sub-basins. *Climatic Change*, **72**, 17–36, <https://doi.org/10.1007/s10584-005-5379-4>.