Climate and hydrologic ensembling lead to differing streamflow and sediment yield predictions

Travis A. Dahl [1,2] ⓘ · Anthony D. Kendall [2] ⓘ · David W. Hyndman [2,3] ⓘ

Abstract

Climate change is leading to alterations of the hydrologic cycle and sediment movement within watersheds, but the details and impacts of these changes are indeterminate. To reduce this uncertainty, many researchers create ensembles by averaging the projected temperature and precipitation from multiple global climate model (GCM) ensemble members before running these as forcing inputs through hydrologic models. There is little research quantifying if these ensembled climate scenarios produce similar hydrologic model results to those based on individual ensemble members. We created multiple sets of ensembled climate inputs for a pair of hydrologic and sediment yield models of adjacent watersheds that drain to the Great Lakes. We then compared the hydrologic and sediment results of the models forced by these ensembled climate scenarios with hydrologic ensembles created by running the individual climate ensemble members through the same hydrologic models. We found that, in all cases, the streamflow and sediment yield results are significantly different at the 5% confidence level and the ensembled climate scenarios can lead to systematic negative biases. We also looked at three subset hydrologic ensembles: all 10 CMIP5 ensemble members from the CSIRO mk3.6 model; a Representative ensemble with high, moderate, and low precipitation predictions; and a Best Fit ensemble based on GCM performance relative to historic climate. We found that the subset ensembles covered a large portion of the range of outputs for the whole set, while producing mean annual streamflows within 5.5% of the full hydrologic ensemble results and sediment yield and sediment discharge results within 12.2%.

✉ Travis A. Dahl
Travis.A.Dahl@usace.army.mil

[1] U.S. Army Engineer Research and Development Center, Coastal and Hydraulics Laboratory, Vicksburg, MS 39180, USA

[2] Department of Earth and Environmental Sciences, Michigan State University, East Lansing, MI 48824, USA

[3] Department of Geosciences, School of Natural Sciences and Mathematics, The University of Texas at Dallas, Richardson, TX 75080, USA

1 Introduction

Changes in climate, particularly precipitation and temperature, can cause a wide range of impacts to our environment. Streamflow and sediment movement are particularly susceptible to changes in both magnitude and timing of precipitation. They are also affected by differences in temperature, although less directly. These changes to streamflow and sediment movement through a watershed in turn impact infrastructure, agriculture, and ecosystems.

Many researchers have looked at the potential impacts of climate change on streamflow and sediment, but most limit their modeling efforts to a handful of climate scenarios (e.g., Cherkauer and Sinha 2010; Johnson et al. 2015; O'Neal et al. 2005; Park et al. 2011; Serpa et al. 2015; Verma et al. 2015). This can be problematic because research has shown there can be a significant difference in streamflows, sediment yield from the landscape, and sediment discharge across climate model ensemble members, even when they are driven by the same forcing conditions (Dahl et al. 2018). One approach to reduce this variability while still minimizing the number of simulations required is to create an ensemble of the climate inputs (e.g., Cotterman et al. 2018; Neupane et al. 2015; Praskievicz 2016; Shrestha et al. 2012; van Liew et al. 2012), with the implicit assumption that the mean behavior across models is representative of the most likely future conditions.

Ensembled climate inputs are typically created by averaging the precipitation and temperature for a given point in time and space across multiple climate ensemble members. This single ensembled climate is then used as input for a hydrologic model to produce a streamflow and sediment discharge projection (Fig. 1). This is in contrast to an ensembled hydrology where the individual climate ensemble members are first downscaled and then run through the hydrologic model separately before the results are averaged together. It is not clear from the literature whether the streamflow and sediment transport resulting from a single, ensembled climate are the same as an ensembled hydrology.

To address this question, we modeled two large, adjacent watersheds in the Great Lakes region using the Soil and Water Assessment Tool (SWAT). We ran these models using both individually downscaled climate model outputs and ensembles of these outputs. We then
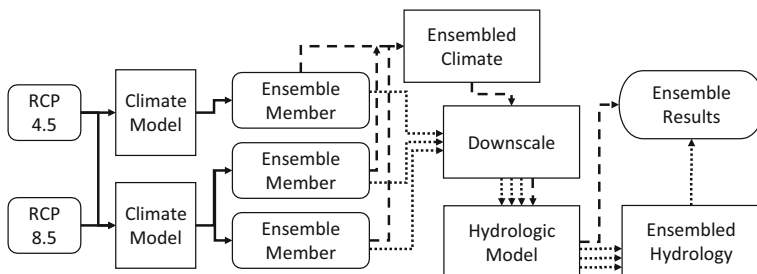

Fig. 1 In this article, we focus on two RCPs (4.5 and 8.5). Outputs associated with each pathway for multiple climate models (with one or more ensemble members) from CMIP5 were used as inputs to our analysis. We define ensembled climate scenarios as being an average of selected climate model ensemble members before they are downscaled and run through a hydrologic model (dashed lines). Ensembled hydrology scenarios are ones where the individual climate ensemble members are downscaled and run through the hydrologic model separately before being averaged together (dotted lines)

compared the streamflow, sediment yield, and sediment transport results to determine whether a single climate ensemble run or a subset of the climate ensemble members can be used to accurately represent the climate model effects of all of the available ensemble members.

2 Methods

2.1 Site description

The Maumee and St. Joseph River watersheds span the lower portion of Michigan, stretching from Lake Erie to Lake Michigan (Fig. 2). The Maumee watershed covers 17,015 km^2 and is primarily agricultural (74.7% row crops and 5.2% pasture), according to the 2006 National Land Cover Database (Fry et al. 2013). It drains portions of northeastern Indiana, southwestern Michigan, and northwestern Ohio to Lake Erie at Toledo, Ohio. The mainstem of the Maumee River has United States Geological Survey (USGS) flow gages at Defiance, Ohio (#04192500), and further downstream at Waterville, Ohio (#04193500). The flow at Waterville averaged 172.6 m^3/s between 1990 and 2009 while the average annual suspended sediment load was 1.2 million tonnes between 1990 and 2003.

The St. Joseph River watershed abuts the northwestern edge of the Maumee watershed and extends westward to its outlet at St. Joseph, Michigan, on Lake Michigan. The St. Joseph watershed is smaller (12,138 km^2) and has a lower proportion of agriculture (49.3% row crops, 12.2% pasture) than the Maumee, but more than twice as much forest (23.8% versus 8.2%).
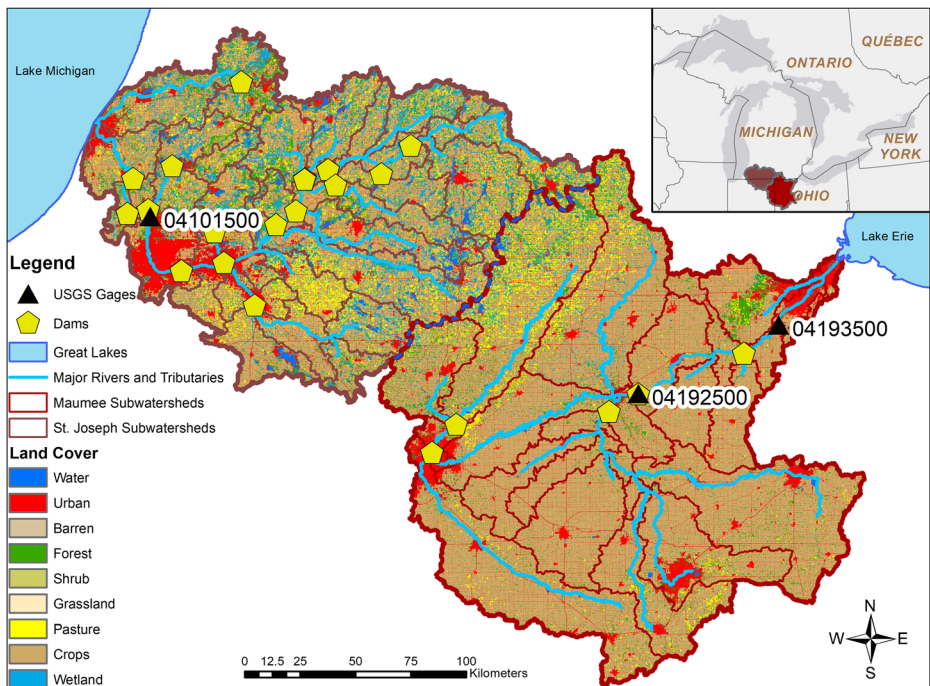

Fig. 2 The Maumee and St. Joseph River watersheds constitute a contiguous block of land between Lake Erie and Lake Michigan, covering portions of Michigan, Ohio, and Indiana (modified from Dahl et al. (2018))

The USGS gage at Niles, Michigan (#04101500), reported average flows of 113.6 m³/s between 1990 and 2009. There is no long-term sediment gaging at Niles, but a relationship between flows at the Niles gage and sediment moving downstream into St. Joseph Harbor was developed by the USACE (2007). While this sediment-discharge relationship is not appropriate for individual events, it showed good long-term agreement with both empirical values and harbor dredging records (USACE 2007).

There are numerous dams throughout both the Maumee and St. Joseph River watersheds. The dams in this region were typically built for small-scale hydropower or recreational purposes and are operated as run-of-river (inflow equal to outflow). Figure 2 indicates the location of dams with more than 1.2 million m³ of storage.

2.2 Soil and Water Assessment Tool models

We used individual SWAT models for each watershed based on 1 arc-second resolution elevation data from the National Elevation Dataset, land use/land cover from the 2006 National Land Cover Database (representative of our selected calibration and validation time periods), and soil data from the Natural Resources Conservation Service's Soil Survey Geographic database. Information on dams included in the models was obtained from the National Inventory of Dams maintained by the U.S. Army Corps of Engineers. The detailed development and calibration of these models is described in Dahl et al. (2018). We ran the models using downscaled climate model data for 2010–2099, with the first 5 years as a warm-up period that was excluded from our analysis.

2.3 Climate data and downscaling

The Fifth Coupled Model Intercomparison Project (CMIP5) (Taylor et al. 2012) resulted in at least 234 ensemble members from 37 different climate models. We selected Representative Concentration Pathways (RCP) 4.5 (Masui et al. 2011) and 8.5 (Riahi et al. 2011) because they were both required by the CMIP5 experimental design and therefore had the largest numbers of available ensemble members. We retrieved bias-corrected, statistically downscaled versions of the CMIP5 climate model ensemble members from a dataset created by the United States Bureau of Reclamation and others (Brekke et al. 2013). The archive of downscaled CMIP5 model runs contains 70 complete runs of both RCP 4.5 and 8.5 with precipitation and temperature available on a monthly basis for the North American Land Data Assimilation (NLDAS) grid. We eliminated seven of the climate ensemble members (access1-3.1.rcp85, fgoals-s2.2.rcp85, fgoals-s2.3.rcp85, noresm1-me.1.rcp85, access1-3.1.rcp45, fgoals-s2.2.rcp45, and noresm1-me.1.rcp45) because they only had average temperatures available and one (hadgem2-es.1.rcp45) because it was missing data from December 2099.

We created four separate climate ensembles for each selected RCP by averaging the precipitation and temperature for all ensemble members at each time step and grid cell. The first pair (RCP 4.5 and RCP 8.5) of climate ensembles used all of the available climate model ensemble members. We then created a pair of climate ensembles based only on the 10 ensemble members submitted from the CSIRO mk3.6 model (Jeffrey et al. 2013), because this submission had the most individual ensemble members of any model in the CMIP5. A third pair of climate ensembles was then generated based on a representative subset of climate model ensemble members in an attempt to create a parsimonious representation of the full range of potential climate forcings. Finally, we created a fourth pair of climate ensembles

based on the subset of climate model ensemble members that did the best job of matching the historical climate of our study region.

To select climate ensemble members representative of the full range, we first totaled the precipitation for each run both spatially (over the NLDAS grid cells centered on 40.1875° N to 42.4375° N and 86.4375° W to 83.0625° W) and temporally (from 2010 to 2099). The ensemble members were then sorted based on the total precipitation and percentiles were assigned to each run. We then selected one ensemble member each from the top and bottom 10% and three from the middle 20% of total precipitation, choosing the same ensemble member from both RCP 4.5 and RCP 8.5 whenever possible. Additionally, we chose the lowest and highest ranked ensemble members from the CSIRO-mk3-6-0 model, because this submission had the largest number of ensemble members submitted to CMIP5. The selected members of the Representative ensembles are all shown in Table 1.

The Best Fit ensemble represents the set of climate ensemble members that best matched the historical climate of the study region. We determined this by comparing the total precipitation and average temperatures over the study region for each ensemble member for 1971–1999 to measured data from NLDAS. We selected the climate ensemble members that were in the most representative quartile for both precipitation and temperature (Supplementary Table S1). Similar to the observation of Knutti et al. (2010), we found that few ensemble members performed well for both temperature and precipitation, despite being bias-corrected to long-term (1961–1990) average temperature and precipitation (Brekke et al. 2013).

We downscaled each selected scenario using the same methodology as Dahl et al. (2018), which is based on the work of other researchers (Maurer and Hidalgo 2008; Wood et al. 2004). This method takes the bias-corrected, spatially downscaled data provided by the United States Bureau of Reclamation and disaggregates it spatially to individual gage locations and temporally to daily time steps.

2.4 Statistical analysis

We tested the differences between the hydrologic outputs of the ensembled climate and the ensembled hydrology of the individual climate ensemble members using analysis of covariance (ANCOVA). ANCOVA tests for differences in both the slope and y-intercept of regression lines fit to the data. We centered the years by subtracting the mean to minimize the effect of the large year values on the slope in the ANCOVA test. We did not quantitatively test for normality and heteroscedasticity of the residuals because the sample sizes ($n = 85$) were sufficient to minimize the impact of non-normality and heteroscedasticity due to the Central Limit Theorem (Ghasemi and Zahediasl 2012; Helsel et al. 2020; Pek et al. 2018;

Table 1 Climate ensemble members selected for use in the Representative climate ensemble

RCP 4.5		RCP 8.5	
Ensemble member	Precipitation percentile	Ensemble member	Precipitation percentile
access1-0.1.rcp45	3%	access1-0.1.rcp85	3%
ccsm4.4.rcp45	45%	csiro-mk3-6-0.10.rcp85	43%
mpi-esm-lr.1.rcp45	51%	mpi-esm-lr.1.rcp85	46%
csiro-mk3-6-0.7.rcp45	55%	ccsm4.4.rcp85	47%
csiro-mk3-6-0.1.rcp45	99%	csiro-mk3-6-0.1.rcp85	93%

Supplementary Fig. S1). We chose to include any potential outliers, although the use of annual averages tended to suppress these (Supplementary Fig. S2). We also tested each line for monotonic trends using the Mann-Kendall test after using the bias-corrected, pre-whitening outlined by Hamed (2009). We used a significance level of $\alpha = 0.05$ for all tests. The p values we present in the main body of this article are not adjusted for multiple comparison testing, but we point out any differences with the results adjusted to limit the false discovery rate (FDR) to no more than 5%. The adjusted p values for the ANCOVA tests, calculated after the method of Benjamini and Yekutieli (2001), are provided in the Supplementary Material (Supplementary Tables S2, S3, and S4).

3 Results and discussion

3.1 SWAT model validation

The SWAT model of the Maumee River watershed was previously calibrated to monthly flow and sediment data for water years 1991–1999 (Dahl et al. 2018). The modeled streamflow has a Nash-Sutcliffe efficiency (NSE) of 0.79 at the Waterville, Ohio, gage and 0.86 at the Defiance, Ohio, gage for the validation time period of October 1987–December 1990 and January 2000–September 2003. Over the same time period, the modeled sediment at Waterville, Ohio, has a NSE of 0.48 and a % Bias of + 2.5%.

The St. Joseph River SWAT model was calibrated to calendar years 1990–1999 and validated using 2000–2009. The monthly model streamflow at Niles, Michigan, has a NSE of 0.72 for the validation time period. The modeled monthly sediment has a NSE of 0.29 and a % Bias of + 10.5% for the validation time period relative to a sediment rating curve at Niles, Michigan (Dahl et al. 2018).

3.2 Streamflow

The mean annual streamflow based on the ensembled climate for the full set of GCM outputs and that from the ensembled hydrology for the same input data are shown in Fig. 3. This figure also shows the range of all of the individual climate ensemble members. The streamflow for the ensembled climate is significantly different from the ensembled hydrology for the Maumee and St. Joseph Rivers under both RCP 4.5 and RCP 8.5 at the 5% level of confidence. Supplementary Table S2 provides both F-statistics and p values for the ANCOVA tests. These significant differences are all based on the y-intercept and not slope, indicating a constant, systematic, negative bias induced by the ensembled climate. This parallel, non-divergent behavior is observable in plots of the streamflow (Fig. 3, Supplementary Fig. S1). This is true even after controlling the FDR to no more than 5% (Supplementary Table S2). The ensembled climate produces a mean annual streamflow that is 16.2 to 17.7 m³/s (12.1 to 13.0%) lower than the ensembled hydrology for the Maumee River and 7.1 to 8.2 m³/s (5.2 to 6.1%) for the St. Joseph River (Table 2).

The Maumee River streamflow from the ensembled hydrology has statistically significant upward trends for both RCP 4.5 ($\tau = 0.258$, $p < 0.001$) and RCP 8.5 ($\tau = 0.514$, $p < 0.001$). The ensembled climate streamflow trend is also significant for both RCP 4.5 ($\tau = 0.168$, $p = 0.024$) and RCP 8.5 ($\tau = 0.363$, $p < 0.001$). While all of the trends are statistically significant, the difference in p values demonstrates that the choice of ensembling methodology can affect
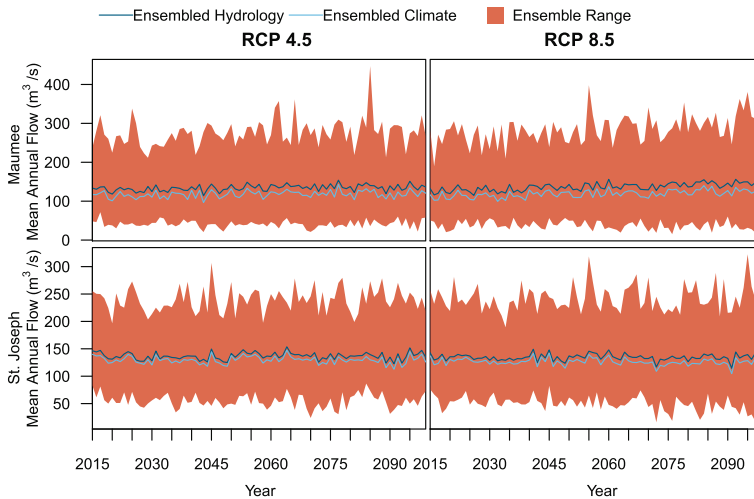
Fig. 3 The mean annual flow of all the climate ensemble members is greater than the mean annual flow of the ensembled climate run through the same hydrologic model. This is true for both the Maumee (top) and St. Joseph Rivers (bottom) and regardless of RCP. The red shaded area represents the full range of the individual climate runs used to create the ensembled hydrology

the detection of projected trends. None of the St. Joseph River streamflows had statistically significant streamflow trends. This disparity is likely due to the difference in land use between the two watersheds. The Maumee has a much higher proportion of agriculture and, as noted by Dahl et al. (2018), this leads to a feedback effect on streamflow under a warming climate. As the climate warms, crops mature and are harvested earlier in the year by the model, reducing the late season transpiration and allowing greater runoff. Increasing crop yields have been noted as a potential effect of climate change (Pryor et al. 2014).

Table 2 The ensembled climate streamflow is consistently biased lower than from the ensembled hydrology, regardless of the choice of ensemble members

	RCP	Ensemble Method	Maumee mean annual streamflow (m³/s)	St. Joseph mean annual streamflow (m³/s)
All members	4.5	Hydrology	134.0	136.7
		Climate	117.8	129.7
	8.5	Hydrology	136.8	134.7
		Climate	119.1	126.4
CSIRO mk3.6	4.5	Hydrology	140.9	143.2
		Climate	125.9	136.4
	8.5	Hydrology	139.9	139.8
		Climate	124.3	132.6
Representative	4.5	Hydrology	136.8	138.0
		Climate	123.5	131.6
	8.5	Hydrology	133.4	130.7
		Climate	121.8	124.2
Best Fit	4.5	Hydrology	134.7	134.8
		Climate	121.1	128.5
	8.5	Hydrology	143.2	137.5
		Climate	128.0	130.1

The difference between the negative bias of the ensembled climate relative to the ensembled hydrology is likely due to the combination of the nonlinear hydrologic processes and a loss of the precipitation signal. Knutti et al. (2010) noted this effect in GCMs and showed that the distribution of precipitation in multi-model ensembles is narrower than any of the individual runs because the differences from average are not co-located in space or time. When the precipitation and temperature differences are translated through the hydrologic model, this effect can be magnified.

3.3 Sediment yield

The annual sediment yield produced by the ensembled climate and ensembled hydrology for all GCM outputs is significantly different ($p < 0.05$) for both the St. Joseph and Maumee Rivers (Fig. 4; Supplementary Table S5). None of the slopes is significantly different between the two ensembling methods, indicating that the difference between the two manifests as a consistent bias, with the ensembled climate resulting in mean annual sediment, yields 900 to 952 kilotonnes (12.4–14.0%) lower in the Maumee and 8.6 to 10.1 kilotonnes (10.8–11.2%) lower in the St. Joseph. The difference in intercepts remains significant even after controlling the FDR to no more than 5% (Supplementary Table S3).

All of the ensembled sediment yields except the RCP 4.5 Best Fit ensembled hydrology in the St. Joseph watershed have statistically significant, upward trends with p values less than 0.02. The ensembled climate results in consistently lower sediment yields than the ensembled hydrology. It is interesting to note that both watersheds show large increases in sediment yield towards the end of the century under RCP 8.5. These increases may be the result of earlier crop harvest leaving behind bare ground for longer periods of the year. Dahl et al. (2018) note that while the model leaves the field fallow, farmers faced with a changing climate may instead choose to plant two crops per year or crops with a longer time to maturity.
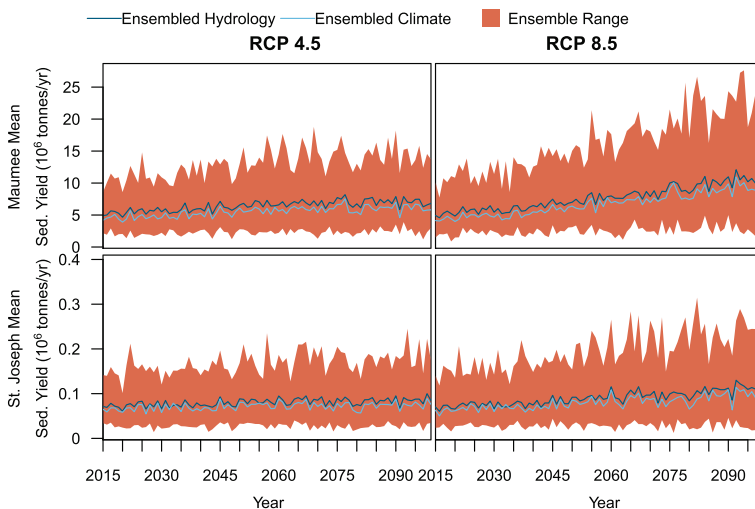

Fig. 4 The mean amount of sediment delivered to the river each year has statistically significant differences between the ensembled climate and ensembled hydrology approaches. The red shaded area represents the full range of the individual climate runs
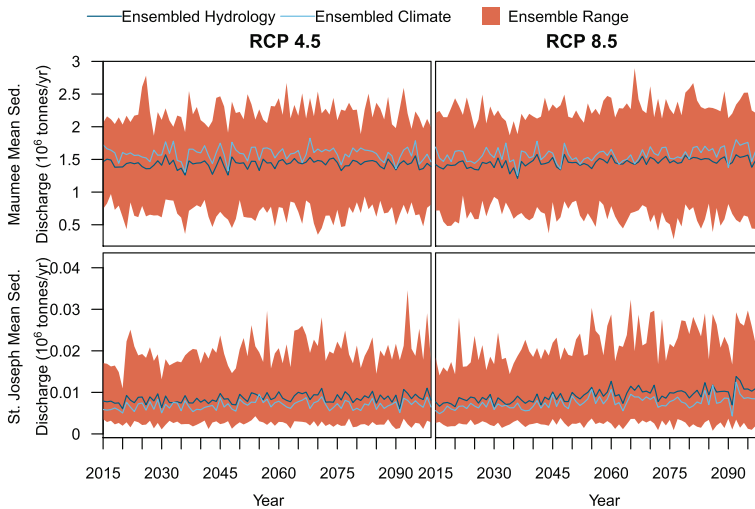
Fig. 5 Sediment discharge differs between the climate and hydrologic ensembling methods. The direction of this difference varies between the watersheds and depends on the reservoir sediment properties

3.4 Sediment discharge

We report the sediment discharge as the amount of sediment exiting the mouth of the river (Fig. 5). The ensembled climate and ensembled hydrology for all climate ensemble members produce statistically different sediment discharges ($p < 0.05$) for all watershed and RCP combinations, based on the intercept (Supplementary Table S4). This is true even after controlling the FDR to no more than 5% (Supplementary Table S4). The total annual sediment discharge at the mouth of the St. Joseph River is 1.6 to 1.7 (~18.7%) kilotonnes lower for the ensembled climate than the ensembled hydrology.

The sediment discharge for the Maumee is the one variable we examined where the ensembled climate is greater than the ensembled hydrology. The total annual sediment discharge at the mouth of the Maumee River is 118.5 to 155.4 kilotonnes (8.1 to 10.8%) higher for the ensembled climate but 1.6 to 1.7 kilotonnes (~18.7%) lower at the mouth of the St. Joseph River. This difference between watersheds is likely due to the presence of reservoirs and the different predominant grain sizes. While all of the reservoirs in the SWAT models are treated as run-of-river, sediment can still settle out in them because of the slower velocities and the physical barrier of the dam. SWAT simulates this settling as a function of the incoming and equilibrium sediment concentrations in the reservoir and the predominant grain size. The median grain size (d50) differs between the St. Joseph and Maumee watersheds and this leads to a large difference in the equilibrium sediment concentration. In the Maumee model, the d50 is 0.041 mm (coarse silt) and the equilibrium sediment concentration is 1135 mg/l. Changing these parameters to match those of the St. Joseph model reservoirs (d50 = 0.265 mm or fine sand; equilibrium sediment concentration = 335 mg/l) causes more sediment to settle out in the reservoirs due to the lower equilibrium concentration and greatly reduces the difference between the Maumee River sediment flux calculations for the ensembled climate and ensembled hydrology results (Supplementary Fig. S3). The importance of dams for sediment movement through the

Maumee watershed has been previously established. Alighalehbabakhani et al. (2017) found that Independence Dam, the second most downstream dam on the mainstem of the Maumee, had the highest sediment accumulation of the 12 dams they studied across the Great Lakes and their modeling results suggest that the reservoir behind this dam may already be filled with sediment.

3.5 Effect of ensemble member choice

We created an ensembled climate and ensembled hydrology based on the 10 members for the CSIRO mk3.6 model (Fig. 6). The resulting ensembled mean annual flows are significantly ($p < 0.05$) different in their y-intercepts, but not their slopes (Supplementary Table S2). The ensembled climate results in consistently lower flows than the ensembled hydrology, with decreases in the Maumee of 15 to 15.7 m³/s (10.6–11.2%) and 6.8 to 7.2 m3/s (4.8 to 5.2%) in the St. Joseph (Table 2). All of the Maumee River ensembles and the RCP 4.5 St. Joseph River ensembles have statistically significant increasing trends. The CSIRO mk3.6 ensemble results for sediment yield (Supplementary Fig. S4, Supplementary Table S3) and sediment discharge (Supplementary Fig. S5, Supplementary Table S4) are similar to those for the ensembles based on the entire range of available climate ensemble members.

The range of streamflows for the 10 ensemble members from the CSIRO mk3.6 model spans most of the range of all the climate ensemble members (Fig. 6). While each ensemble member is from the same global climate model, they were initialized at different times from the control run (Jeffrey et al. 2013), producing significant variability over the duration of the climate change run. Earlier work found that the intermodel variability is the largest source of uncertainty in climate change ensemble members (Giorgi and Francisco 2000a, b), but the wide range of streamflow and sediment results produced by the CSIRO mk3.6 ensemble members indicates that it may no longer be the case for modern GCMs.

The Representative ensembled climate streamflow is significantly different from the Representative ensembled hydrology across both models and RCPs (Fig. 7, Supplementary
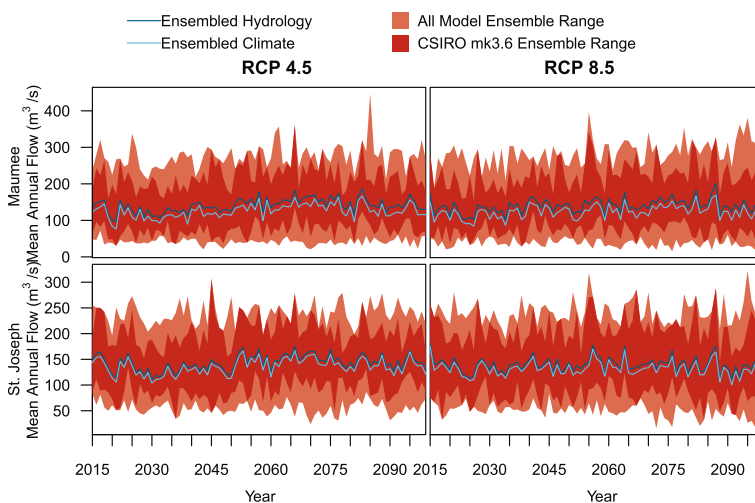

Fig. 6 The ten ensemble members from the CSIRO mk3.6 model have almost as large a range as all the available climate ensemble members
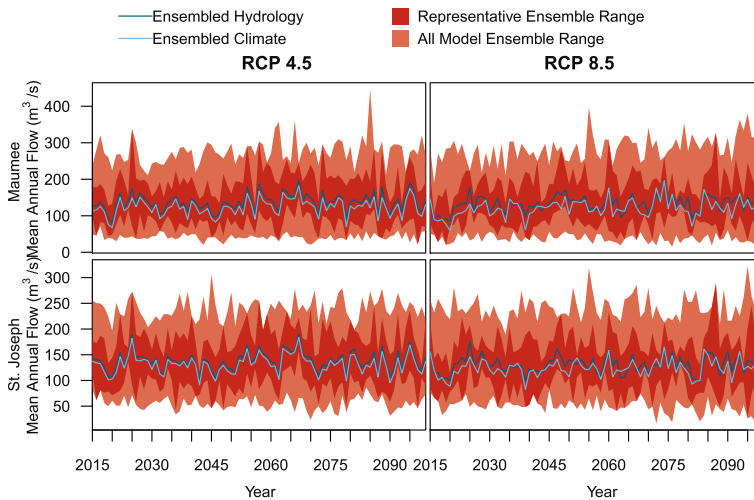
Fig. 7 The ensemble of a small number of representative climate model ensemble members does not capture the full range of variability and suffers from the same bias between ensembled climate and ensembled hydrology

Table S2). This difference is due to the *y*-intercept and there is no statistically significant difference in slope. After controlling the FDR to no more than 5%, all of the intercepts for the Maumee River remain significant, but not those of the St. Joseph River (Supplementary Table S2). The ensembled climate results in streamflows that are 11.6 to 13.3 m³/s (8.7–9.7%) lower in the Maumee and 6.4 to 6.5 m³/s (4.75–5.0%) lower in the St. Joseph than the ensembled hydrology (Table 2). Both ensembling methods for the Representative climate ensemble members have statistically significant increasing trends for the Maumee River under RCP8.5. The St. Joseph River only has a statistically significant trend for the RCP 8.5 ensembled climate. Even though we selected the ensemble members to be representative, there are very different patterns of statistically significant trends between this method and the ensembles made up of the full suite of climate ensemble members. The impacts of ensembling using representative climate ensemble members on sediment yield and sediment discharge can be seen in Supplementary Fig. S6 and Supplementary Fig. S7, respectively. The ANCOVA results are available in Supplementary Table S3 and Supplementary Table S4.

The Best Fit ensembles were similar to the Representative ensembles (Fig. 8). Again, the streamflow associated with the ensembled climate was lower than the ensembled hydrology and significantly different in intercept but not in slope (Table 2, Supplementary Table S2). After controlling the FDR to no more than 5%, the intercepts for the Maumee River remain significant, but only the intercept for RCP 8.5 is significant in the St. Joseph River (Supplementary Table S2). The ensembled climate mean annual streamflow was 13.6 to 15.3 m³/s (10.1–10.7%) lower in the Maumee and 6.3 to 7.4 m³/s (4.7–5.4%) lower in the St. Joseph than the streamflow for the ensembled hydrology. In spite of these ensemble members being among the best at matching the historic climate, they still account for some of the most extreme streamflows in the all model ensembles and cover a large portion of the range of results. This is not surprising because GCMs that perform well for historic climate can produce divergent outputs under future climate scenarios (Knutti et al. 2010). The impacts of ensembling using the Best Fit climate ensemble members on sediment yield and sediment discharge can be seen in Supplementary Figs. S8 and S9, respectively.
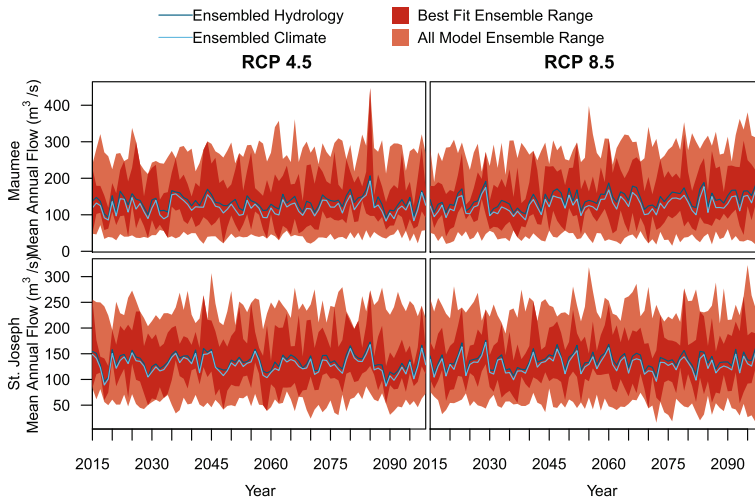
Fig. 8 The Best Fit ensemble, based on the climate model ensemble members that most closely match historical climate results, is still susceptible to differences between ensembled hydrology and ensembled climate

These results clearly show that the ensembled climate and ensembled hydrology do not produce the same results. Since many studies have limited computational resources, there are likely to be questions about which, if any, of the ensemble subsets (e.g., CSIRO mk3.6, Representative, or Best Fit) should be used. When we compare the results for the three ensemble subsets presented here to the results from the full set of ensemble members, the mean streamflows are all within 5% (Table 3). The mean Representative ensemble streamflow, sediment yield, and sediment discharge are all within 3% of the ensembled hydrology for the full set of GCM outputs. Based on these results, we suggest that researchers looking to mimic the full range of ensemble members with a limited subset should consider something akin to our Representative ensemble where a limited subset of high, average, and low precipitation scenarios are combined.

Comparing the subset ensembles to the full ensemble raises some philosophical questions about what the correct ensemble is. A common assumption in climate change studies is that climate models that predict the past well will continue to do so in the future. This is far from certain, because climate models that agree on historical conditions often diverge for future predictions (Knutti et al. 2010; Pierce et al. 2009). The Reliability Ensemble Averaging (REA) technique (Giorgi and Mearns 2002) attempts to address this by assigning weights based on

Table 3 The difference from the mean ensembled hydrology for all GCM outputs shows that the three different subset ensembles all produce similar results to the full set of ensembles

		Streamflow		Sediment yield		Sediment discharge	
	RCP	Maumee	St. Joseph	Maumee	St. Joseph	Maumee	St. Joseph
CSIRO mk3.6	4.5	+5.2%	+4.7%	+9.5%	+11.3%	+3.9%	+12.2%
	8.5	+2.3%	+3.8%	+5.0%	+7.5%	+3.4%	+7.8%
Representative	4.5	+2.1%	+1.0%	+2.1%	+2.5%	+1.9%	+1.9%
	8.5	−2.5%	−2.9%	+0.8%	+0.0%	−1.3%	−0.8%
Best Fit	4.5	+0.5%	−1.4%	+1.1%	−0.9%	+0.5%	−2.0%
	8.5	+4.7%	+2.1%	+7.0%	+2.8%	+2.3%	+3.3%

both historical accuracy and how close they are to the mean of all future model predictions. The REA method reduces the weight of outliers, implicitly assuming that they are much less likely to occur and may be the product of a flawed model. In contrast, our Representative ensemble equally weights all of the ensemble members based solely on historical accuracy.

There are numerous other methods published in the literature for selecting subset ensembles. Karmalkar et al. (2019) provides an example of how to select a limited subset that still includes inter-model variability. Ross and Najjar (2019) go a step further by comparing multiple subset methods that have varying goals and testing them for sensitivity to ensemble size. Their recommended method, KKZ, required at least 5 to 11 members to capture 75% of the variability of the full ensemble, similar to our 5-member Representative ensemble. We suggest that if the purpose of a study is to look at the potential range of hydrologic and sediment impacts that may occur, a Representative ensemble is a good surrogate for the full range of climate ensemble members.

4 Conclusions

The use of an ensembled climate scenario as the input to a hydrologic model biases the result relative to the ensembled hydrologic output based on the individual climate ensemble members, producing results that are significantly different ($p < 0.05$). This is most likely due to the loss of the precipitation signal that is offset temporally and spatially in the individual ensemble members. When possible, ensembling should be done using the outputs of hydrological models rather than their inputs. Avoiding or acknowledging the potential for biasing the results will help scientists and policymakers formulate better responses to the changing climate.

It is often necessary to use an ensembled climate or a limited number of climate runs due to computational limitations or time constraints, but it is important to do so in an informed manner. Here, we have confirmed that it is possible to capture a significant amount of the range of all climate ensemble members using a limited number (5–10) of members. Selection of the ensemble members to encompass the full range of temperature and precipitation can result in a parsimonious ensemble that produces mean annual hydrologic and sediment model results similar to the full suite of potential ensemble members. While the selection of the particular ensemble members has been the subject of a number of recent articles, it deserves further study and should consider whether the goal is to achieve a consistent mean or account for the likely range of future scenarios.

Declarations

References

Alighalehbabakhani F, Miller CJ, Baskaran M, Selegean JP, Barkach JH, Dahl T, Abkenar SMS (2017) Forecasting the remaining reservoir capacity in the Laurentian Great Lakes watershed. J Hydrol 555:926–937. https://doi.org/10.1016/j.jhydrol.2017.10.052

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. Ann Stat 29:1165–1188

Brekke L, Thrasher BL, Maurer EP, Pruitt T (2013) Downscaled CMIP3 and CMIP5 climate and hydrology projections: release of downscaled CMIP5 climate projections, comparison with preceding information, and summary of user needs. U.S. Department of the Interior, Bureau of Reclamation,

Cherkauer KA, Sinha T (2010) Hydrologic impacts of projected future climate change in the Lake Michigan region. J Great Lakes Res 36:33–50. https://doi.org/10.1016/j.jglr.2009.11.012

Cotterman KA, Kendall AD, Basso B, Hyndman DW (2018) Groundwater depletion and climate change: future prospects of crop production in the Central High Plains Aquifer. Clim Chang 146:187–200. https://doi.org/10.1007/s10584-017-1947-7

Dahl TA, Kendall AD, Hyndman DW (2018) Impacts of projected climate change on sediment yield and dredging costs. Hydrol Process 32:1223–1234. https://doi.org/10.1002/hyp.11486

Fry JA et al (2013) Completion of the 2006 National Land Cover Database for the conterminous United States. Photogramm Eng Remote Sens 130:294–304

Ghasemi A, Zahediasl S (2012) Normality tests for statistical analysis: a guide for non-statisticians. Int J Endocrinol Metab 10:486–489. https://doi.org/10.5812/ijem.3505

Giorgi F, Francisco R (2000a) Evaluating uncertainties in the prediction of regional climate change. Geophys Res Lett 27:1295–1298. https://doi.org/10.1029/1999gl011016

Giorgi F, Francisco R (2000b) Uncertainties in regional climate change prediction: a regional analysis of ensemble simulations with the HADCM2 coupled AOGCM Climate Dynamics 16:169-182 doi:https://doi.org/10.1007/pl00013733

Giorgi F, Mearns LO (2002) Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the "reliability ensemble averaging"(REA) method. J Clim 15:1141–1158

Hamed KH (2009) Enhancing the effectiveness of prewhitening in trend analysis of hydrologic data. J Hydrol 368:143–155. https://doi.org/10.1016/j.jhydrol.2009.01.040

Helsel DR, Hirsch RM, Ryberg KR, Archfield SA, Gilroy EJ (2020) Statistical methods in water resources. Reston, VA doi:https://doi.org/10.3133/tm4A3

Jeffrey S et al. (2013) Australia's CMIP5 submission using the CSIRO Mk3. 6 model Aust Meteor Oceanogr J 63:1-13

Johnson T et al (2015) Modeling streamflow and water quality sensitivity to climate change and urban development in 20 US watersheds. J Am Water Resour Assoc 51:1321–1341. https://doi.org/10.1111/1752-1688.12308

Karmalkar AV, Thibeault JM, Bryan AM, Seth A (2019) Identifying credible and diverse GCMs for regional climate change studies—case study: Northeastern United States. Clim Chang 154:367–386

Knutti R, Furrer R, Tebaldi C, Cermak J, Meehl GA (2010) Challenges in Combining Projections from Multiple Climate Models J Clim 23:2739–2758 doi:https://doi.org/10.1175/2009jcli3361.1

Masui T et al (2011) An emission pathway for stabilization at 6 Wm($-$2) radiative forcing. Clim Chang 109:59–76. https://doi.org/10.1007/s10584-011-0150-5

Maurer EP, Hidalgo HG (2008) Utility of daily vs. monthly large-scale climate data: an intercomparison of two statistical downscaling methods. Hydrol Earth Syst Sci 12:551–563

Neupane RP, White JD, Alexander SE (2015) Projected hydrologic changes in monsoon-dominated Himalaya Mountain basins with changing climate and deforestation. J Hydrol 525:216–230. https://doi.org/10.1016/j.jhydrol.2015.03.048

O'Neal MR, Nearing MA, Vining RC, Southworth J, Pfeifer RA (2005) Climate change impacts on soil erosion in Midwest United States with changes in crop management. Catena 61:165–184. https://doi.org/10.1016/j.catena.2005.03.003

Park JY et al (2011) Assessment of Future Climate Change Impacts on Water Quantity and Quality for a Mountainous Dam Watershed Using SWAT. Trans ASABE 54:1725–1737

Pek J, Wong O, Wong ACM (2018) How to Address Non-normality: A Taxonomy of Approaches, Reviewed, and Illustrated. Front Psychol 9:2140. https://doi.org/10.3389/fpsyg.2018.02104

Pierce DW, Barnett TP, Santer BD, Gleckler PJ (2009) Selecting global climate models for regional climate change studies. Proc Natl Acad Sci 106:8441–8446

Praskievicz S (2016) Impacts of projected climate changes on streamflow and sediment transport for three snowmelt-dominated rivers in the interior Pacific Northwest. River Res Appl 32:4–17. https://doi.org/10.1002/rra.2841

Pryor SC et al. (2014) Ch. 18: Midwest. In: Melillo JM, Richmond TC, Yohe GW (eds) Climate Change Impacts in the United States: The Third National Climate Assessment. U.S. Global Change Research Program, pp 418-440. doi:https://doi.org/10.7930/J0J1012N

Riahi K et al (2011) RCP 8.5-A scenario of comparatively high greenhouse gas emissions. Clim Chang 109:33–57. https://doi.org/10.1007/s10584-011-0149-y

Ross AC, Najjar RG (2019) Evaluation of methods for selecting climate models to simulate future hydrological change. Clim Chang 157:407–428

Serpa D et al (2015) Impacts of climate and land use changes on the hydrological and erosion processes of two contrasting Mediterranean catchments. Sci Total Environ 538:64–77. https://doi.org/10.1016/j.scitotenv.2015.08.033

Shrestha RR, Dibike YB, Prowse TD (2012) Modelling of climate-induced hydrologic changes in the Lake Winnipeg watershed. J Great Lakes Res 38:83–94. https://doi.org/10.1016/j.jglr.2011.02.004

Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experiment design. Bull Am Meteorol Soc 93:485–498. https://doi.org/10.1175/bams-d-11-00094.1

USACE (2007) St. Joseph River sediment transport modeling study. Detroit, MI

van Liew MW, Feng S, Pathak TB (2012) Climate change impacts on streamflow, water quality, and best management practices for the Shell and Logan Creek watersheds in Nebraska, USA. J Agric Biol Eng 5:13–34

Verma S, Bhattarai R, Bosch NS, Cooke RC, Kalita PK, Markus M (2015) Climate change impacts on flow, sediment and nutrient export in a Great Lakes watershed using SWAT. Clean-Soil Air Water 43:1464–1474. https://doi.org/10.1002/clen.201400724

Wood AW, Leung LR, Sridhar V, Lettenmaier DP (2004) Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. Clim Chang 62:189–216. https://doi.org/10.1023/B:CLIM.0000013685.99609.9e

Supplementary Material

Dahl et al. (2021). Climate and Hydrologic Ensembling Lead to Differing Streamflow and Sediment Yield Predictions.

Supplementary Table S1. Climate ensemble members selected for the best fit climate ensemble were chosen based on being in the most accurate quartile for both total precipitation and average temperature over the period 1971-1999.

RCP 4.5			RCP 8.5		
Ensemble Member	Precip. %	Temp. %	Ensemble Member	Precip. %	Temp. %
noresm1-m.1.rcp45	6%	15%	noresm1-m.1.rcp85	6%	16%
miroc-esm-chem.1.rcp45	8%	23%	miroc-esm-chem.1.rcp85	8%	23%
hadgem2-es.2.rcp45	13%	7%	hadgem2-es.2.rcp85	11%	7%
mri-cgcm3.1.rcp45	17%	8%	mri-cgcm3.1.rcp85	15%	9%
ccsm4.2.rcp45	25%	21%	ccsm4.2.rcp85	23%	21%
			miroc5.1.rcp85	24%	10%
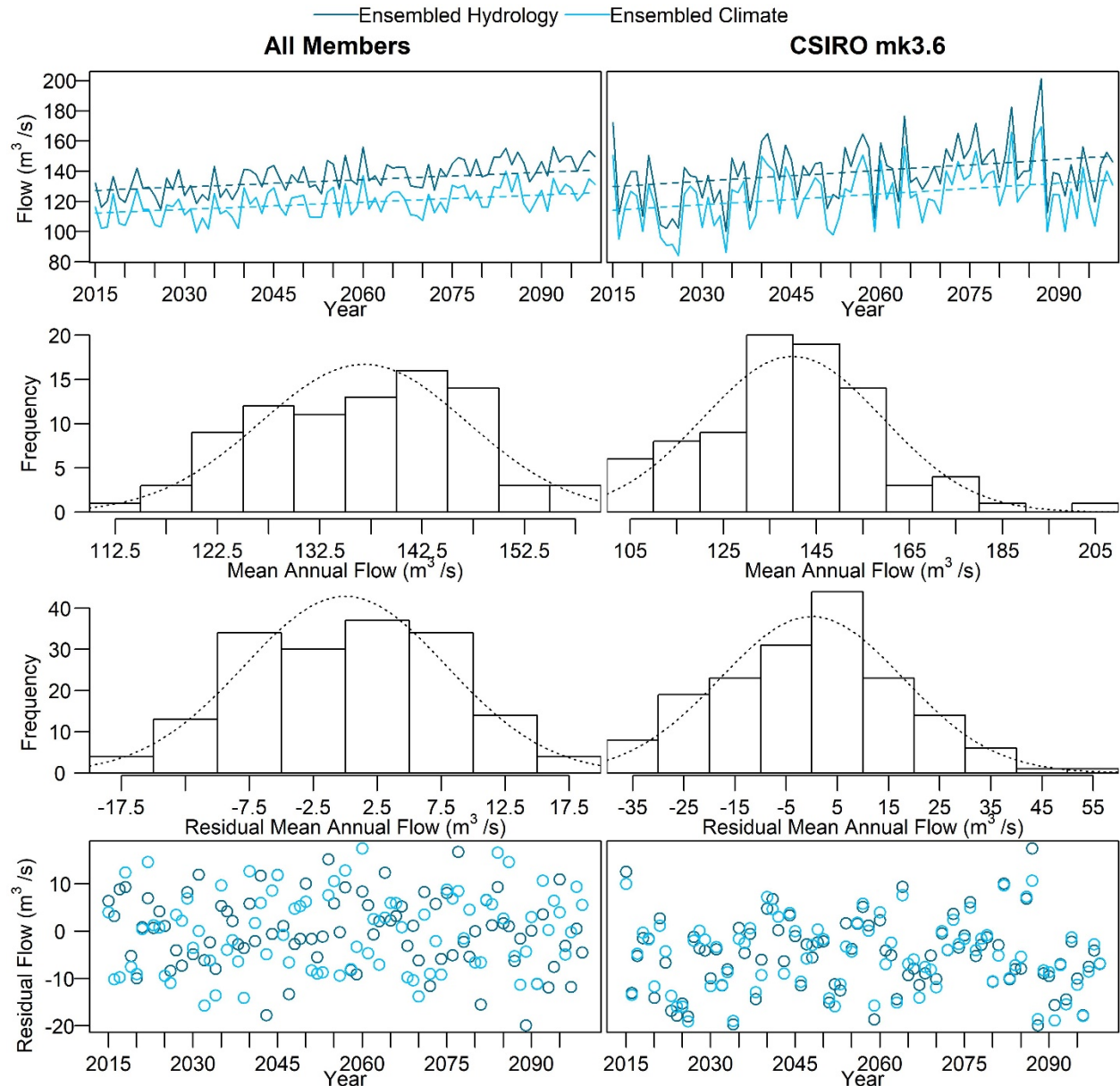
Supplementary Figure S1. This figure shows Maumee River streamflow simulated using ensembled hydrology (dark blue) and ensembled climate (light blue) for All Ensemble Members (left column) and for the CSIRO mk3.6 subset (right column) for illustrative purposes. The top row demonstrates that both the modeled streamflow (solid lines) and regression (dashed) lines are qualitatively parallel. The histograms of mean annual flow data (2nd row) and residual flow data after the ANCOVA regression (3rd row) are similar to normal distributions with the same mean and standard deviation (dotted lines), as predicted by the Central Limit Theorem (CLT). Plotting the residual flow data (4th row) shows no obvious trends in variance, indicating homescedasticity. These plots demonstrate that it is reasonable to apply the CLT for the mean annual hydrologic variables used in this study.
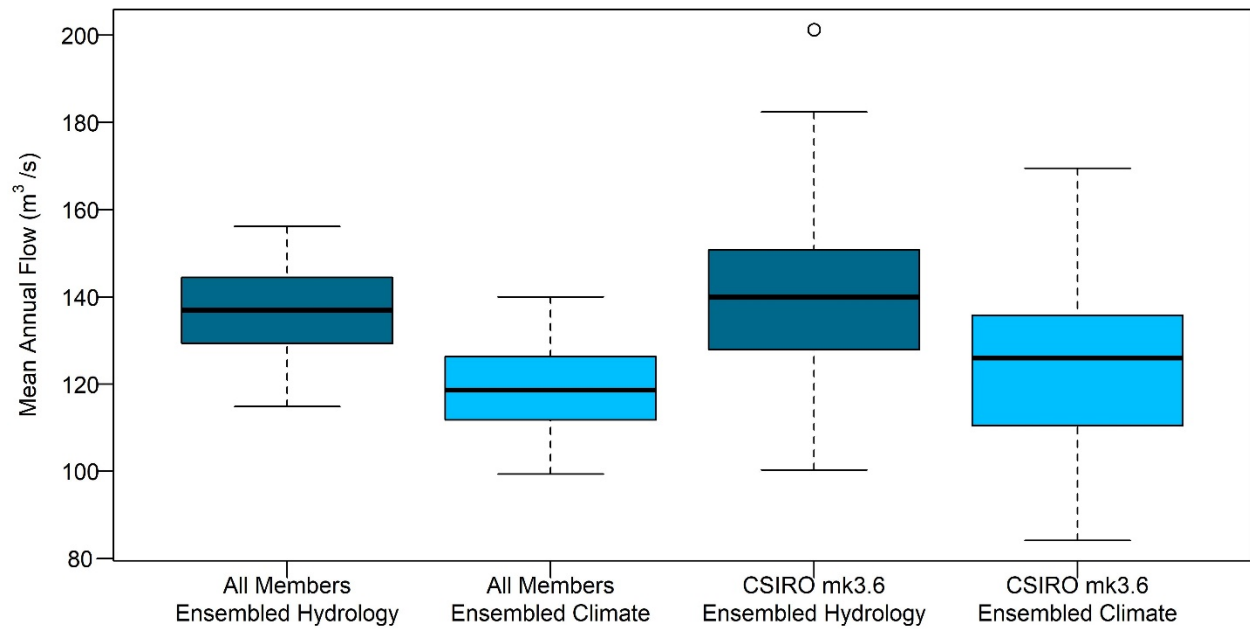
Supplementary Figure S2. Few outliers are present in the data and we decided to include all of them in the analysis. Shown here are boxplots of the Maumee River mean annual flow for the same sets of ensembles presented in Supplementary Figure S1. Only one outlier is indicated.

Supplementary Table S2. ANCOVA tests for streamflow show that the ensembled climate is significantly ($p < 0.05$, in bold) different from the ensembled hydrology in all cases, based on the intercept. After controlling the FDR to less than 5%, all of the Maumee intercepts remain significant ($p_{adj} < 0.05$, in bold). Five of the eight St. Joseph intercepts also remain significant (differences italicized). None of the slopes are significantly different.
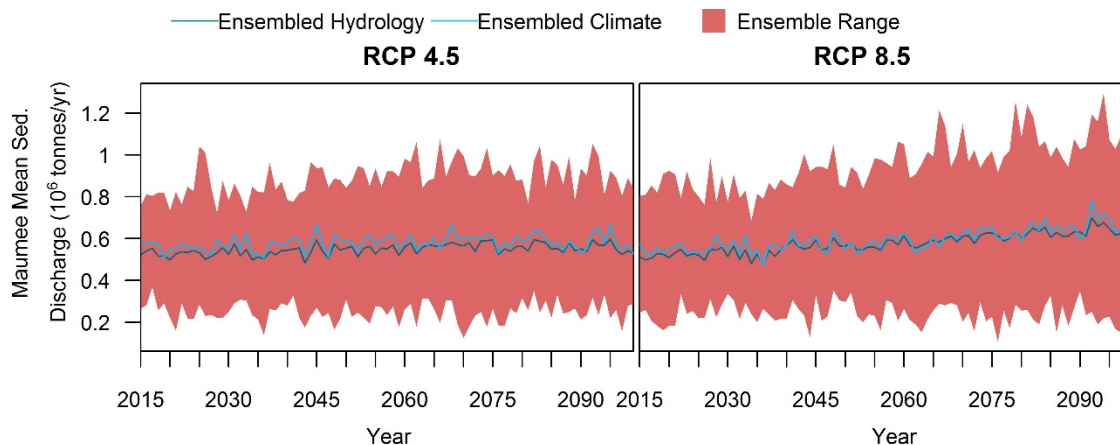
	RCP	Maumee ANCOVA		St. Joe ANCOVA	
		Slope	Intercept	Slope	Intercept
All Members	**4.5**	$F(1, 166) = 0.36$ $p = 0.550$	$F(1, 167) = 167.54$ **$p < 0.001$**	$F(1, 166) = 0.50$ $p = 0.479$	$F(1, 167) = 46.44$ **$p < 0.001$**
	8.5	$F(1, 166) = 3.58$ $p = 0.060$	$F(1, 167) = 149.82$ **$p < 0.001$**	$F(1, 166) = 1.27$ $p = 0.261$	$F(1, 167) = 105.46$ **$p < 0.001$**
CSIRO mk3.6	**4.5**	$F(1, 166) = 0.07$ $p = 0.792$	$F(1, 167) = 30.41$ **$p < 0.001$**	$F(1, 166) = 0.03$ $p = 0.855$	$F(1, 167) = 9.16$ **$p = 0.003$**
	8.5	$F(1, 166) = 0.05$ $p = 0.822$	$F(1, 167) = 32.35$ **$p < 0.001$**	$F(1, 166) = 0.05$ $p = 0.829$	$F(1, 167) = 10.45$ **$p < 0.001$**
Representative	**4.5**	$F(1, 166) = 0.03$ $p = 0.873$	$F(1, 167) = 13.50$ **$p < 0.001$**	$F(1, 166) = 0.06$ $p = 0.810$	$F(1, 167) = 5.04$ **$p = 0.026$**
	8.5	$F(1, 166) = 0.30$ $p = 0.583$	$F(1, 167) = 11.20$ **$p = 0.001$**	$F(1, 166) = 0.66$ $p = 0.416$	$F(1, 167) = 6.58$ **$p = 0.001$**
Best Fit	**4.5**	$F(1, 166) = 0.00$ $p = 0.985$	$F(1, 167) = 18.75$ **$p < 0.001$**	$F(1, 166) = 0.07$ $p = 0.797$	$F(1, 167) = 6.28$ **$p = 0.013$**
	8.5	$F(1, 166) = 0.23$ $p = 0.635$	$F(1, 167) = 23.13$ **$p < 0.001$**	$F(1, 166) = 0.23$ $p = 0.629$	$F(1, 167) = 9.38$ **$p < 0.001$**

Supplementary Table S3. ANCOVA tests on the mean annual sediment yield demonstrate that the ensembled climate is significantly (p < 0.05, in bold) than the ensembled hydrology in intercept but not slope, for all cases. After controlling the FDR to less than 5%, all of the Maumee intercepts remain significant (p_{adj} < 0.05, in bold). Seven of the eight St. Joseph intercepts also remain significant (difference italicized).

|		Maumee		St. Joe	
		ANCOVA		**ANCOVA**	
	RCP	**Slope**	**Intercept**	**Slope**	**Intercept**
All Members	**4.5**	F(1, 166) = 0.35 p = 0.556	F(1, 167) = 115.92 **p < 0.001**	F(1, 166) = 1.42 p = 0.234	F(1, 167) = 54.89 **p < 0.001**
	8.5	F(1, 166) = 1.12 p = 0.292	F(1, 167) = 97.40 **p < 0.001**	F(1, 166) = 2.78 p = 0.097	F(1, 167) = 47.77 **p < 0.001**
CSIRO mk3.6	**4.5**	F(1, 166) = 0.05 p = 0.830	F(1, 167) = 29.21 **p < 0.001**	F(1, 166) = 0.65 p = 0.420	F(1, 167) = 15.97 **p < 0.001**
	8.5	F(1, 166) = 0.01 p = 0.933	F(1, 167) = 28.74 **p < 0.001**	F(1, 166) = 0.49 p = 0.486	F(1, 167) = 14.54 **p < 0.001**
Representative	**4.5**	F(1, 166) = 0.05 p = 0.831	F(1, 167) = 16.98 **p < 0.001**	F(1, 166) = 0.06 p = 0.799	F(1, 167) = 7.34 **p = 0.007**
	8.5	F(1, 166) = 0.45 p = 0.505	F(1, 167) = 10.24 **p = 0.002**	F(1, 166) = 0.29 p = 0.590	F(1, 167) = 3.85 **p = 0.002**
Best Fit	**4.5**	F(1, 166) = 0.01 p = 0.944	F(1, 167) = 30.26 **p < 0.001**	F(1, 166) = 0.00 p = 0.996	F(1, 167) = 15.92 **p < 0.001**
	8.5	F(1, 166) = 0.05 p = 0.822	F(1, 167) = 20.94 **p < 0.001**	F(1, 166) = 0.78 p = 0.380	F(1, 167) = 12.79 **p < 0.001**

Supplementary Table S4. ANCOVA tests on the mean total annual sediment discharge show that the ensembled climate is significantly ($p < 0.05$, in bold) different from the ensembled hydrology for all cases. This difference manifests in the intercepts but not the slopes. After controlling the FDR to less than 5%, all of the intercepts remain significant ($p_{adj} < 0.05$, in bold).

|	RCP	Maumee		St. Joe	
		ANCOVA		ANCOVA	
		Slope	Intercept	Slope	Intercept
All Members	**4.5**	$F(1, 166) = 0.63$ $p = 0.427$	$F(1, 167) = 144.50$ **$p < 0.001$**	$F(1, 166) = 0.43$ $p = 0.513$	$F(1, 167) = 106.43$ **$p < 0.001$**
	8.5	$F(1, 166) = 0.16$ $p = 0.690$	$F(1, 167) = 88.15$ **$p < 0.001$**	$F(1, 166) = 2.38$ $p = 0.125$	$F(1, 167) = 79.06$ **$p < 0.001$**
CSIRO mk3.6	**4.5**	$F(1, 166) = 0.05$ $p = 0.832$	$F(1, 167) = 46.61$ **$p < 0.001$**	$F(1, 166) = 0.38$ $p = 0.536$	$F(1, 167) = 36.40$ **$p < 0.001$**
	8.5	$F(1, 166) = 0.09$ $p = 0.770$	$F(1, 167) = 23.65$ **$p < 0.001$**	$F(1, 166) = 0.09$ $p = 0.769$	$F(1, 167) = 31.93$ **$p < 0.001$**
Representative	**4.5**	$F(1, 166) = 0.10$ $p = 0.754$	$F(1, 167) = 9.79$ **$p = 0.002$**	$F(1, 166) = 0.00$ $p = 0.976$	$F(1, 167) = 19.56$ **$p < 0.001$**
	8.5	$F(1, 166) = 1.48$ $p = 0.225$	$F(1, 167) = 15.97$ **$p < 0.001$**	$F(1, 166) = 0.44$ $p = 0.509$	$F(1, 167) = 10.09$ **$p < 0.001$**
Best Fit	**4.5**	$F(1, 166) = 0.93$ $p = 0.338$	$F(1, 167) = 17.61$ **$p < 0.001$**	$F(1, 166) = 0.64$ $p = 0.426$	$F(1, 167) = 29.65$ **$p < 0.001$**
	8.5	$F(1, 166) = 2.65$ $p = 0.105$	$F(1, 167) = 15.02$ **$p < 0.001$**	$F(1, 166) = 0.60$ $p = 0.440$	$F(1, 167) = 27.12$ **$p < 0.001$**

Supplementary Figure S3. The reservoir sediment processes play an important role in sediment discharge, particularly under climate change. Increasing the reservoir sediment d_{50} and reducing the normal concentration for the Maumee model to match those used for the St. Joseph River reservoirs causes the ensembled hydrology and ensembled climate to approach each other.
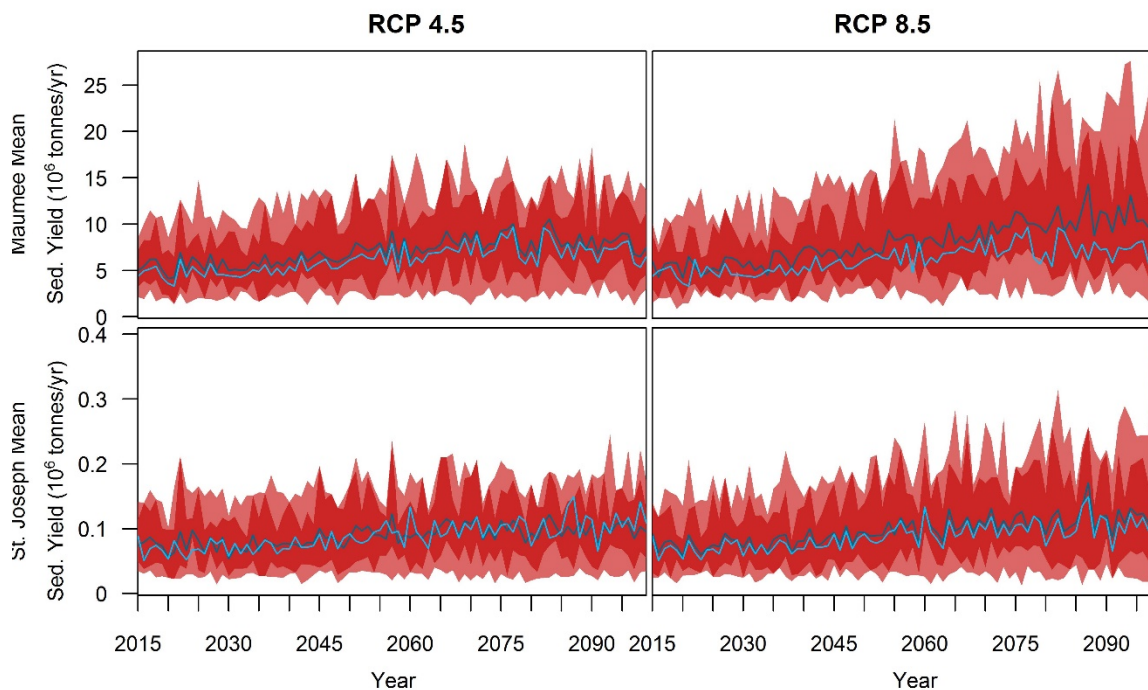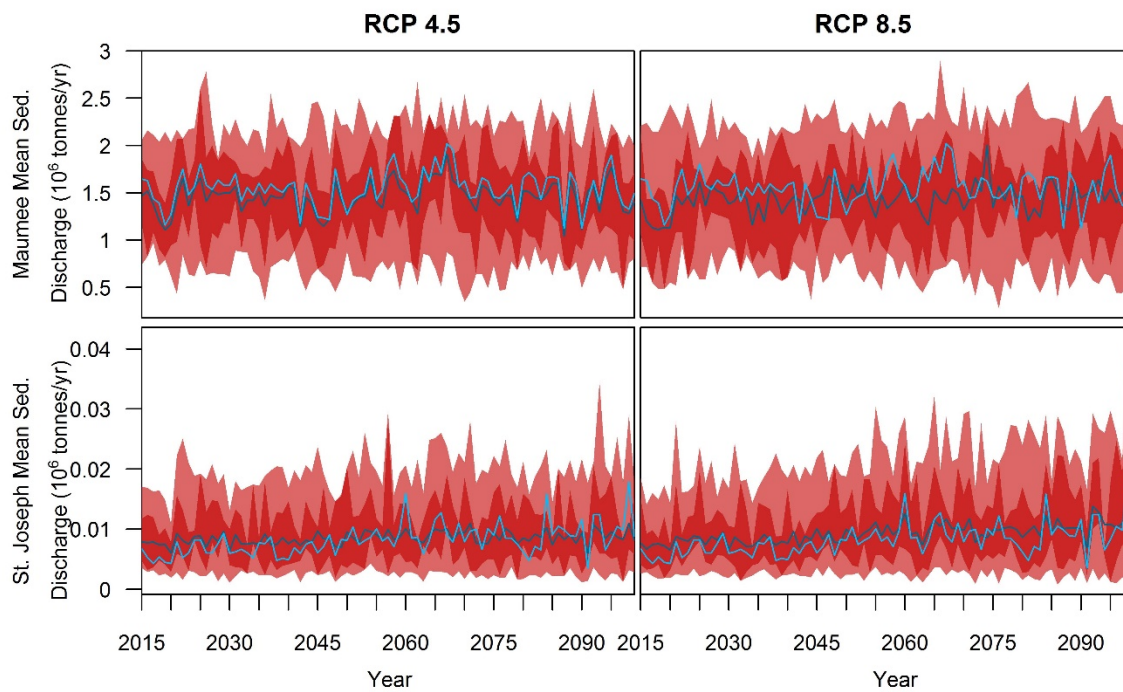

Supplementary Figure S4. Sediment yields based on climate inputs from the CSIRO mk3.6 global climate model.
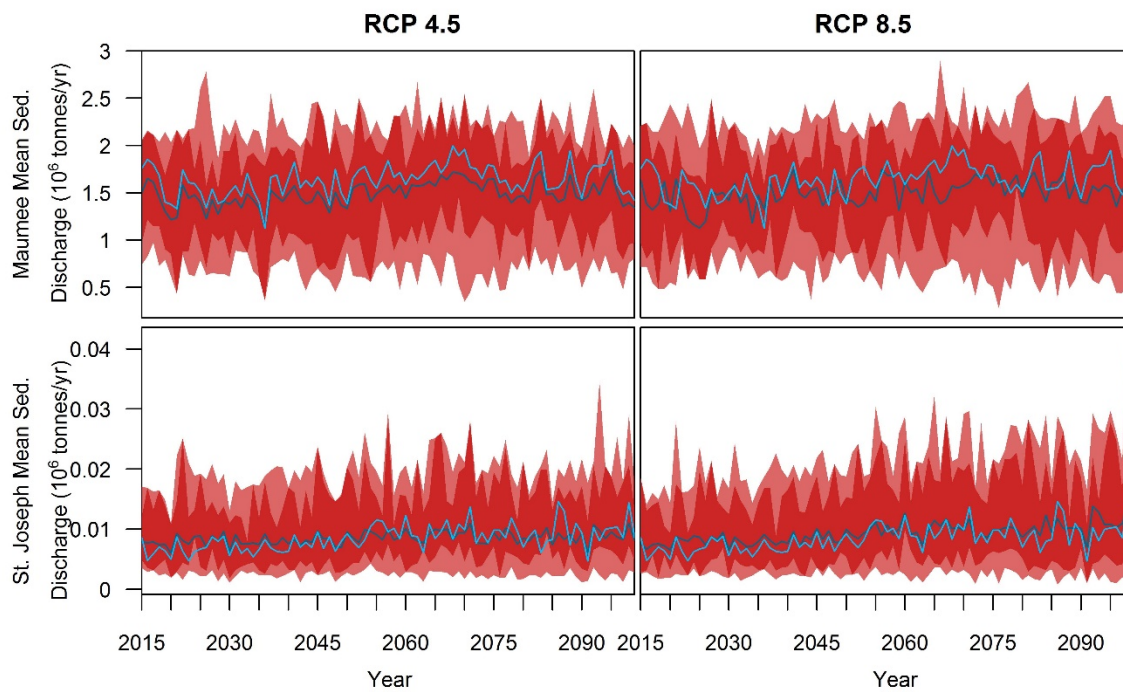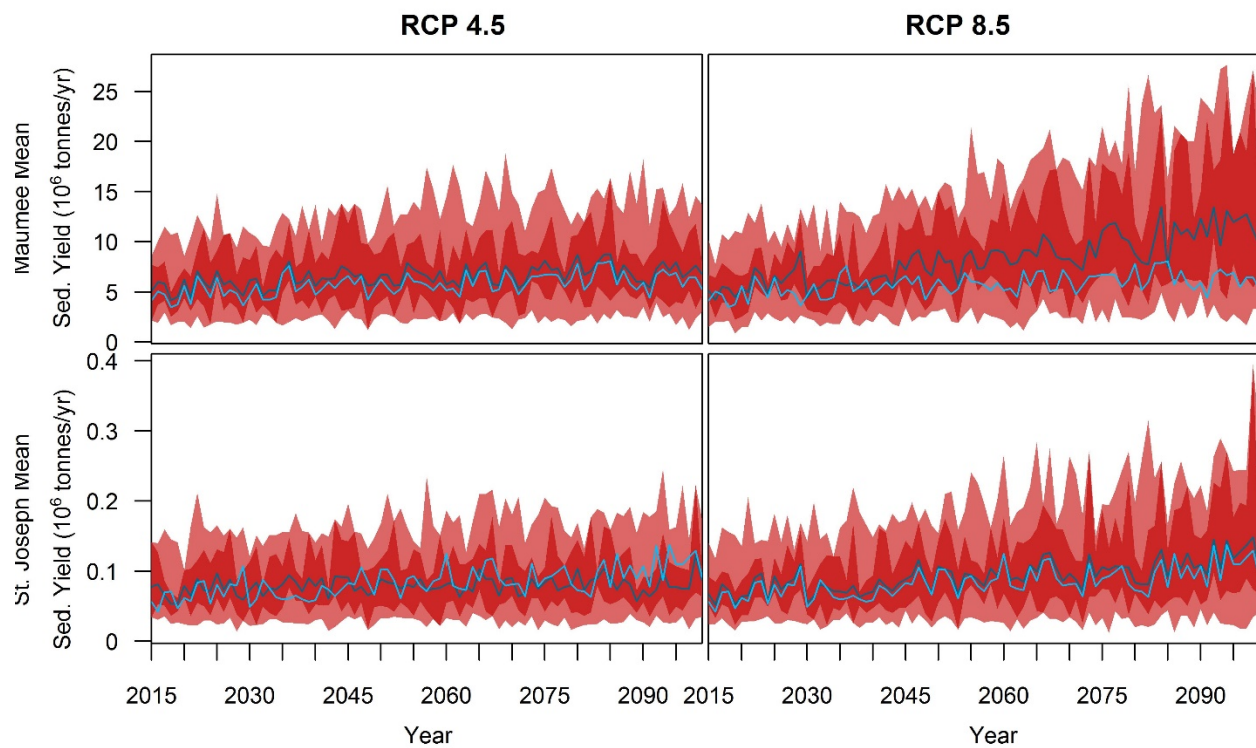
Supplementary Figure S5. Sediment discharge based on climate inputs from the CSIRO mk3.6 global climate model.
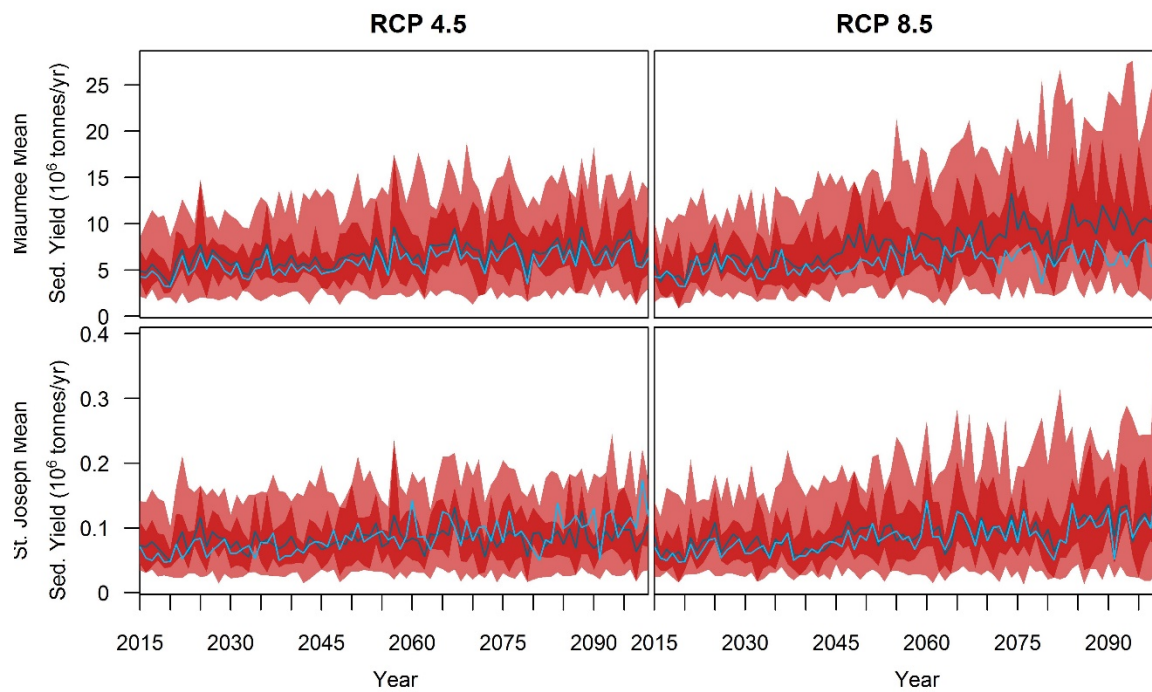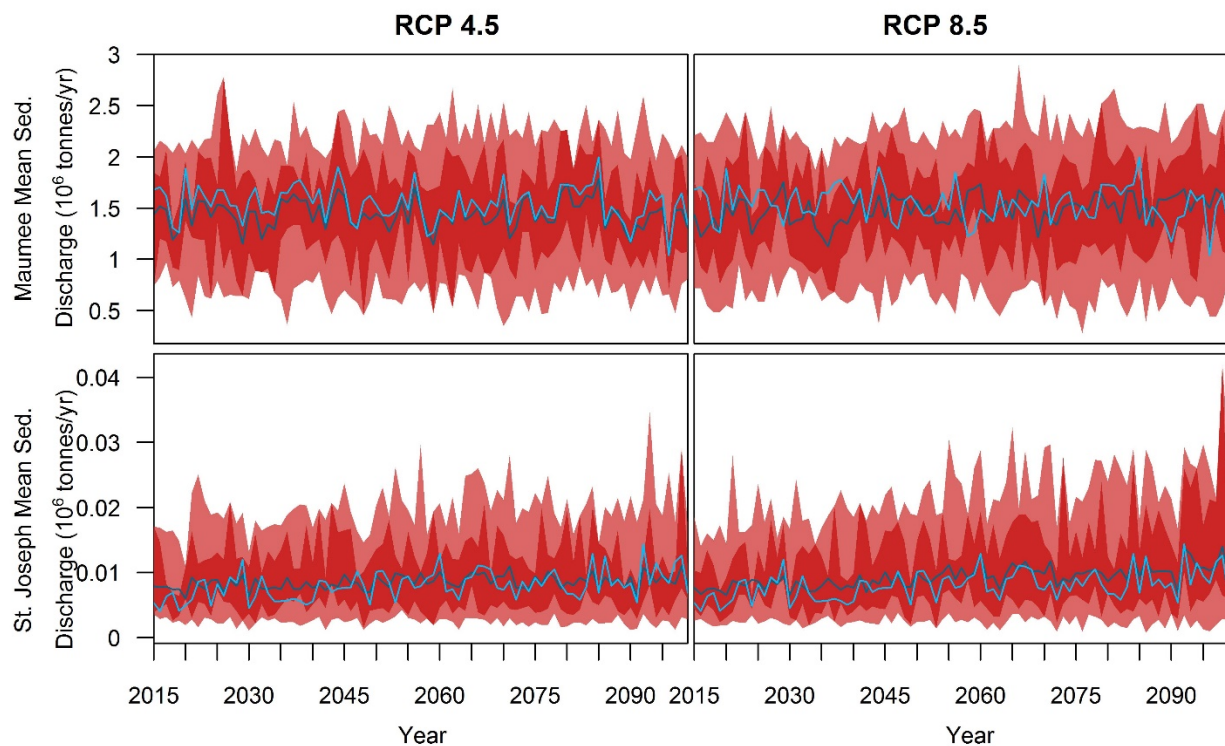
Supplementary Figure S6. Sediment yields based on climate inputs from a representative selection of available climate ensemble members.

Supplementary Figure S7. Sediment discharge based on climate inputs from a representative selection of available climate ensemble members.

Supplementary Figure S8. Sediment yields based on climate inputs from the best fit climate ensemble members.

Supplementary Figure S9. Sediment discharge based on climate inputs from the best fit climate ensemble members.

Supplementary Table S5. The ensembled climate sediment yield is consistently biased lower than the ensembled hydrology, regardless of the choice of ensemble members.

	RCP	Ensemble Method	Maumee Total Annual Sediment Yield (tonnes)	St. Joseph Total Annual Sediment Yield (tonnes)
All Members	4.5	Hydrology	6,434,215	80,081
		Climate	5,534,487	71,464
	8.5	Hydrology	7,699,588	90,527
		Climate	6,747,477	80,397
CSIRO mk3.6	4.5	Hydrology	7,044,665	89,116
		Climate	6,214,455	81,784
	8.5	Hydrology	8,085,969	97,351
		Climate	7,250,606	88,792
Representative	4.5	Hydrology	6,570,391	82,052
		Climate	5,865,431	76,019
	8.5	Hydrology	7,757,722	90,552
		Climate	7,113,726	85,084
Best Fit	4.5	Hydrology	6,506,819	79,372
		Climate	5,749,773	72,359
	8.5	Hydrology	8,239,787	93,046
		Climate	7,372,572	84,225

Supplementary Table S6. The ensembled climate sediment discharge is consistently biased lower than the ensembled hydrology for the St. Joseph River, regardless of the choice of ensemble members. This difference is reversed for the Maumee River, where the ensembled hydrology produces lower annual sediment discharges than the ensembled climate.

	RCP	Ensemble Method	Maumee Total Annual Sediment Discharge (tonnes)	St. Joseph Total Annual Sediment Discharge (tonnes)
All Members	4.5	Hydrology	1,433,531	8,519
		Climate	1,588,889	6,929
	8.5	Hydrology	1,454,865	9,312
		Climate	1,573,345	7,568
CSIRO mk3.6	4.5	Hydrology	1,488,785	9,556
		Climate	1,637,137	8,095
	8.5	Hydrology	1,503,671	10,034
		Climate	1,614,038	8,499
Representative	4.5	Hydrology	1,461,130	8,685
		Climate	1,545,479	7,464
	8.5	Hydrology	1,436,010	9,239
		Climate	1,537,315	8,178
Best Fit	4.5	Hydrology	1,440,338	8,350
		Climate	1,542,808	7,133
	8.5	Hydrology	1,487,714	9,619
		Climate	1,573,732	8,046