Reconciling Precipitation with Runoff: Observed Hydrological Change in the Midlatitudes

JOE M. OSBORNE,* F. HUGO LAMBERT,* MARGRIET GROENENDIJK,[+] ANNA B. HARPER,*
CHARLES D. KOVEN,[#] BENJAMIN POULTER,[@] THOMAS A. M. PUGH,[&] STEPHEN SITCH,[+]
BENJAMIN D. STOCKER,**·[@@] ANDY WILTSHIRE,[++] AND SÖNKE ZAEHLE[##]

*College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, United Kingdom
[+]College of Life and Environmental Sciences, University of Exeter, Exeter, United Kingdom
[#]Lawrence Berkeley National Laboratory, Berkeley, California
[@]Department of Ecology, Montana State University, Bozeman, Montana
[&]Institute of Meteorology and Climate Research, Atmospheric Environmental Research, Karlsruhe Institute
of Technology, Garmisch-Partenkirchen, Germany
**Climate and Environmental Physics, Physics Institute, and Oeschger Centre for Climate Change Research,
University of Bern, Bern, Switzerland
[++]Met Office Hadley Centre, Exeter, United Kingdom
[##]Biogeochemical Integration Department, Max Planck Institute for Biogeochemistry, Jena, Germany

ABSTRACT

Century-long observed gridded land precipitation datasets are a cornerstone of hydrometeorological research. But recent work has suggested that observed Northern Hemisphere midlatitude (NHML) land mean precipitation does not show evidence of an expected negative response to mid-twentieth-century aerosol forcing. Utilizing observed river discharges, the observed runoff is calculated and compared with observed land precipitation. The results show a near-zero twentieth-century trend in observed NHML land mean runoff, in contrast to the significant positive trend in observed NHML land mean precipitation. However, precipitation and runoff share common interannual and decadal variability. An obvious split, or breakpoint, is found in the NHML land mean runoff–precipitation relationship in the 1930s. Using runoff simulated by six land surface models (LSMs), which are driven by the observed precipitation dataset, such breakpoints are absent. These findings support previous hypotheses that inhomogeneities exist in the early-twentieth-century NHML land mean precipitation record. Adjusting the observed precipitation record according to the observed runoff record largely accounts for the departure of the observed precipitation response from that predicted given the real-world aerosol forcing estimate, more than halving the discrepancy from about 6 to around $2 \, \mathrm{W \, m^{-2}}$. Consideration of complementary observed runoff adds support to the suggestion that NHML-wide early-twentieth-century precipitation observations are unsuitable for climate change studies. The agreement between precipitation and runoff over Europe, however, is excellent, supporting the use of whole-twentieth-century observed precipitation datasets here.

[@@]Current affiliation: Department of Life Sciences, Imperial College London, Silwood Park Campus, Ascot, United Kingdom.

Corresponding author address: Joe M. Osborne, Laver Building, University of Exeter, North Park Road, Exeter EX4 4QE, United Kingdom.
E-mail: j.m.osborne@exeter.ac.uk

1. Introduction

Precipitation and runoff are two intrinsically related components of the land hydrological cycle, but they are often studied in isolation (Xu et al. 2010). Consideration of observed runoff can complement observed precipitation and add extra weight to hypotheses for recent hydrological changes. Runoff observations can also be used to validate observed precipitation changes, allowing for the identification of inhomogeneities that cannot be explained by natural variability or anthropogenic influences on the hydrological cycle.

Recent work by Osborne and Lambert (2014, hereafter OL14) showed that observed Northern Hemisphere mid-latitude (NHML) land mean precipitation changes contain no evidence of a response to sharply increasing aerosol emissions in the mid-twentieth century (Lamarque et al. 2010). The NHML region was the main source of these emissions, which led to a well understood decrease in observed precipitation over the remote Northern Hemisphere tropical land region. This pattern is robustly simulated by the latest general circulation models (GCMs; Hwang et al. 2013; Wu et al. 2013). Theory suggests that such a decrease should also be evident locally in observed NHML land mean precipitation. Indeed, GCMs typically simulate this decrease, with the magnitude of the decrease dependent on the strength of NHML aerosol forcing within the model. Conversely, observed NHML land mean precipitation increases rather than decreases. This could be explained by a major misunderstanding of aerosol radiative effects, contradicting a long-established scientific literature (Charlson et al. 1992). Alternatively, it is possible that precipitation in the NHML land region is poorly simulated by the latest models and that they fail to capture important physical processes, such as land–atmosphere coupling (Koster et al. 2004). A final possible explanation concerns the quality of existing precipitation observations. Inhomogeneities in widely used gridded observed land precipitation datasets have been extensively documented (Groisman et al. 1991; Legates 1995), with records from the early twentieth century particularly vulnerable to inhomogeneities.

In the absence of impacts of human activity and where natural changes in storage terms, such as soil infiltration and deep percolation, can be neglected—typically true on annual or longer time scales (Walter et al. 2004; Jung et al. 2013)—we can apply a simple water budget equation to a region consisting of one or more river catchments:

$$P = Q + \text{ET}, \tag{1}$$

where P, Q, and ET are precipitation, runoff, and evapotranspiration, respectively. Changes in runoff in many river catchments reflect changes in precipitation (Milliman et al. 2008). However, human activities can influence this water budget by changing the partitioning of precipitation into evapotranspiration and runoff and through changes in water storage, giving

$$P = Q + \text{ET} + \Delta S, \tag{2}$$

where ΔS is the net change in water storage due to human activities (Adam and Lettenmaier 2008). Dam and reservoir construction can affect seasonal water storage, but they appear to have little effect at annual time scales (Adam et al. 2007; Adam and Lettenmaier 2008). Melting ground ice in permafrost, due to human-induced warming (Bindoff et al. 2013), could lead to decreased water storage and has been invoked in explaining the disparity between decreasing precipitation and increasing runoff in northern Eurasia river catchments (Adam and Lettenmaier 2008).

Humans can have a number of direct, but also poorly quantified, influences on the hydrological cycle, affecting the partitioning of precipitation into evapotranspiration and runoff. A major influence is water withdrawal for irrigation, which can have a big impact on runoff at the catchment scale (Milliman et al. 2008; Xu et al. 2010). A fraction of this withdrawal is not returned to runoff, defined as water consumption (Xu et al. 2010; Sterling et al. 2013). Although the significance of irrigation and other water withdrawals at the catchment scale is undisputed, there is less agreement over its influence on global mean runoff (e.g., Gerten et al. 2008; Sterling et al. 2013). Some research suggests that decreases in global runoff resulting from irrigation are cancelled out by increases resulting from changes in land cover or land use (predominantly deforestation; Gordon et al. 2005; Sterling et al. 2013).

Further to these direct influences of human activity are a number of indirect influences. Rising CO_2 can cause a transpiration-reducing plant physiological response and an accompanying increase in runoff (Gerten et al. 2008). Gedney et al. (2006) suggested this effect as the reason for increases in global runoff, as described by Labat et al. (2004), before contradictory subsequent work found no significant change in global runoff over the twentieth century (Milliman et al. 2008; Dai et al. 2009). Consideration of expanding vegetation is also required to quantify the overall CO_2 effect on global runoff in the twentieth century, the sign of which is, again, uncertain. Greenhouse gas–driven global warming (Bindoff et al. 2013) has caused a slight decrease in global runoff over the twentieth century, largely through higher summer evapotranspiration (Gerten et al. 2008), although this is dependent on sufficient moisture supply (Jung et al. 2010). Solar dimming associated with mid-twentieth-century aerosol emissions (Wild 2012) has also been proposed as an indirect human influence on runoff, through forcing a decrease in evapotranspiration (Oliveira et al. 2011; Gedney

et al. 2014). The net aerosol impact is to increase runoff, with Gedney et al. (2014) finding a detectable increase in local river flow in response to increased NHML aerosol emissions in the 1960s and 1970s.

Despite the complicating mix of direct and indirect human influences on the hydrological cycle, the dominant driver of changes in runoff is changes in precipitation (Gerten et al. 2008; Milliman et al. 2008), which itself has been affected by human activities (Zhang et al. 2007; Wu et al. 2013; OL14). Gradual changes in the large-scale runoff–precipitation relationship can be expected because of the more transient human influences, such as the CO_2 physiological effect. For example, increases in evapotranspiration on the North American continent have been documented (Milly and Dunne 2001; Walter et al. 2004). But we still expect an approximately linear runoff–precipitation relationship, with sudden shifts in this relationship only at scales at, or close to, the catchment scale, perhaps due to deforestation or major dam construction. To the best of our knowledge, there is no physical explanation behind any such shifts at the scales we are investigating, which would require, for example, a temporally concentrated, continental-wide dam construction scheme.

Fekete et al. (2002) suggested that river discharge data, used to calculate catchment runoff, are more accurate than other water cycle components, with errors thought to be within 10%–20%. This is an improvement on typical precipitation errors (Hagemann and Dümenil 1997), with underestimates common because of wind-induced undercatch (Legates 1995). Snowfall undercatch can lead to biases in winter higher-midlatitude precipitation totals of 20%–40% (Adam and Lettenmaier 2003; Yang et al. 2005). Rainfall uncertainty is also dependent on the spatial scale being studied, but scale is not as important for river discharge uncertainty because of the integrated nature of measurement (McMillan et al. 2012).

At the large scale, such as the NHML land mean, different precipitation datasets share more common variability and trends (Polson et al. 2013; OL14). Therefore, any major inhomogeneities that still exist must be common across all gridded precipitation datasets. Such inhomogeneities can be a consequence of changes in measurement practice, instrument changes, and the implementation of wetting corrections (Groisman and Rankova 2001). Producing homogeneous records relies on quality observations of other meteorological variables (e.g., temperature, wind speed, and precipitation type) and metainformation about individual stations (New et al. 2000; Groisman and Rankova 2001). These are both difficult to obtain, particularly in the early twentieth century. As such, traditional approaches to correcting observations may prove fruitless. The framework presented in OL14 offers a basis for discovering inhomogeneities given the

expected response to best estimates of real-world climate forcing. Yet there is still significant uncertainty in estimates of twentieth-century aerosol forcing (Myhre et al. 2013) and a still-developing physical understanding of the aerosol effect on precipitation, particularly with regards to complex aerosol–cloud interactions (Boucher et al. 2013). Here, we develop and strengthen this physical framework by also considering observed NHML mean runoff.

Using the conclusions of OL14, we hypothesize that inhomogeneities in observed NHML land mean precipitation make the early-twentieth-century data unsuitable for climate change studies. Observed NHML mean runoff, as far as we are aware, provides the only valid test of this hypothesis. This test relies on two key assumptions: 1) that runoff observations are immune to inhomogeneities of the type and magnitude common in precipitation observations (particularly those due to changes in spatial coverage), and 2) that in the absence of inhomogeneities the runoff–precipitation relationship is approximately linear. We use a statistical model to search for sudden shifts, which we define as breakpoints, in the NHML land mean runoff–precipitation relationship, also repeating this approach at the continental scale to further pinpoint where any inhomogeneities in observed precipitation may exist. We also analyze NHML land mean runoff in six land surface models (LSMs) that form part of the Trends in Net Land–Atmosphere Carbon Exchanges (TRENDY) land surface model intercomparison project (Sitch et al. 2015). These models were all driven by observed NHML land mean precipitation and so offer the opportunity to search for breakpoints in the runoff–precipitation relationship that have arisen through a simulated physical process, thus serving to test the second assumption above. Observed NHML land mean precipitation is then adjusted using its runoff counterpart and the implications for our understanding of the response of NHML land mean precipitation to mid-twentieth-century aerosol forcing is discussed.

Section 2 details the data used; section 3 considers the observed and modeled runoff–precipitation relationship, including the approach to testing for breakpoints; and section 4 describes the adjustment applied to observed NHML land mean precipitation made on the basis of our findings. Section 5 discusses possible reasons for a breakpoint in the runoff–precipitation relationship, and in section 6 we discuss our conclusions for the effect of aerosols on NHML land mean precipitation.

2. Data and methods

a. Observed NHML land mean runoff

We use the Dai et al. (2009) continental discharge dataset to calculate observed runoff. Discharge, or
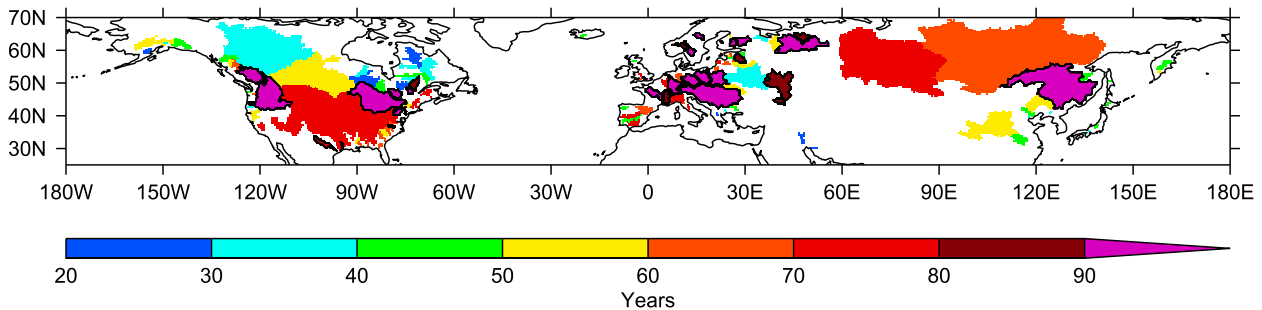
FIG. 1. NHML catchment coverage of the 148 rivers that meet the conditions set out in section 2, with the color of the catchment representative of the (not always continuous) record length from the Dai et al. (2009) discharge dataset. River catchments with discharge records at least 80 years or longer are marked with black borders.

streamflow, is the temporally lagged, spatial integral of runoff over a river basin (Milly et al. 2005). It is possible to derive estimates of runoff for a catchment by dividing river discharge at a gauging station by the upstream catchment area (Fekete et al. 2002). Dai et al. (2009) produced a temporally complete streamflow dataset for 925 of the world's largest ocean-reaching rivers for 1948–2004, by infilling data gaps through considering nearby streamflow data, where possible. Any remaining data gaps were reconstructed by using a hydrological model and the correlation between precipitation and streamflow. Because we want to contrast mid-twentieth-century aerosol-driven hydrological changes with overall twentieth-century changes, we instead use the temporally incomplete, observations-only, dataset that extends back to 1900.

We follow OL14 and use 1905–2004 as our twentieth-century period, but instead use the water year, October–September, with the first month contributing toward our period being October 1904. This minimizes the lag effect associated with early winter snowfall being locked within a catchment until the spring melt of the following calendar year (Dai et al. 2009). Owing to this lag effect, we calculate mean water-year runoff for a given catchment only where all 12 months' worth of data are available. Further, the streamflow records are conditioned so that rivers must contain a minimum record length of 20 years in the 30-yr 1961–90 period, to which data are anomalized to throughout this work. In OL14, the NHML region is defined as the latitude band between 30° and 65°N. We stipulate that >80% of the catchment area upstream of the farthest downstream gauging station must fall within this band for a river to contribute. This leaves us with a total of 148 rivers. Figure 1 shows the spatial coverage and record length of the 148 rivers.

b. Observed NHML land mean precipitation

In OL14, four gridded observational precipitation datasets were used in total. We ignored spatially interpolated versions of two of the datasets, which offer complete global terrestrial coverage. Here, such a product is necessary to allow for direct comparison with the spatial and temporal coverage of observed NHML land mean runoff. To ensure an accurate comparison between precipitation and runoff in each catchment, we produce high-resolution catchment masks on a $0.5° \times 0.5°$ grid to match that of the precipitation dataset used. We select the latest Climatic Research Unit Time Series, version 3.21 (CRU TS3.21), high-resolution precipitation dataset (Harris et al. 2014), which has been modified for use in driving the TRENDY models. This modification merged the monthly climatology of CRU TS3.21 with the National Centers for Environmental Prediction (NCEP)–National Center for Atmospheric Research (NCAR) reanalyses (from 1948 to present) dataset (Kalnay et al. 1996), bilinearly interpolated to the CRU TS3.21 resolution, to generate the diurnal and daily variability required to drive the TRENDY models at 6-hourly time steps [see Fisher et al. (2013) for further details]. The merged dataset (CRU–NCEP, hereafter CRUNCEP) offers a finer temporal resolution for modeling purposes, but monthly precipitation totals remain the same. Figure 2 shows the twentieth-century NHML annual spatial coverage for both runoff and the precipitation (with and without the time-varying runoff mask applied) station network. During some later decades, runoff observations offer roughly a fivefold increase in coverage over precipitation. This is, of course, an overestimate for precipitation, since gauges only provide a point measurement and many grid boxes are poorly gauged, especially in mountainous regions where precipitation is highly variable (Adam et al. 2006).

c. TRENDY models

We use an ensemble of six LSMs, summarized in Table 1. The models were forced over the historical period with CRUNCEP precipitation, as well as other observed climate variables that were derived from merging CRU and NCEP data and changing CO_2
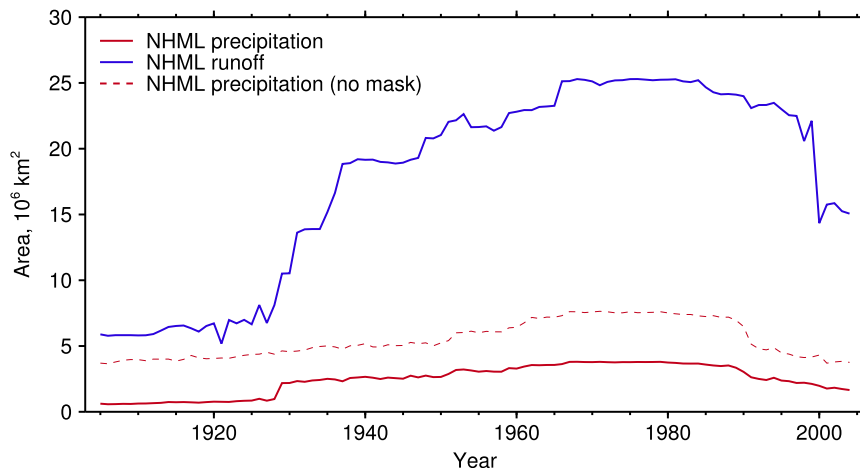
FIG. 2. Max annual NHML land area coverage of observed runoff and precipitation using the Dai et al. (2009) discharge dataset and the CRU TS3.21 precipitation (using the raw product, which only considers grid boxes where real observations exist), respectively. Both river catchment masks and the CRU TS3.21 dataset are computed on a high-resolution $0.5° \times 0.5°$ grid.

concentrations (Sitch et al. 2015; Le Quéré et al. 2014). These runs were also driven by historical land-use changes, calculated from the History Database of the Global Environment (HYDE; Klein Goldewijk and Verburg 2013). Although Yang et al. (2015) report some sensitivity of runoff trend in TRENDY models to land-use change at the catchment scale, our conclusions are insensitive to the inclusion or exclusion of the historical land-use dataset. Simulated runoff is available as output at a monthly frequency. Model output not already at $0.5° \times 0.5°$ resolution is regridded using bilinear interpolation. Simulated runoff is then masked to the varying spatial and temporal coverage of the runoff (streamflow) observations.

d. CMIP5 models

We compare observed NHML land mean precipitation with that modeled by 15 models participating in phase 5 of the Coupled Model Intercomparison Project (CMIP5; Taylor et al. 2012). In contrast to OL14, we have omitted GCMs with a historical (all forcings) experiment ensemble size of just one, as well as the two GISS GCMs, from our analyses. OL14 found the GISS

models to have anomalously strong NHML surface longwave forcing given their NHML surface aerosol forcing. Therefore, shortwave-induced surface cooling is counteracted by significant longwave-induced surface warming. Also, it has been subsequently noted that GISS-E2-R does not produce the interhemispheric forcing asymmetry expected from its strongly negative NHML surface aerosol forcing because it overestimates negative Southern Hemisphere nitrate aerosol and ozone forcing (Shindell 2014). This clearly influenced the results presented by OL14. Simulated precipitation in the CMIP5 models is also regridded to $0.5° \times 0.5°$ resolution and masked to the availability of runoff observations.

Using these data we calculate the precipitation response to aerosol forcing, which we call the offset (see section 2e), and plot this against the strength of NHML surface aerosol forcing—calculated using the Forster et al. (2013) linear forcing–feedback model and using shortwave forcing as a proxy for total aerosol forcing—for each of the 15 CMIP5 models, as well as the observations (an estimate of real-world- or IPCC-equivalent NHML surface aerosol forcing is used for the observations; see

TABLE 1. Overview of the ensemble of six TRENDY LSMs.

Model name	Abbreviation	Original resolution	Reference
Community Land Model, version 4.5	CLM4.5	$1.25° \times 0.9375°$	Oleson et al. (2013)
Joint UK Land Environment Simulator	JULES	$1.875° \times 1.25°$	Clark et al. (2011); Best et al. (2011)
Lund–Potsdam–Jena Land Model	LPJ	$0.5° \times 0.5°$	Sitch et al. (2003)
Lund–Potsdam–Jena General Ecosystem Simulator	LPJG	$0.5° \times 0.5°$	Smith et al. (2001); Ahlström et al. (2012)
Land Surface Processes and Exchanges–Bern	LPX	$1.0° \times 1.0°$	Stocker et al. (2013)
ORCHIDEE–Carbon and Nitrogen	OCN	$1.0° \times 1.0°$	Zaehle and Friend (2010); Zaehle et al. (2010)

OL14). Figure 3a shows this updated version of Fig. 3c from OL14. Overall, there is little difference between the plots, with the stark contrast between the observed precipitation response and the modeled precipitation responses still evident. Removing this mask and instead considering all grid boxes in the CRUNCEP dataset between 30° and 65°N, we see an even more robust relationship between the CMIP5 precipitation response to aerosol forcing and the CMIP5 NHML surface aerosol forcing (Fig. 3b). This suggests that masking the precipitation dataset to the runoff observations weakens this robust signal, but it remains significant.

e. NHML land mean precipitation offset

In OL14 it is shown that the NHML land mean precipitation response to aerosol is stronger than that to greenhouse gases in GCMs, but the global mean temperature response to both is similar. Therefore, with a step in aerosol emissions around the 1960s (Lamarque et al. 2010), a simple linear regression approach is used to model the NHML land mean precipitation–global mean temperature relationship as a greenhouse gas–driven straight line offset from pre-1960 data by a mid-twentieth-century increase in aerosol forcing. To calculate the size of the NHML land mean precipitation offset β_0, we fit the following model to the data:

$$P_{60-04i} = \beta_0 + \beta_1 T_{60-04i} + \varepsilon_{60-04i}, \qquad (3)$$

where P_{60-04} are 5-yr mean NHML land mean precipitation anomalies for the 1960–2004 period with respect to the 1905–59 mean NHML land mean precipitation anomaly and T_{60-04} are 5-yr mean global mean temperature anomalies for the 1960–2004 period with respect to the 1905–59 mean global mean temperature anomaly. The intercept term β_0 is the NHML land mean precipitation offset, a measure of the precipitation response to aerosol forcing, and β_1 is the slope from the linear regression fit to 1960–2004 five-year means. The residuals ε_{60-04} are assumed to be identically and independently distributed with zero mean. This model is fitted to the observations and each of the CMIP5 models in turn. Further details on this offset and physical mechanisms behind it are discussed in OL14 (and references therein).

3. Runoff–precipitation relationship

Ideally, we would be able to reproduce Fig. 3c of OL14 using just NHML land mean runoff, with an expectation that interannual, decadal, and multidecadal changes in runoff should match those in precipitation. However, we find the representation of runoff in CMIP5 models to be suboptimal, with the runoff–precipitation
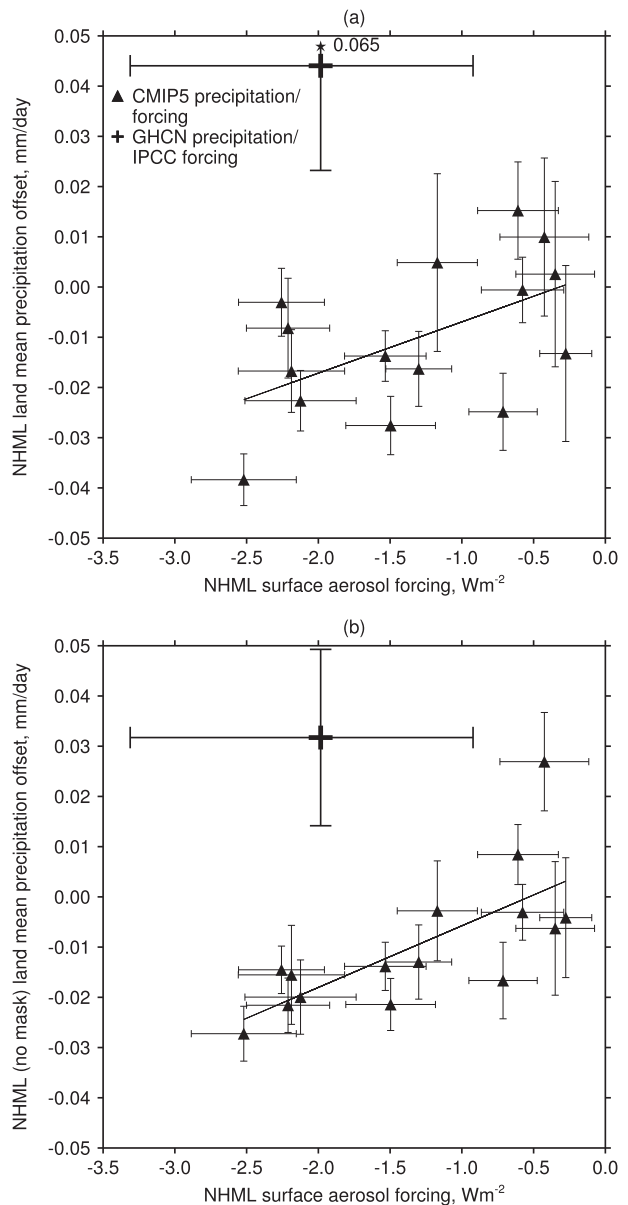

FIG. 3. NHML land mean precipitation offset against NHML surface aerosol forcing. Error bars represent the 5%–95% uncertainty range, with real-world- or IPCC-equivalent NHML surface aerosol forcing and the associated 5%–95% uncertainty range used for the observed data point. (a) CRUNCEP masked to the availability of runoff observations and (b) all CRUNCEP grid boxes between 30° and 65°N. The value next to the error bar in (a) indicates where the uncertainty extends beyond the scale. There is a significant positive correlation between precipitation offset and aerosol forcing across the 15 CMIP5 GCMs in both (a) and (b) ($r = 0.54$, $p < 0.05$ and $r = 0.87$, $p < 0.01$, respectively). For further details on the methods used to calculate the offsets, aerosol forcings, and associated uncertainty ranges, refer to OL14.
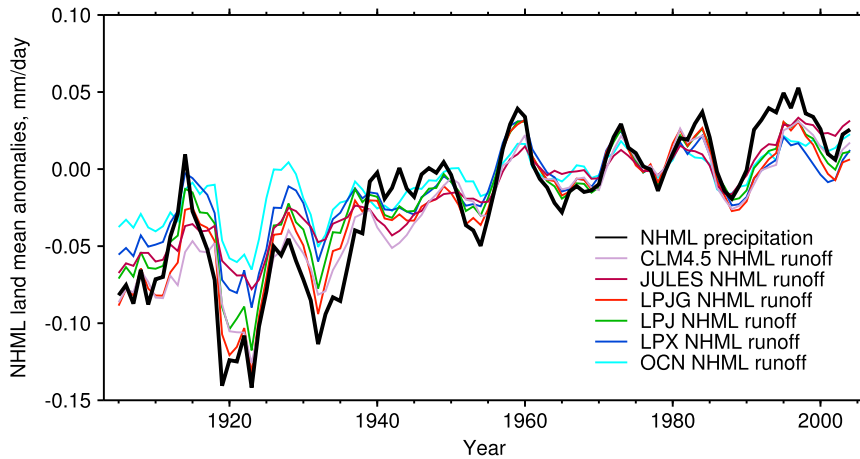
FIG. 4. The 5-yr running mean NHML land mean model-simulated runoff and observed precipitation anomalies. Data are from the six TRENDY LSMs (colored lines) and CRUNCEP dataset (thick black line) for runoff and precipitation, respectively.

relationship frequently nonlinear. Bias in the simulated precipitation climatology leads to bias in aridity and, in turn, the runoff coefficient (runoff divided by precipitation), in line with the Budyko framework (Budyko 1974). Runoff outputs from GCMs have also been described as biased because of the representation of runoff processes (Hirabayashi et al. 2013). This may be a consequence of low resolution or insufficient subgrid parameterization (Sperna Weiland et al. 2012). In some GCMs, runoff may be simply advected straight to the ocean, meaning that runoff is poorly represented in larger river catchments because time lags between runoff generation and return to the ocean are not considered (Falloon et al. 2011; Sperna Weiland et al. 2012). Because of the coarse resolution and reduced complexity of land surface schemes in GCMs relative to offline LSMs, changes in runoff are often investigated using direct GCM runoff output integrated to river discharges using a river routing scheme (Milly et al. 2005; Hirabayashi et al. 2013), or through using observed precipitation or GCM precipitation output to drive LSMs (Gerten et al. 2008; Schewe et al. 2014), as we have done with the TRENDY models.

Although we consider direct runoff output, rather than routed discharge, the six TRENDY LSMs all simulate NHML land mean runoff that is strongly correlated with observed NHML land mean precipitation used to drive the experiments (Fig. 4). The twentieth-century observed CRUNCEP precipitation trend is found to be positive and significant, consistent with the four precipitation datasets used in OL14 (Table 2), which use only actual observations (no spatial interpolation) at a coarser resolution and are not constrained to the availability of runoff observations. Trends in NHML land mean precipitation are robust to

the choice of dataset, some of which favor fewer long-term homogenized records over more short-term records, infilling of data through interpolation and spatial coverage. Interestingly, the runoff trend is positive and significant in all six models, but in each case less than the trend in observed CRUNCEP precipitation (Table 3). With the increase in precipitation not being matched by an increase in runoff, it would appear that the LSMs are all able to capture the already mentioned, well-documented increase in evapotranspiration (Milly and Dunne 2001; Walter et al. 2004).

Observed NHML land mean runoff shows a near-zero trend over the twentieth century (Table 2), in line with what is expected given mid-twentieth-century aerosol forcing. Annual mean runoff is strongly correlated with

TABLE 2. Trends in observed NHML land mean precipitation and runoff, with precipitation trends shown for the four datasets used in OL14 (observations only and calendar year), as well as the CRUNCEP dataset used in the current study (spatially interpolated and water year). Errors are the 5%–95% confidence intervals, with values in boldface significantly different from zero ($p < 0.05$).

Dataset	Reference	Trend (mm yr^{-1} century^{-1})
Precipitation (OL14)		
CRU	Harris et al. (2014)	**30.5 ± 9.1**
GPCC	Becker et al. (2013)	**28.8 ± 8.3**
GHCN	Vose et al. (1992)	**32.8 ± 9.6**
Zhang	Zhang et al. (2007)	**45.4 ± 10.8**
Precipitation		
CRUNCEP	Fisher et al. (2013)	**44.8 ± 12.6**
Runoff		
Dai	Dai et al. (2009)	3.2 ± 8.5

TABLE 3. Trends in NHML land mean runoff in the TRENDY LSMs and TRENDY multimodel mean, as well as the NHML land mean precipitation trend in the observed CRUNCEP dataset used to drive the LSMs. Errors are the 5%–95% confidence intervals, with values in boldface significantly different from zero ($p < 0.05$).

Model/dataset	Trend (mm yr^{-1} century^{-1})
CLM	**40.5** ± 7.5
JULES	**32.5** ± 4.5
LPJG	**37.4** ± 10.2
LPJ	**32.1** ± 9.1
LPX	**22.4** ± 8.1
OCN	**19.0** ± 6.1
TRENDY mean	**30.6** ± 7.2
CRUNCEP	**44.8** ± 12.6

annual mean precipitation ($r = 0.62$, $p < 0.01$). However, inspection of the NHML land mean runoff and precipitation time series appears to show a split in the relationship around the 1940s (Fig. 5). Simply adjusting the pre-1960 precipitation, as in OL14, to the pre-1960 runoff, using the linear regression fit of post-1960 five-year mean NHML land mean runoff against five-year NHML land mean precipitation, we find that the positive trend in precipitation is reduced and becomes nonsignificant. Arbitrarily selecting 1940 as a breakpoint in the runoff–precipitation relationship, we plot 5-yr mean NHML land mean runoff anomalies against 5-yr mean NHML land mean precipitation anomalies, split into a pre-1940 group and a post-1940 group (Fig. 6). On first impression there would appear to be a clear breakpoint, with two lines with similar slopes but different intercepts.

Determining whether a breakpoint is a real feature of this relationship requires a statistical test. Testing for breakpoints, also referred to as inhomogeneities (Peterson et al. 1998), change points (Reeves et al. 2007), and step changes (McCabe and Wolock 2002), in hydrometeorological variables is a long-established practice. Many different breakpoint detection methods are available and they vary in the number of breakpoints they can detect, the assumptions made on data distribution, and whether the timing of the breakpoint is known before testing [see Reeves et al. (2007) and Ferguson and Villarini (2012) for more in depth discussions]. Our approach differs from many of these studies in that we want to detect breakpoints in the relationship between two intrinsically linked hydrometeorological variables, rather than the time series of just one variable.

Therefore, we test for a breakpoint by fitting a model with four parameters, representing two lines with unique slopes and intercepts:

$$Q_i = \beta_0 + \beta_1 P_i + \beta_2 z_i + \beta_3 z_i P_i + \varepsilon_i, \qquad (4)$$

where Q are the 5-yr mean NHML land mean runoff anomalies, P are the 5-yr mean NHML land mean precipitation anomalies, and z is a binary variable coded 0 for data points in the first time segment and coded 1 for data points in the second time segment. The β coefficients are parameters to be estimated. The residuals ε are assumed to be identically and independently distributed with zero mean. By varying the length of the two time segments, it is possible to test for breakpoints at a number of 5-yr
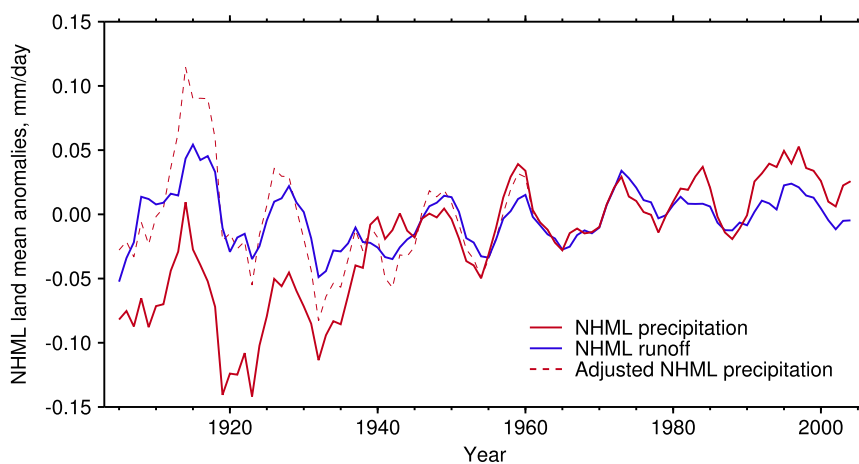

FIG. 5. The 5-yr running mean NHML land mean runoff and precipitation anomalies. Data are from the Dai et al. (2009) discharge dataset and CRUNCEP dataset for runoff and precipitation, respectively. The adjusted NHML land mean runoff time series is constructed using pre-1960 runoff values and the post-1960 five-year mean runoff–precipitation linear regression fit. All data are for the water year October–September during 1905–2004, and are anomalized with respect to the period 1961–90.
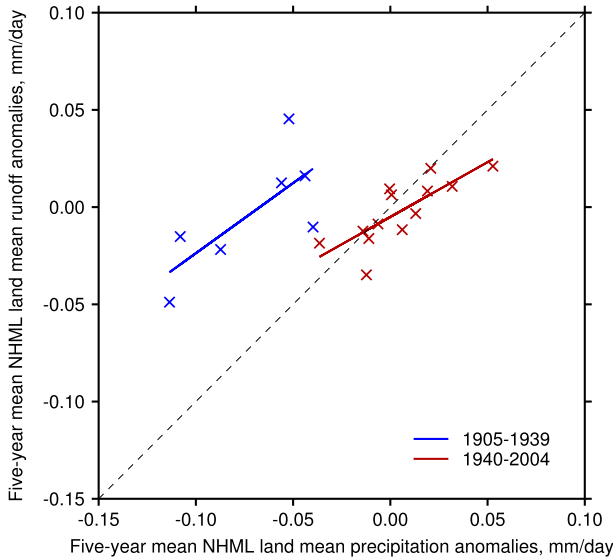
FIG. 6. The 5-yr mean NHML land mean runoff–precipitation relationship, split into a pre-1940 group and a post-1940 group.

intervals. The minimum length of each time segment, or group, is set at 25 years (a sample size of five). Therefore, we search for breakpoints between 1930 and 1980 inclusive. An analysis of variance (ANOVA) F test is used to determine whether Eq. (4), at any given division, offers a significantly better fit to the data than a simple one slope, one intercept model:

$$Q_i = \beta_0 + \beta_1 P_i + \varepsilon_i. \qquad (5)$$

However, because we are performing multiple tests—we perform 11 tests at the $p = 0.05$ level—there is a probability greater than 0.05 of finding at least one breakpoint. Because there is a probability greater than 0.05 that evidence against the null hypothesis [that Eq. (5) is the best fit] appears by chance, the significance level at which each of the 11 tests is performed requires adjustment to reduce the probability of a type I error back to 0.05.

To investigate this problem further, we simulate pseudo data, creating 10 000 datasets of runoff and precipitation anomalies. Random data for both precipitation and runoff are generated with a normal distribution with mean 0 and variance 1 (the variance is assumed to be constant between time segments). The probability of finding at least one breakpoint is 0.213. We find that performing each of the 11 tests at the $p = 0.0078$ level gives a probability of finding at least one breakpoint of 0.05. Therefore, if the F statistic at any of the 11 intervals is significant at the 0.78% level, we proceed with fitting the more complex model [Eq. (4)]. If β_3 in Eq. (4), known as the interaction parameter, is

TABLE 4. The estimated β coefficients from a model fit to a 1930 breakpoint for both Eqs. (4) and (6). Values in boldface are significant at the $p = 0.05$ level.

	β_0	β_1	β_2	β_3
Eq. (4)	**0.063**	**0.794**	**−0.066**	−0.364
Eq. (6)	**0.040**	**0.475**	**−0.044**	—

not significant ($p > 0.05$) we can assume $\beta_3 = 0$ and fit a more parsimonious, additive model

$$Q_i = \beta_0 + \beta_1 P_i + \beta_2 z_i + \varepsilon_i \qquad (6)$$

used for separate regression lines with the same slopes. However, we might expect some change in slope due to the more transient, indirect human influences discussed that could be prevalent at such large scales (e.g., Walter et al. 2004; Gedney et al. 2014). To avoid including the effect of slope changes in the breakpoint estimate, we take β_2 from Eq. (6) to be the breakpoint magnitude. The difference in β_2 values from Eqs. (4) and (6) is due to slope changes between the two time segments. This residual is typically small and our conclusions are not dependent on the choice of either Eq. (4) or (6). Although there could be more than one inhomogeneity in the observed precipitation and runoff records, the assumption is made that only one major breakpoint exists in the runoff–precipitation relationship. Further, this approach requires the assumption that any major inhomogeneity occurs on a time scale of years (Groisman and Rankova 2001) rather than decades. Where possible breakpoints are identified at more than one time division, the breakpoint is chosen as the model fit that minimizes the residual sum of squares (RSS).

A possible negative breakpoint in the observed NHML land mean runoff–precipitation relationship is identified at three time divisions—1930, 1935, and 1940—with the model fit to the 1930 breakpoint found to minimize the RSS (Table 4). Because this is the earliest interval at which a breakpoint can be detected using 5-yr means, the analysis is repeated using 3-yr means. Using 5-yr means is common when analyzing changes or trends in precipitation and in detection and attribution studies (Min et al. 2011; Wu et al. 2013; OL14) as it reduces the signal to noise and avoids problems with autocorrelation. The autocorrelation function for both the observed runoff and precipitation decays toward zero at a time lag of 3 yr (not shown). Repeating the analysis for 3-yr means—offering a detection range between 1920 and 1990—gives near-identical results with the breakpoint identified in 1932 (accounting for the multiple tests problem). This suggests the result is robust. Because of the highly variable nature of precipitation, at smaller
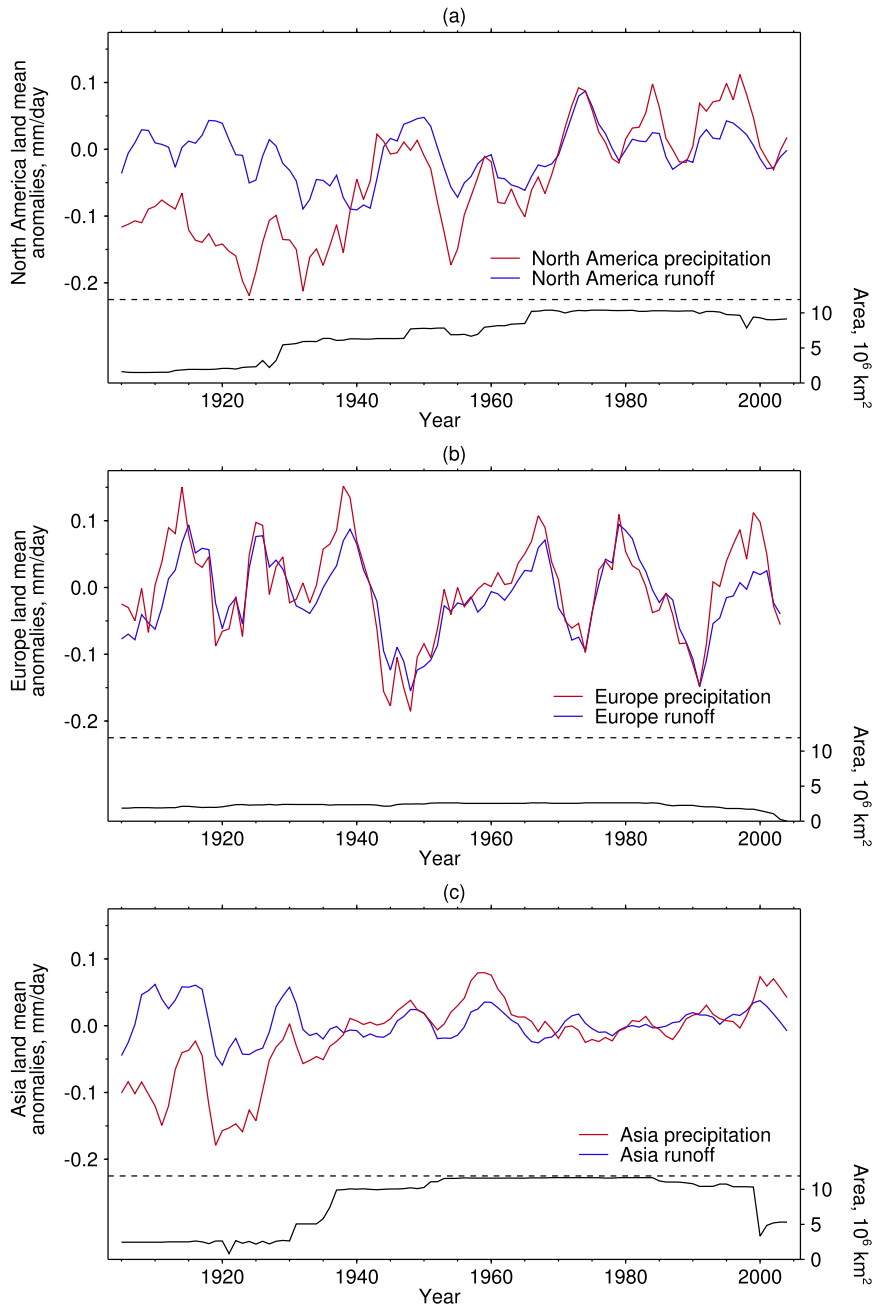
FIG. 7. The 5-yr running mean land mean runoff and precipitation anomalies and annual land area coverage of observed runoff: (a) North America, (b) Europe, and (c) Asia.

scales—continental scale and below—we only consider 5-yr means in further analyses to improve the signal to noise.

With a breakpoint detected, the approach is extended to determine where geographically the breakpoint occurs. While downscaling to the catchment scale is limited by temporal coverage and poor signal to noise, downscaling to the continental scale should offer more insight (Fig. 7). The continental-scale analysis points toward North America as the likely source of the breakpoint (Fig. 7a). A North American breakpoint is detected in 1935. The agreement between observed Europe land mean runoff and precipitation is excellent and the relationship is free of a breakpoint (Fig. 7b). The Asia land mean runoff and precipitation time series, which include all Russian river catchments (Fig. 7c),
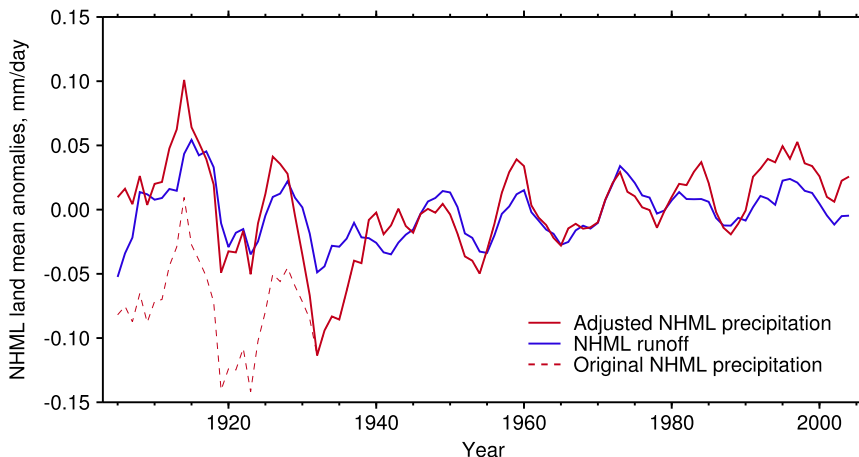
FIG. 8. The 5-yr running mean NHML land mean runoff and precipitation anomalies. This is an updated version of Fig. 5, showing the adjusted precipitation record, given a 1930 breakpoint in the runoff–precipitation relationship. The original precipitation record is also shown for reference.

suggest that a breakpoint might exist there in the first 30–40 years. Although our approach fails to detect one, reducing the minimum time segment length to 20 years, instead of 25 years, detects a breakpoint in 1925. The first ~25 years of the Asia time series are dominated by the Amur River, a catchment with particularly poor precipitation station spatial coverage at that time. This will undoubtedly contribute toward the 1930 breakpoint found in the NHML land mean runoff–precipitation relationship. We address the role of the Amur River and poor spatial coverage of the early-twentieth-century rain gauge network in section 4. Beyond about 1930, the agreement between precipitation and runoff in Asia is reasonable. However, we also note a contrast between decreasing precipitation and increasing runoff between the 1950s and 1980s, as found in other studies (Adam and Lettenmaier 2008; Milliman et al. 2008).

The breakpoint model is also applied to the TRENDY LSM NHML land mean runoff–precipitation relationships. Breakpoints are detected in two models, CLM4.5 and JULES, both in 1950. However, the breakpoints are of opposite sign (positive) to the observed breakpoint (negative) and are about half the size. Also, using the multimodel mean NHML land mean runoff no breakpoint is detected. The CLM4.5 and JULES simulations here have the lowest resolution of the six TRENDY models. Previous work has shown that a model performs much better in simulating the hydrological balance of various European regions with increased resolution (Elguindi et al. 2011). Investigating the differences between observed and simulated runoff spatially, the North American continent again stands out (not shown).

4. Adjusting observed precipitation

Having detected a 1930 breakpoint in the observed NHML land mean runoff–precipitation relationship, we now investigate the effect of adjusting the precipitation time series to bring it in line with runoff. This gives a hypothesized time series of early-twentieth-century precipitation based on the assumptions laid out in section 1. It is not to be interpreted as a true precipitation time series—the adjustment is only a crude estimate. However, it is instructive to see the extent to which the adjusted series is consistent with the expected precipitation response to aerosol forcing. The adjustment technique preserves the more subtle changes in the runoff–precipitation relationship, so that the slopes of the two lines in Eq. (4) are unchanged. Instead we simply increase pre-1930 precipitation so that $\beta_2 = 0$ in Eq. (6). The effect of this is to reduce the positive trend in twentieth-century precipitation from 44.8 ± 12.6 to 7.2 ± 13.6 mm yr^{-1} century^{-1} (Fig. 8). Reproducing Fig. 3 with adjusted observed precipitation, the observed NHML land mean precipitation offset of 0.002 ± 0.021 mm day^{-1} is equivalent to an aerosol forcing of -0.1 ± 2.0 W m^{-2}, calculated using the NHML land mean precipitation offset–aerosol forcing regression coefficients (Fig. 9). Prior to this precipitation adjustment, an offset of 0.044 ± 0.021 mm day^{-1} gave an estimate of aerosol forcing of 4.0 ± 2.0 W m^{-2}.

This adjusted precipitation gives an offset consistent with negative aerosol forcing. The difference between the best estimate of real-world NHML surface aerosol forcing of -2.0 (from -3.3 to -0.9) W m^{-2} and that suggested by the precipitation offset has reduced from
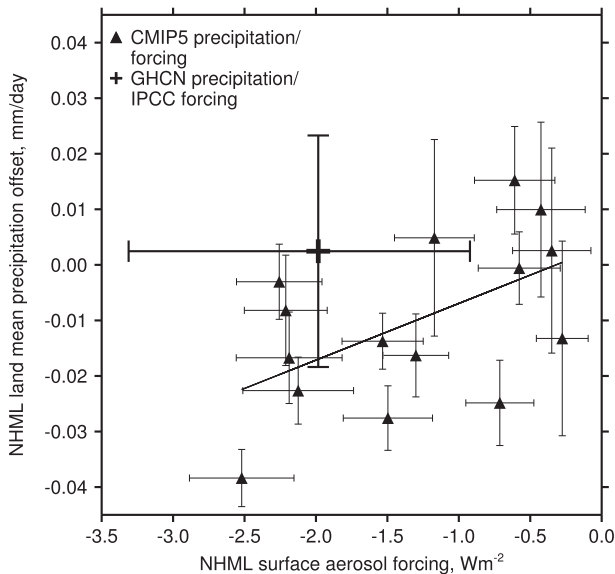
FIG. 9. NHML land mean precipitation offset against NHML surface aerosol forcing. This is an updated version of Fig. 3, but the adjusted precipitation record, given a 1930 breakpoint in the runoff–precipitation relationship, is used to calculate the observed NHML land mean precipitation offset.

6.0 to $1.9 \, \text{W m}^{-2}$. We find quantitatively similar results using the simple technique outlined in section 3 (Fig. 5), where we simply adjust the entire precipitation time series given the post-1960 runoff–precipitation relationship. This reduces the discrepancy from 6.0 to $1.7 \, \text{W m}^{-2}$. These results still somewhat contradict similar analyses performed for observed Northern Hemisphere tropical land mean precipitation and observed temperature gradient (between the NHML region and Southern Hemisphere extratropical region), which both show an aerosol response in agreement with expectations (see OL14).

5. Explaining the breakpoint

Meteorological datasets will still contain inherent inhomogeneities, owing to their changing spatial coverage in time, in addition to the various systematic errors listed above. These inhomogeneities and biases may be shared between different gridded precipitation datasets, as different groups use many common records. From Figs. 1, 2, and 7, it can be seen that the annual NHML land area coverage increases around the time that the breakpoint is found. This is when the three great Arctic Ocean draining Eurasian rivers (Ob, Yenisey, and Lena) and the Mississippi River begin contributing to NHML mean runoff. A near-constant spatial mask between 1905 and 2004 can be considered by selecting only river catchments with discharge records at least 80 years or longer (catchments with black borders in Fig. 1). Any

missing years for rivers with long discharge records are not noticeably clustered at the beginning of the twentieth century and are instead most obvious post-1990 (not shown). Because we are working with a spatially interpolated gridded precipitation dataset, the true spatial coverage of the underlying station network will still vary somewhat. This approach greatly reduces the spatial coverage but the breakpoint remains a robust feature, although 1935 is now the most likely timing (not shown). Changing spatial coverage in time could possibly lead to a breakpoint as different climatic regions are incorporated into the NHML mean. This is because a certain change in precipitation should not be expected to produce the same change in runoff in an arid region as in a humid region (Budyko 1974). However, the TRENDY LSMs realistically represent the range in the runoff coefficient across different climates, but do not contain breakpoints of the sign or magnitude of the observed breakpoint.

In the introduction, we dismissed the impact of dam and reservoir construction on the NHML mean runoff time series. At the catchment level, the influence of flow regulation (the ratio of total reservoir storage capacity to mean annual discharge) could be important. Indeed, Vörösmarty et al. (1997) showed that 20%–30% of global mean annual discharge was now stored in reservoirs constructed in the late twentieth century. Water impoundments can increase evaporative losses and therefore decrease runoff. Evaporative losses have been shown to have an effect on individual drainage basins (e.g., Vörösmarty and Moore 1991), but appear to be insignificant at continental scales (Vörösmarty et al. 1997; McClelland et al. 2004). Schwalm et al. (2014) recently found that including a representation of water management structures in terrestrial biosphere models made little difference to simulated runoff over the United States. Of course, the filling of a large reservoir following the construction of a dam causes an abrupt decrease in catchment runoff. This could lead to an anomalous annual mean runoff value. But following filling, streamflow should return to predam levels, so such impoundments are unlikely to cause breakpoints.

Dam and reservoir construction allows for easier access to water resources and typically leads to more water withdrawal, often for irrigation. Water withdrawn from reservoirs for irrigation can cause substantial decreases in runoff because of enhanced evapotranspiration. Irrigation is likely to account for a significant fraction of the stark runoff decreases observed in the Yellow River catchment in China since the 1960s (Xu et al. 2010). The Yellow River is one of the 148 included in our analyses. To determine what effect highly irrigated catchments have on our results, we take irrigation

index values [the ratio of area equipped for irrigation, from Siebert et al. (2005), to naturalized discharge (discharge prior to direct human influences)] from the Nilsson et al. (2005) study. Values are available for 42 out of 52 large rivers considered here with an upstream catchment area of greater than 50 000 km^2. These 42 large rivers account for 88% of the total area of all 148 river catchments. As found by Milliman et al. (2008), irrigation and flow regulation indices [the latter are also taken from the Nilsson et al. (2005) study] are strongly correlated (not shown). We find a total of eight rivers (Sacramento, Dnepr, Yellow, Don, Ebro, Colorado, Tagus, and Liao) that exceed both an irrigation index of 200 and a flow regulation of 20, defined by Milliman et al. (2008) as thresholds above which deficit rivers (rivers where changes in runoff are considerably less than changes in precipitation) typically fall. Removing these rivers from the study and considering only the remaining 34 does not change our conclusions. We still find a significant breakpoint, although a 1935 breakpoint provides a marginally better fit to the data than a 1930 breakpoint.

A subject worthy of consideration is the quality of the interpolated CRUNCEP precipitation observations. While complete spatial coverage is desirable, it is important to consider the true spatial coverage of the underlying station network. The area represented per station within a catchment, defined as the area per station (APS), is much greater across the Asian continent than the North American and European continents (Table 5), particularly so in the 1905–29 period. However, as already discussed, this is also a time period when spatial coverage of runoff observations across the Asian continent was largely limited to the Amur River, the fifth-largest river of the 148 in terms of catchment area and ranking fifth in terms of 1905–2004 mean annual streamflow. There are no available streamflow observations for the three great Arctic Ocean–draining Eurasian rivers (Ob, Yenisey, and Lena) between 1905 and 1929, when their APS values were extremely high. As such, Asia mean APS values are noticeably different in this period depending on whether a fixed temporal mask of all rivers is used or whether precipitation is masked to the availability of runoff observations. When masked to the availability of runoff observations, the Amur catchment has the worst mean annual APS of all rivers. The poor APS of the Amur River means that Asia land mean precipitation is likely to be more vulnerable to inhomogeneities in the pre-1930 period.

Although the APS might explain the Asian continent contribution toward the 1930 NHML land mean runoff–precipitation breakpoint, it has less merit in explaining the more prominent 1935 North American breakpoint.

TABLE 5. The APS of the North American, European, and Asian continents for the 1905–29 and 1930–2004 periods (before and after the NHML land mean runoff–precipitation breakpoint is detected). Values (km^2) are shown for the variable mask, where precipitation is masked to the availability of runoff observations (the spatial mask varies in time), and the fixed mask, where precipitation is masked to the catchments of all 148 rivers (the spatial mask is constant in time).

	North America	Europe	Asia
Variable mask			
1905–29	14 866	31 677	155 338
1930–2004	18 474	16 384	36 038
Fixed mask			
1905–29	24 107	27 918	245 153
1930–2004	20 790	17 455	42 151

Further, the North American and European APS values are comparable, yet there is good agreement between observed precipitation and runoff for Europe. Up until 1935, the St. Lawrence and Columbia River catchments account for much of the spatial coverage of North American runoff observations. The St. Lawrence and Columbia are the fourth- and fifth-largest river catchments of the 72 in North America, respectively, with the former draining into the Atlantic and the latter draining into the Pacific. Combined aggregated Pacific and Atlantic draining river catchments contain a runoff–precipitation breakpoint in 1935, although the β_1 and β_2 coefficients in Eq. (4) are not significant ($p = 0.25$ and $p = 0.29$, respectively). No breakpoint is found when considering the aggregated Gulf of Mexico draining rivers (dominated by the Mississippi catchment). While attributing the exact location and cause behind a breakpoint in the runoff–precipitation relationship is beyond the scope of this study, this does help narrow the focus of future work into precipitation inhomogeneities.

The 1930s saw a notable hydroclimatic event across North America in the form of the Dust Bowl, a period characterized by low precipitation amounts and high temperatures (e.g., Schubert et al. 2004). This precipitation deficit period is seemingly captured by the North America precipitation time series (Fig. 7a) and in a hypothetical time series of observed North America land mean precipitation (following the methodology of section 4; not shown). The LSMs—driven by observed precipitation—do not appear to simulate this breakpoint. A breakpoint is detected in one model (LPX) in 1940, which is of the same sign as the observed breakpoint but roughly half the magnitude. Interestingly, this is no longer detected when considering only river catchments with discharge records at least 80 years or longer (as above), implying some sensitivity to spatial coverage. Overall, the LSMs do not support a real

change in the runoff–precipitation relationship over the North American continent around this time.

6. Summary and conclusions

In OL14, observed NHML land mean precipitation was found to be missing an expected response (Wu et al. 2013) to well-documented mid-twentieth-century aerosol forcing (Lamarque et al. 2010; Wild 2012). This is despite a robust signal across CMIP5 GCMs, with the magnitude of the response dependent on the strength of aerosol forcing. It was suggested that inhomogeneities in the precipitation record could be part of the reason for this disagreement, with some studies questioning the quality of precipitation observations, particularly in the first half of the twentieth century (Groisman and Rankova 2001). Here, we have attempted to reconcile these precipitation observations with complementary runoff observations, which represent a key land surface component of the hydrological cycle.

Variations in interannual, decadal, and multidecadal runoff are known to be predominantly forced by variations in precipitation (Dai et al. 2009). Occasionally, long-term trends in runoff are shown to differ from long-term trends in precipitation (Milliman et al. 2008). This is particularly prevalent at the catchment scale, where direct human influences, such as water withdrawals for irrigation (Milliman et al. 2008; Xu et al. 2010) and land-cover or land-use change (Piao et al. 2007; Sterling et al. 2013) can have significant impacts. At larger scales, differences in runoff and precipitation trends can be attributed to changes in evapotranspiration (Milly and Dunne 2001; Walter et al. 2004), perhaps because of increasing temperatures due to rising greenhouse gas concentrations (Huntington 2008; Krakauer and Fung 2008) and augmented, at least locally, by aerosol-induced dimming (Gedney et al. 2014). Consideration of the plant physiological response to rising CO_2 levels is also necessary (Gedney et al. 2006). However, while these influences may have an effect on the runoff–precipitation relationship in the NHML land mean, they are not expected to cause breakpoints.

Testing for breakpoints in the NHML land mean runoff–precipitation relationship, three potential candidates at 5-yr intervals were found (1930, 1935, and 1940), with a negative breakpoint in 1930 found to provide the best fit to the data. Runoff anomalies pre-1940 are noticeably greater than precipitation anomalies. This lends support to the hypothesis suggested by OL14 that pre-1960 precipitation anomalies should be greater. They found that an increase of 2.9% in monthly total precipitation before 1960 in the CRU TS3.21 dataset would generate a negative NHML land mean

precipitation offset consistent with estimated aerosol forcing. Our results do not support an increase in precipitation totals across the entire pre-1960 period or the magnitude of the increase suggested. However, adjusting the precipitation record before the 1930 breakpoint using the runoff record, we do account for some of the discrepancy between the (lack of) observed precipitation aerosol response and that suggested given our most recent estimates of real-world aerosol forcing (Myhre et al. 2013). Importantly, this adjustment does allow the precipitation aerosol offset to be negative and means that the range of aerosol forcing suggested by this response is within the real-world aerosol forcing range.

Using six LSMs as part of the TRENDY intercomparison project (Sitch et al. 2015), it is possible to show that this breakpoint is unlikely to be a consequence of known physical processes. Driven by observed precipitation, four of six models show no breakpoint, whereas two show small, positive breakpoints in 1950. Although not nearly as apparent as the breakpoint in the observed runoff–precipitation relationship, and perhaps just a consequence of the choice of statistical model, this warrants further attention. However, we conclude that the modeled runoff–precipitation relationship is free of breakpoints of the type and magnitude found in observations. Perhaps the most obvious shortcoming in this technique of using observed precipitation to drive LSMs is that land surface feedbacks on the atmosphere are neglected, because of the absence of coupling (Harding et al. 2014; Schewe et al. 2014), limiting the potential accuracy of results. Future efforts to incorporate more complex land surface schemes into GCMs should improve our understanding of the importance of such interactions. The LSMs used here also do not include a representation of direct human influences, such as irrigation, with the exception of historical land-use changes, although we argue in section 1 that these influences are unlikely to cause breakpoints at continental scales.

A continental-scale search for breakpoints found that the North American continent was the likely source for the observed NHML-wide breakpoint. Changes in measurement practice and gauge type are just two of many explanations for such an inhomogeneity in precipitation records (Groisman and Rankova 2001). We have assumed that runoff records are less susceptible to inhomogeneities (Hagemann and Dümenil 1997), but some will undoubtedly exist and should also be investigated in much the same way as precipitation records. In addition, it would be unwise to wholly discount the influence of dams, with dam construction in the Columbia River catchment, for example, beginning in the 1930s, notably with the Bonneville Dam. Although most work suggests that the influence of dams on annual

mean discharge at large scales is minimal (Vörösmarty et al. 1997; McClelland et al. 2004), the St. Lawrence and Columbia River catchments deserve particular attention as their runoff and precipitation observations contribute significantly to the pre-1935 North American land means. Although the LSMs do not simulate this North American breakpoint, there remains the possibility that the unique hydroclimatological precursors that lead to the 1930s Dust Bowl event are not well captured by other meteorological variables used to drive the models.

Further, we stress some shortcomings in our methodology. First, not all precipitation inhomogeneities can be expected to appear as an obvious discontinuity, with some biases likely to emerge slowly over time, perhaps because of changes in vegetation near to a rain gauge (Legates 1995). Also, we have used just one statistical model to test for breakpoints. It would be useful to see how robust our findings are to different techniques, while also considering the runoff and precipitation records independently. Problems associated with using spatially interpolated data have also been highlighted. The Amur River catchment was identified as a region with a sparse precipitation station network in the early twentieth century. Therefore, a direct comparison between runoff and precipitation here would contrast the complete spatial coverage of runoff (inherent in streamflow measurements) with just a few point measurements of precipitation.

We are also unable to provide a robust explanation for a 1930s breakpoint. One of the more widely documented systematic changes in precipitation measurement occurred in the former USSR. In 1966–67 a wetting correction was introduced that added 5%–15% to annual totals across the former USSR. In the far north of Russia in winter months, totals increased by up to 30% (Groisman et al. 1991; Groisman and Rankova 2001), but no such systematic changes are documented in the 1930s. Instead, it might be the case that the spatial coverage of early-twentieth-century observations, combined with the underrepresentation of, say, mountainous regions, leaves NHML land mean precipitation observations prone to inhomogeneities during this period. Without targeted studies, however, this remains speculative, but we believe we have here eliminated some possibilities behind such a breakpoint (section 5), including the influence of dams and reservoirs, as well as irrigation. This is to be expected since the 1930s preceded major direct human influences on the land surface component of the hydrological cycle.

Although changes in runoff closely track changes in precipitation, the relationship between these two variables is complicated. For example, recent research has shown that a precipitation shift from snow to rain may lead to a decrease in runoff (Berghuijs et al. 2014), and

this may play a role in explaining the steady evapotranspiration increases seen over the NHML region and particularly the North American continent. Regardless, by incorporating runoff into the precipitation–aerosol physical framework of OL14, we have found more evidence suggesting NHML land mean precipitation observations to be in error. Early-twentieth-century precipitation observations typically do not lend themselves well to robust climate change research, although the European continent is a notable exception. Here, the runoff–precipitation agreement is excellent. But even after incorporating runoff into this physical framework, the discrepancy between observed and modeled precipitation over the NHML land region cannot be wholly accounted for. So, while inhomogeneities are likely to exist in precipitation records, we conclude that model-simulated precipitation response to aerosol forcing should still be questioned.

REFERENCES

Adam, J. C., and D. P. Lettenmaier, 2003: Adjustment of global gridded precipitation for systematic bias. *J. Geophys. Res.*, **108**, 4257, doi:10.1029/2002JD002499.

——, and ——, 2008: Application of new precipitation and reconstructed streamflow products to streamflow trend attribution in Northern Eurasia. *J. Climate*, **21**, 1807–1828, doi:10.1175/2007JCLI1535.1.

——, E. A. Clark, D. P. Lettenmaier, and E. F. Wood, 2006: Correction of global precipitation products for orographic effects. *J. Climate*, **19**, 15–38, doi:10.1175/JCLI3604.1.

——, I. Haddeland, F. Su, and D. P. Lettenmaier, 2007: Simulation of reservoir influences on annual and seasonal streamflow changes for the Lena, Yenisei, and Ob' rivers. *J. Geophys. Res.*, **112**, D24114, doi:10.1029/2007JD008525.

Ahlström, A., G. Schurgers, A. Arneth, and B. Smith, 2012: Robustness and uncertainty in terrestrial ecosystem carbon response to CMIP5 climate change projections. *Environ. Res. Lett.*, **7**, 044008, doi:10.1088/1748-9326/7/4/044008.

Becker, A., P. Finger, A. Meyer-Christoffer, B. Rudolf, K. Schamm, U. Schneider, and M. Ziese, 2013: A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901–present. *Earth Syst. Sci. Data*, **5**, 71–99, doi:10.5194/essd-5-71-2013.

Berghuijs, W. R., R. A. Woods, and M. Hrachowitz, 2014: A precipitation shift from snow towards rain leads to a decrease in streamflow. *Nat. Climate Change*, **4**, 583–586, doi:10.1038/nclimate2246.

Best, M. J., and Coauthors, 2011: The Joint UK Land Environment Simulator (JULES), model description—Part 1: Energy and water fluxes. *Geosci. Model Dev.*, **4**, 677–699, doi:10.5194/gmd-4-677-2011.

Bindoff, N. L., and Coauthors, 2013: Detection and attribution of climate change: From global to regional. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 867–952.

Boucher, O., and Coauthors, 2013: Clouds and aerosols. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 571–658.

Budyko, M. I., 1974: *Climate and Life*. Academic Press, 508 pp.

Charlson, R. J., S. E. Schwartz, J. M. Hales, R. D. Cess, J. A. Coakley, J. E. Hansen, and D. J. Hofmann, 1992: Climate forcing by anthropogenic aerosols. *Science*, **255**, 423–430, doi:10.1126/science.255.5043.423.

Clark, D. B., and Coauthors, 2011: The Joint UK Land Environment Simulator (JULES), model description—Part 2: Carbon fluxes and vegetation dynamics. *Geosci. Model Dev.*, **4**, 701–722, doi:10.5194/gmd-4-701-2011.

Dai, A., T. Qian, K. E. Trenberth, and J. D. Milliman, 2009: Changes in continental freshwater discharge from 1948 to 2004. *J. Climate*, **22**, 2773–2792, doi:10.1175/2008JCLI2592.1.

Elguindi, N., S. Somot, M. Déqué, and W. Ludwig, 2011: Climate change evolution of the hydrological balance of the Mediterranean, Black and Caspian Seas: Impact of climate model resolution. *Climate Dyn.*, **36**, 205–228, doi:10.1007/s00382-009-0715-4.

Falloon, P., R. Betts, A. Wiltshire, R. Dankers, C. Mathison, D. McNeall, P. Bates, and M. Trigg, 2011: Validation of river flows in HadGEM1 and HadCM3 with the TRIP river flow model. *J. Hydrometeor.*, **12**, 1157–1180, doi:10.1175/2011JHM1388.1.

Fekete, B. M., C. J. Vörösmarty, and W. Grabs, 2002: High-resolution fields of global runoff combining observed river discharge and simulated water balances. *Global Biogeochem. Cycles*, **16**, 1042, doi:10.1029/1999GB001254.

Ferguson, C. R., and G. Villarini, 2012: Detecting inhomogeneities in the twentieth century reanalysis over the central United States. *J. Geophys. Res.*, **117**, D05123, doi:10.1029/2011JD016988.

Fisher, J. B., and Coauthors, 2013: African tropical rainforest net carbon dioxide fluxes in the twentieth century. *Philos. Trans. Roy. Soc. London*, **B368**, 20120376, doi:10.1098/rstb.2012.0376.

Forster, P. M., T. Andrews, P. Good, J. M. Gregory, L. S. Jackson, and M. Zelinka, 2013: Evaluating adjusted forcing and model spread for historical and future scenarios in the CMIP5 generation of climate models. *J. Geophys. Res. Atmos.*, **118**, 1139–1150, doi:10.1002/jgrd.50174.

Gedney, N., P. M. Cox, R. A. Betts, O. Boucher, C. Huntingford, and P. A. Stott, 2006: Detection of a direct carbon dioxide effect in continental river runoff records. *Nature*, **439**, 835–838, doi:10.1038/nature04504.

——, C. Huntingford, G. P. Weedon, N. Bellouin, O. Boucher, and P. M. Cox, 2014: Detection of solar dimming and brightening effects on Northern Hemisphere river flow. *Nat. Geosci.*, **7**, 796–800, doi:10.1038/ngeo2263.

Gerten, D., S. Rost, W. von Bloh, and W. Lucht, 2008: Causes of change in 20th century global river discharge. *Geophys. Res. Lett.*, **35**, L20405, doi:10.1029/2008GL035258.

Gordon, L. J., W. Steffen, B. F. Jönsson, C. Folke, M. Falkenmark, and Å. Johannessen, 2005: Human modification of global water vapor flows from the land surface. *Proc. Natl. Acad. Sci. USA*, **102**, 7612–7617, doi:10.1073/pnas.0500208102.

Groisman, P. Ya., and E. Y. Rankova, 2001: Precipitation trends over the Russian permafrost-free zone: Removing the artifacts of pre-processing. *Int. J. Climatol.*, **21**, 657–678, doi:10.1002/joc.627.

Groisman, P. Y., V. V. Koknaeva, T. A. Belokrylova, and T. R. Karl, 1991: Overcoming biases of precipitation measurement: A history of the USSR experience. *Bull. Amer. Meteor. Soc.*, **72**, 1725–1733, doi:10.1175/1520-0477(1991)072<1725:OBOPMA>2.0.CO;2.

Hagemann, S., and L. Dümenil, 1997: A parametrization of the lateral waterflow for the global scale. *Climate Dyn.*, **14**, 17–31, doi:10.1007/s003820050205.

Harding, R. J., G. P. Weedon, H. A. van Lanen, and D. B. Clark, 2014: The future for global water assessment. *J. Hydrol.*, **518**, 186–193, doi:10.1016/j.jhydrol.2014.05.014.

Harris, I., P. Jones, T. Osborn, and D. Lister, 2014: Updated high-resolution grids of monthly climatic observations—The CRU TS3.10 dataset. *Int. J. Climatol.*, **34**, 623–642, doi:10.1002/joc.3711.

Hirabayashi, Y., R. Mahendran, S. Koirala, L. Konoshima, D. Yamazaki, S. Watanabe, H. Kim, and S. Kanae, 2013: Global flood risk under climate change. *Nat. Climate Change*, **3**, 816–821, doi:10.1038/nclimate1911.

Huntington, T. G., 2008: CO_2-induced suppression of transpiration cannot explain increasing runoff. *Hydrol. Processes*, **22**, 311–314, doi:10.1002/hyp.6925.

Hwang, Y.-T., D. M. W. Frierson, and S. M. Kang, 2013: Anthropogenic sulfate aerosol and the southward shift of tropical precipitation in the late 20th century. *Geophys. Res. Lett.*, **40**, 2845–2850, doi:10.1002/grl.50502.

Jung, I. W., H. Chang, and J. Risley, 2013: Effects of runoff sensitivity and catchment characteristics on regional actual evapotranspiration trends in the conterminous US. *Environ. Res. Lett.*, **8**, 044002, doi:10.1088/1748-9326/8/4/044002.

Jung, M., and Coauthors, 2010: Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature*, **467**, 951–954, doi:10.1038/nature09396.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471, doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Klein Goldewijk, K., and P. H. Verburg, 2013: Uncertainties in global-scale reconstructions of historical land use: An illustration using the HYDE data set. *Landscape Ecol.*, **28**, 861–877, doi:10.1007/s10980-013-9877-x.

Koster, R. D., and Coauthors, 2004: Regions of strong coupling between soil moisture and precipitation. *Science*, **305**, 1138–1140, doi:10.1126/science.1100217.

Krakauer, N. Y., and I. Fung, 2008: Mapping and attribution of change in streamflow in the conterminous United States. *Hydrol. Earth Syst. Sci.*, **12**, 1111–1120, doi:10.5194/hess-12-1111-2008.

Labat, D., Y. Goddéris, J. L. Probst, and J. L. Guyot, 2004: Evidence for global runoff increase related to climate warming. *Adv. Water Resour.*, **27**, 631–642, doi:10.1016/j.advwatres.2004.02.020.

Lamarque, J.-F., and Coauthors, 2010: Historical (1850–2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: Methodology and application. *Atmos. Chem. Phys.*, **10**, 7017–7039, doi:10.5194/acp-10-7017-2010.

Legates, D. R., 1995: Global and terrestrial precipitation: A comparative assessment of existing climatologies. *Int. J. Climatol.*, **15**, 237–258, doi:10.1002/joc.3370150302.

Le Quéré, C., and Coauthors, 2014: Global carbon budget 2013. *Earth Syst. Sci. Data*, **6**, 235–263, doi:10.5194/essd-6-235-2014.

McCabe, G. J., and D. M. Wolock, 2002: A step increase in streamflow in the conterminous United States. *Geophys. Res. Lett.*, **29**, 2185, doi:10.1029/2002GL015999.

McClelland, J. W., R. M. Holmes, B. J. Peterson, and M. Stieglitz, 2004: Increasing river discharge in the Eurasian Arctic: Consideration of dams, permafrost thaw, and fires as potential agents of change. *J. Geophys. Res.*, **109**, D18102, doi:10.1029/2004JD004583.

McMillan, H., T. Krueger, and J. Freer, 2012: Benchmarking observational uncertainties for hydrology: Rainfall, river discharge and water quality. *Hydrol. Processes*, **26**, 4078–4111, doi:10.1002/hyp.9384.

Milliman, J., K. Farnsworth, P. Jones, K. Xu, and L. Smith, 2008: Climatic and anthropogenic factors affecting river discharge to the global ocean, 1951–2000. *Global Planet. Change*, **62**, 187–194, doi:10.1016/j.gloplacha.2008.03.001.

Milly, P. C. D., and K. A. Dunne, 2001: Trends in evaporation and surface cooling in the Mississippi River basin. *Geophys. Res. Lett.*, **28**, 1219–1222, doi:10.1029/2000GL012321.

——, ——, and A. V. Vecchia, 2005: Global pattern of trends in streamflow and water availability in a changing climate. *Nature*, **438**, 347–350, doi:10.1038/nature04312.

Min, S.-K., X. Zhang, F. W. Zwiers, and G. C. Hegerl, 2011: Human contribution to more-intense precipitation extremes. *Nature*, **470**, 378–381, doi:10.1038/nature09763.

Myhre, G., and Coauthors, 2013: Anthropogenic and natural radiative forcing. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 659–740.

New, M., M. Hulme, and P. Jones, 2000: Representing twentieth-century space–time climate variability. Part II: Development of 1901–96 monthly grids of terrestrial surface climate. *J. Climate*, **13**, 2217–2238, doi:10.1175/1520-0442(2000)013<2217:RTCSTC>2.0.CO;2.

Nilsson, C., C. A. Reidy, M. Dynesius, and C. Revenga, 2005: Fragmentation and flow regulation of the world's large river systems. *Science*, **308**, 405–408, doi:10.1126/science.1107887.

Oleson, K. W., and Coauthors, 2013: Technical description of version 4.5 of the Community Land Model (CLM). NCAR Tech. Note NCAR/TN-503+STR, 420 pp., doi:10.5065/D6RR1W7M.

Oliveira, P. J. C., E. L. Davin, S. Levis, and S. I. Seneviratne, 2011: Vegetation-mediated impacts of trends in global radiation on land hydrology: A global sensitivity study. *Global Change Biol.*, **17**, 3453–3467, doi:10.1111/j.1365-2486.2011.02506.x.

Osborne, J. M., and F. H. Lambert, 2014: The missing aerosol response in twentieth-century mid-latitude precipitation observations. *Nat. Climate Change*, **4**, 374–378, doi:10.1038/nclimate2173.

Peterson, T. C., and Coauthors, 1998: Homogeneity adjustments of in situ atmospheric climate data: A review. *Int. J. Climatol.*, **18**, 1493–1517, doi:10.1002/(SICI)1097-0088(19981115)18:13<1493::AID-JOC329>3.0.CO;2-T.

Piao, S., P. Friedlingstein, P. Ciais, N. de Noblet-Ducoudré, D. Labat, and S. Zaehle, 2007: Changes in climate and land use

have a larger direct impact than rising CO_2 on global river runoff trends. *Proc. Natl. Acad. Sci. USA*, **104**, 15 242–15 247, doi:10.1073/pnas.0707213104.

Polson, D., G. C. Hegerl, X. Zhang, and T. J. Osborn, 2013: Causes of robust seasonal land precipitation changes. *J. Climate*, **26**, 6679–6697, doi:10.1175/JCLI-D-12-00474.1.

Reeves, J., J. Chen, X. L. Wang, R. Lund, and Q. Q. Lu, 2007: A review and comparison of changepoint detection techniques for climate data. *J. Appl. Meteor. Climatol.*, **46**, 900–915, doi:10.1175/JAM2493.1.

Schewe, J., and Coauthors, 2014: Multimodel assessment of water scarcity under climate change. *Proc. Natl. Acad. Sci. USA*, **111**, 3245–3250, doi:10.1073/pnas.1222460110.

Schubert, S. D., M. J. Suarez, P. J. Pegion, R. D. Koster, and J. T. Bacmeister, 2004: On the cause of the 1930s Dust Bowl. *Science*, **303**, 1855–1859, doi:10.1126/science.1095048.

Schwalm, C. R., and Coauthors, 2014: A model–data intercomparison of simulated runoff in the contiguous United States: Results from the North America Carbon Regional and Continental Interim-Synthesis. *Biogeosci. Discuss.*, **11**, 1801–1826, doi:10.5194/bgd-11-1801-2014.

Shindell, D. T., 2014: Inhomogeneous forcing and transient climate sensitivity. *Nat. Climate Change*, **4**, 274–277, doi:10.1038/nclimate2136.

Siebert, S., P. Döll, J. Hoogeveen, J.-M. Faures, K. Frenken, and S. Feick, 2005: Development and validation of the global map of irrigation areas. *Hydrol. Earth Syst. Sci.*, **9**, 535–547, doi:10.5194/hess-9-535-2005.

Sitch, S., and Coauthors, 2003: Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. *Global Change Biol.*, **9**, 161–185, doi:10.1046/j.1365-2486.2003.00569.x.

——, and Coauthors, 2015: Recent trends and drivers of regional sources and sinks of carbon dioxide. *Biogeosciences*, **12**, 653–679, doi:10.5194/bg-12-653-2015.

Smith, B., I. C. Prentice, and M. T. Sykes, 2001: Representation of vegetation dynamics in the modelling of terrestrial ecosystems: Comparing two contrasting approaches within European climate space. *Global Ecol. Biogeogr.*, **10**, 621–637, doi:10.1046/j.1466-822X.2001.00256.x.

Sperna Weiland, F. C., L. P. H. van Beek, J. C. J. Kwadijk, and M. F. P. Bierkens, 2012: Global patterns of change in discharge regimes for 2100. *Hydrol. Earth Syst. Sci.*, **16**, 1047–1062, doi:10.5194/hess-16-1047-2012.

Sterling, S. M., A. Ducharne, and J. Polcher, 2013: The impact of global land-cover change on the terrestrial water cycle. *Nat. Climate Change*, **3**, 385–390, doi:10.1038/nclimate1690.

Stocker, B. D., and Coauthors, 2013: Multiple greenhouse-gas feedbacks from the land biosphere under future climate change scenarios. *Nat. Climate Change*, **3**, 666–672, doi:10.1038/nclimate1864.

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, doi:10.1175/BAMS-D-11-00094.1.

Vörösmarty, C. J., and B. Moore III, 1991: Modeling basin-scale hydrology in support of physical climate and global biogeochemical studies: An example using the Zambezi River. *Surv. Geophys.*, **12**, 271–311, doi:10.1007/BF01903422.

——, K. Sharma, B. M. Fekete, A. H. Copeland, J. Holden, J. Marble, and J. A. Lough, 1997: The storage and aging of continental runoff in large reservoir systems of the world. *Ambio*, **26**, 210–219.

Vose, R. S., R. L. Schmoyer, P. M. Steurer, T. C. Peterson, R. Heim, T. R. Karl, and J. K. Eischeid, 1992: The Global Historical Climatology Network: Long-term monthly temperature, precipitation, sea level pressure, and station pressure data. Environmental Sciences Division Publ. 3912, ORNL/CDIAC-53 NDP-041, Oak Ridge National Laboratory, 324 pp., doi:10.3334/CDIAC/cli.ndp041.

Walter, M. T., D. S. Wilks, J.-Y. Parlange, and R. L. Schneider, 2004: Increasing evapotranspiration from the conterminous United States. *J. Hydrometeor.*, **5**, 405–408, doi:10.1175/1525-7541(2004)005<0405:IEFTCU>2.0.CO;2.

Wild, M., 2012: Enlightening global dimming and brightening. *Bull. Amer. Meteor. Soc.*, **93**, 27–37, doi:10.1175/BAMS-D-11-00074.1.

Wu, P., N. Christidis, and P. Stott, 2013: Anthropogenic impact on Earth's hydrological cycle. *Nat. Climate Change*, **3**, 807–810, doi:10.1038/nclimate1932.

Xu, K., J. D. Milliman, and H. Xu, 2010: Temporal trend of precipitation and runoff in major Chinese rivers since 1951. *Global Planet. Change*, **73**, 219–232, doi:10.1016/j.gloplacha.2010.07.002.

Yang, D., D. Kane, Z. Zhang, D. Legates, and B. Goodison, 2005: Bias corrections of long-term (1973–2004) daily precipitation data over the northern regions. *Geophys. Res. Lett.*, **32**, L19501, doi:10.1029/2005GL024057.

Yang, H., and Coauthors, 2015: Multi-criteria evaluation of discharge simulation in Dynamic Global Vegetation Models. *J. Geophys. Res. Atmos.*, **120**, 7488–7505, doi:10.1002/2015JD023129.

Zaehle, S., and A. D. Friend, 2010: Carbon and nitrogen cycle dynamics in the O-CN land surface model: 1. Model description, site-scale evaluation, and sensitivity to parameter estimates. *Global Biogeochem. Cycles*, **24**, GB1005, doi:10.1029/2009GB003521.

——, ——, P. Friedlingstein, F. Dentener, P. Peylin, and M. Schulz, 2010: Carbon and nitrogen cycle dynamics in the O-CN land surface model: 2. Role of the nitrogen cycle in the historical terrestrial carbon balance. *Global Biogeochem. Cycles*, **24**, GB1006, doi:10.1029/2009GB003522.

Zhang, X., F. W. Zwiers, G. C. Hegerl, F. H. Lambert, N. P. Gillett, S. Solomon, P. A. Stott, and T. Nozawa, 2007: Detection of human influence on twentieth-century precipitation trends. *Nature*, **448**, 461–465, doi:10.1038/nature06025.