Detection of climate change impacts on projected water resource system performance

Matthew Chen, Jonathan D. Herman

1 Abstract

Simulation modeling informed by downscaled GCM ensemble predictions coupled with hydrologic modeling is commonly used to gauge future effects of climate change on water resource systems performance. However, climate models are highly uncertain which is reflected by significant variability in simulation results. In this study, the performance of flooding and reliability objectives in the Sacramento-San Joaquin system projected until the year 2100 were tested against the historical period (1951-2000) using non-parametric methods to determine when a significant change can be detected. Model scenarios and subsequent significant detections were sorted by different climate and land-use scenarios and by GCM. Two approaches to analyzing significant detections were implemented: first, finding the proportion of the entire ensemble that detected a significant change before 2100 and second, studying the distribution of first detection years of individual scenarios. It was found that the reliability objective was far more likely to exhibit a significant change within the century, whereas detectable change in flood volume occurs after the year 2100. These results can be used to inform water resource planning management decisions, such as in a dynamic policy-tree algorithm.

2 Introduction

Water resources may be the most impacted natural resource by climate change and atmospheric warming (Minville et al., 2008). Climate change potentially increases both drought severity and duration and increases flood risk by altering natural hydrologic cycles. The most common physically grounded method of studying such impacts on the regional scale is accomplished through analyzing global climate model (GCM) outputs to predict changes in hydrologic variables (Karmalkar et al., 2019). However, these methods suffer from “deep uncertainty,” where many sources of uncertainties can propagate through modeling efforts at multiple levels including but not limited to the GCM, GCM downscaling, greenhouse gas emissions scenario, and natural variability (Chen et al., 2011; Lebel et al., 2000; Minville et al., 2008; Shen et al., 2018; Trinh et al., 2017).

Given the nature of such uncertainty, studies have attempted to statistically detect and characterize long term trends in GCM-grounded results to gauge possible climate change impacts on hydrologic variables. For example, (Mcbean & Motiee, 2006; Xu et al., 2004) have identified possible long-term climate change driven trends applying the nonparametric Mann-Kendall test to the Tarim River basin in China and the Great Lakes in the United States, respectively.

Similarly, this study seeks to detect climate change impacts through statistical testing and aims to study patterns in the temporal distribution of when detections occur. Instead of studying long-

term trends, rolling windows of 30 years were repeatedly tested against the historical baseline to obtain a year-by-year analysis. Furthermore, testing is conducted on system performance objectives (flood volume, reliability) rather than raw hydrologic variables such as precipitation, temperature, and streamflow to characterize climate impacts on water resources planning and management.

3 Methods

3.1 Non-parametric significance detection

Projected reservoir system objectives of cumulative flood volume and reliability south of the San Joaquin Delta were evaluated for 3492 scenarios for the historical period (1951-2000 water years) up to the year 2100 in the future. Reliability was obtained through the simulation modeling of estimated water demand and releases and calculated as the percent time in which releases met or surpassed demand. Cumulative flood volume was found relative to the aforementioned releases and downstream levee capacities. Scenarios were defined by a choice of 31 global circulation models (GCM) and 4 representative concentration pathways (RCP 2.6, 4.5, 6.0, and 8.5) that make up CMIP5 climate scenarios. Climate projections were then downscaled and adapted to a hydrologic simulation model to estimate system objectives. 36 different land use scenarios (LULC) were considered for each GCM/RCP pair to estimate water demand. Finally, 30 year moving averages were considered across the dataset to reduce the effect of natural variability and noise.

To detect non-parametric significance, a one-sided Mann Whitney U-test was conducted progressively for each rolling window against the historical sample distribution. At $p<0.05$, it was said that a significant change in the system objective was detected with 95% confidence. A significant change is defined as the moment the tested window and the historical window are hypothesized to be from different distributions. Furthermore, it was found that the ensuing results are not sensitive to the choice of window size in taking the moving averages. Scenarios that never detected a significant change in the projection window were deemed “no detect scenarios,” where detection is thought to possibly occur after the year 2100 given ongoing climate change impact.

3.2 Single Scenario Analysis

The single scenario analysis sought to consider the first significant detection year across each of the 3492 scenarios, combinations of choices in GCM, RCP, and LULC, individually. The advantage of this methodology is that it provides a direct link between a specific choice in GCM, RCP, and LULC and when significant detection occurs in the projection, which may provide insight in how these choices control results. The resulting distribution of first-detection years was then analyzed in total, including counts of “no detect scenarios.” Results are also sorted by

GCM, RCP, or land use scenario choice. Descriptive statistics for each scenario choice and type was also computed, including the median, standard deviation, and sample size. Pathways incurring a significant detection were also mapped to its actual objective severity at the end of the projection period to gauge if the most severe scenarios are likely to go undetected.

3.3 Multiple Scenario Analysis

The multiple scenarios analysis was concerned with significance detection across the ensemble in timeseries. This analysis characterized how the relative count of significant detection for each year, or the ratio of scenarios incurring a significant detection and those that did not change with time. In essence, if each characterization of the future and every combination of scenarios, is equally likely to occur, the multiple scenario analysis represents an idealized measure of detection probability with time within the context of the sample population of defined possible scenarios included in the analysis. It is important to note that every scenario and model choice only seeks to describe the future, thus outcomes for each scenario are possible, but the model ensemble does not represent a collection of all possible futures. Results were then sorted by GCM, RCP, and LULC's, with relative counts computed relative to the scenarios within a set GCM, RCP, or LULC. Since detections climbed steadily with time, the distribution of relative counts at the end of the projection period (2098) was also considered.

3.4 Global Sensitivity Analysis

Since model scenarios show significant variability, it is interesting to characterize that variability with its input variables. The first, second, and total order global sensitivity of first detection years for choices in different GCM, LULC, or RCPs was also formally computed using Sobol sensitivity analysis, providing a measure of how much variability in results is provided by each parameter both independently and through interactions with other parameters.

4 Results

4.1 Simulation Model Projections

Figure 1 below plots the projected system objectives for all scenarios. For reliability, it is apparent that the mean of all scenarios is decreasing, from about 0.9 at the start of the projection to 0.7 at the end, though variability is also increasing with time. For flood volume, the mean of all scenarios remains relatively constant, though variability is also increasing with time as with reliability. It is difficult to discern whether flood volume is predicted to increase or decrease.

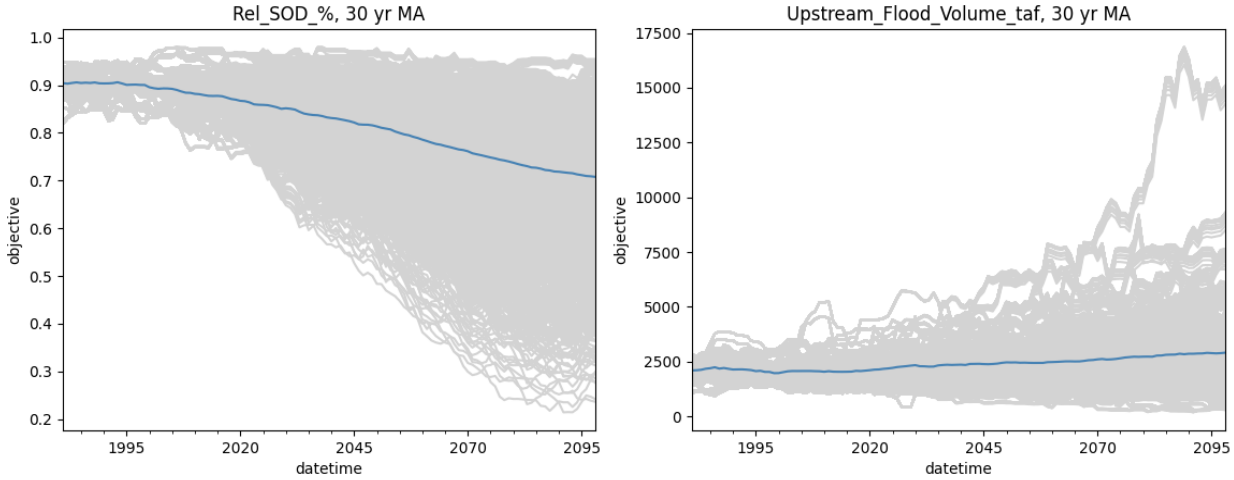

Figure 1. *30-year moving averages of ensemble projections of water supply reliability (left) and cumulative flood volume (right) for 3492 GCM/RCP/LULC scenarios. The blue line represents the mean of all scenarios.*

4.2 Single Scenario Analysis

For reliability, the distribution of first detection years includes 110 or 0.3% no detects out of the sample size of 3492 scenarios, which are not plotted in Figure 2. The distribution is also heavily right skewed, with many detections occurring earlier in the century. In all, most scenarios show a significant detection, and detections are more likely to happen earlier. For flood volume, the distribution includes 2274 or 65% no detects out of the sample size of 3492. The distribution is also right skewed, though there is also a jump in detections at the end of the century. In all, the majority of scenarios did not detect a significant change in the projection period, and the ones that did are more likely to detect at either early or late extremes.

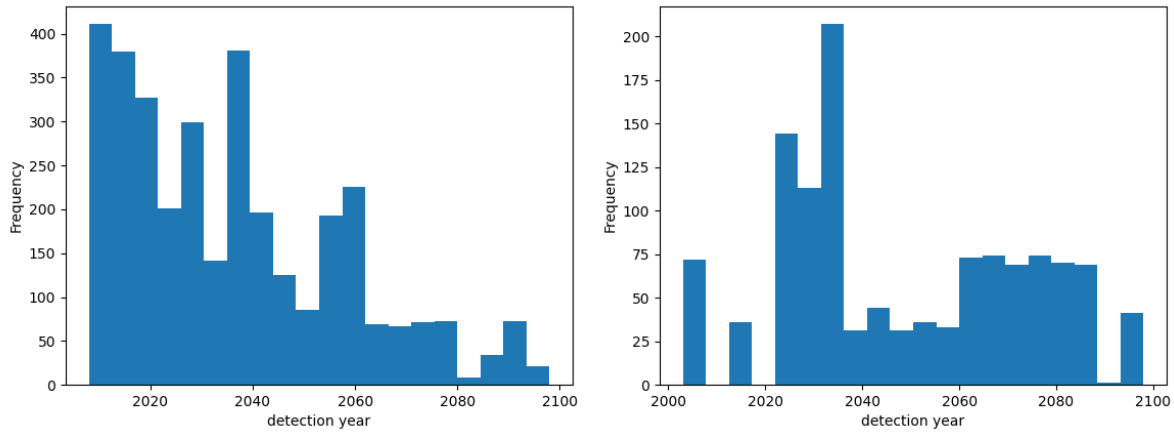

Figure 2. *Distributions of first detection years for reliability (left) and flood volume (right) for all scenarios. No-detect scenarios are not shown.*

When first detection years are sorted by GCM's, RCP's, and LULC's respectively for reliability (Table 1), it is apparent that GCM's are the most variable in both the median, as first detections occur all over the century, and the standard deviation. For LULC, though standard deviations are somewhat larger than from the GCM sorting, the median exhibits a much smaller range centered on about the year 2033. Interestingly, RCP's, which represents the varying effects of climate change, show increasing medians as the RCP number increases.

Table 1. *Distributions of first detection years sorted by choice of GCM's, LULC's and RCP's for reliability.*

GCM	Median	Standard deviation	Sample size
access1-0	2024	21.2	72
bcc-csm1-1	2013	2.5	144
bcc-csm1-1-m	2057	8.8	72
canesm2	2018	18.6	108
ccsm4	2033.5	15.5	144
cesm1-bgc	2053	8.2	72
cesm1-cam5	2020.5	9.5	144
cmcc-cm	2059	10.0	72
cnrm-cm5	2075	9.1	72
csiro-mk3-6-0	2030	13.5	144
fgoals-g2	2062	9.7	108
fio-esm	2087	13.5	77
gfdl-cm3	2030	9.3	144
gfdl-esm2g	2015.5	11.8	144
gfdl-esm2m	2012.5	4.7	144
giss-e2-h-cc	2027	0.0	36
giss-e2-r	2085	11.4	105
giss-e2-r-cc	2059	5.3	36
hadgem2-ao	2027.5	12.3	144
hadgem2-cc	2025	21.7	72
hadgem2-es	2027	8.6	144
inmcm4	2052.5	18.3	68
ipsl-cm5a-mr	2016	11.5	144
ipsl-cm5b-lr	2011.5	1.5	72
miroc5	2038	8.4	144
miroc-esm	2032	11.5	144
miroc-esm-chem	2034	11.9	144
mpi-esm-lr	2040	15.6	108
mpi-esm-mr	2035	11.9	108
mri-cgcm3	2020	9.6	108
noresm1-m	2032.5	12.9	144

LULC	Median	Standard deviation	Sample size
FORE-SCE-A1B	2032	17.5	97
FORE-SCE-A2	2030	21.5	96
FORE-SCE-B1	2037	21.6	95
GCAM-26	2029	12.2	97
GCAM-45	2029	14.1	97
GCAM-ref	2030	16.2	97
LUCAS-BAU_High-0001	2033	22.9	94
LUCAS-BAU_High-0002	2034	21.7	94
LUCAS-BAU_High-0003	2031	21.3	94
LUCAS-BAU_High-0004	2034	21.8	94
LUCAS-BAU_High-0005	2033	22.3	94
LUCAS-BAU_High-0006	2032	21.6	94
LUCAS-BAU_High-0007	2031	21.7	94
LUCAS-BAU_High-0008	2033	22.9	94
LUCAS-BAU_High-0009	2033	22.2	94
LUCAS-BAU_High-0010	2034	22.0	94
LUCAS-BAU_Low-0001	2033	22.0	91
LUCAS-BAU_Low-0002	2032.5	22.5	92
LUCAS-BAU_Low-0003	2033	23.3	93
LUCAS-BAU_Low-0004	2035	23.3	93
LUCAS-BAU_Low-0005	2035	23.3	93
LUCAS-BAU_Low-0006	2033	22.1	91
LUCAS-BAU_Low-0007	2033	23.0	94
LUCAS-BAU_Low-0008	2033	22.4	91
LUCAS-BAU_Low-0009	2034	22.3	94
LUCAS-BAU_Low-0010	2033	21.9	91
LUCAS-BAU_Med-0001	2034	22.9	94
LUCAS-BAU_Med-0002	2035	22.4	94
LUCAS-BAU_Med-0003	2033	22.9	94
LUCAS-BAU_Med-0004	2035	23.7	94
LUCAS-BAU_Med-0005	2034	22.3	94
LUCAS-BAU_Med-0006	2032.5	22.1	94
LUCAS-BAU_Med-0007	2033	22.1	94
LUCAS-BAU_Med-0008	2032.5	22.1	94
LUCAS-BAU_Med-0009	2034	22.1	94
LUCAS-BAU_Med-0010	2032.5	22.3	94

RCP	Median	Standard deviation	Sample size
2.6	2024	15.2	693
4.5	2029	23.9	1099
6.0	2030	19.8	546
8.5	2040	21.8	1044

Sorting as above was also conducted for flood volume (Table 2). 15 out of 31 GCM's did not detect a significant change in the projection period, and the variability of the median first detection year for those with detections is high. Results for standard deviations is similar to the reliability results, with a range of 0.2 to 30.3 years despite relatively high sample sizes. In contrast, LULC again shows little effect on median, centered on 2042, and standard deviations are high but consistent. RCP's exhibit a similar pattern as with reliability, although the 2.6 pathway detected much later in the projection.

The above results show that the choice in GCM contributes much variability in when detections occur, which can be further confirmed and demonstrated with a formalized sensitivity analysis.

Table 2. *Distributions of first detection years sorted by choice of GCM's, LULC's and RCP's for cumulative flood volume. Categories with no detections are highlighted.*

GCM	Median	Standard deviation	Sample size
access1-0			0
bcc-csm1-1			0
bcc-csm1-1-m	2023.5	1.5	72
canesm2	2067	4.5	72
ccsm4	2095	0.2	36
cesm1-bgc	2052	30.3	72
cesm1-cam5			0
cmcc-cm			0
cnrm-cm5	2031	1.5	72
csiro-mk3-6-0			0
fgoals-g2	2006	4.2	108
fio-esm	2030	21.1	140
gfdl-cm3	2084	0.5	34
gfdl-esm2g	2040	0.8	36
gfdl-esm2m			0
giss-e2-h-cc	2058	2.2	36
giss-e2-r	2033	5.3	144
giss-e2-r-cc			0
hadgem2-ao			0
hadgem2-cc	2066	0.9	72
hadgem2-es			0
inmcm4	2051.5	25.8	72
ipsl-cm5a-mr	2062	25.2	72
ipsl-cm5b-lr			0
miroc5			0
miroc-esm			0
miroc-esm-chem			0
mpi-esm-lr			0
mpi-esm-mr	2062	15.7	108
mri-cgcm3	2065	12.4	72
noresm1-m			0

LULC	Median	Standard deviation	Sample size
FORE-SCE-A1B	2042	27.0	33
FORE-SCE-A2	2043	26.4	34
FORE-SCE-B1	2042	25.1	34
GCAM-26	2043.5	26.9	32
GCAM-45	2040	26.9	32
GCAM-ref	2042	27.1	33
LUCAS-BAU_High-0001	2042	25.1	34
LUCAS-BAU_High-0002	2042	25.1	34
LUCAS-BAU_High-0003	2043	25.1	34
LUCAS-BAU_High-0004	2042	25.1	34
LUCAS-BAU_High-0005	2042	25.1	34
LUCAS-BAU_High-0006	2042	25.1	34
LUCAS-BAU_High-0007	2043	25.1	34
LUCAS-BAU_High-0008	2042	25.1	34
LUCAS-BAU_High-0009	2042	25.1	34
LUCAS-BAU_High-0010	2042	25.1	34
LUCAS-BAU_Low-0001	2042	24.9	34
LUCAS-BAU_Low-0002	2042	24.9	34
LUCAS-BAU_Low-0003	2042	25.0	34
LUCAS-BAU_Low-0004	2042	24.9	34
LUCAS-BAU_Low-0005	2042	25.0	34
LUCAS-BAU_Low-0006	2042	24.9	34
LUCAS-BAU_Low-0007	2042	25.0	34
LUCAS-BAU_Low-0008	2041.5	24.9	34
LUCAS-BAU_Low-0009	2042	25.1	34
LUCAS-BAU_Low-0010	2042	25.0	34
LUCAS-BAU_Med-0001	2042	25.1	34
LUCAS-BAU_Med-0002	2042	25.1	34
LUCAS-BAU_Med-0003	2042	25.1	34
LUCAS-BAU_Med-0004	2042	24.9	34
LUCAS-BAU_Med-0005	2042	25.1	34
LUCAS-BAU_Med-0006	2042	25.1	34
LUCAS-BAU_Med-0007	2042	25.1	34
LUCAS-BAU_Med-0008	2043	25.0	34
LUCAS-BAU_Med-0009	2042	25.1	34
LUCAS-BAU_Med-0010	2042	25.1	34

RCP	Median	Standard deviation	Sample size
2.6	2047	25.8	180
4.5	2033	22.3	428
6.0	2044	28.1	108
8.5	2048	25.0	502

4.3 Multiple Scenario Analysis

The multiple scenario analysis, presented as relative counts between scenarios with detection and the total number of relevant scenarios, is shown below (Figure 3). Considering the entire ensemble, it is abundantly clear that the reliability objective showed far greater detection rates

compared with flood volume. For reliability, 89.8% of scenarios showed a detection at the end of the projection period whereas only 13.3% scenarios for flood volume show significant change. Additionally, detections climb steadily throughout the projection for both scenarios, which is evidence that scenarios are unlikely to fall out of detection once detection occurs. Additionally, relative counts for reliability show less year-to-year variability.

Multiple scenario relative counts at the end of the projection period (2098) were also sorted by choice of GCM, RCP, and LULC (Figure 3). For reliability, there is little variability in results for RCP and LULC choice and all relative counts are greater than 80%. In other words, this shows that 80% or more of scenarios containing any given RCP or LULC will detect a significant change at the end of the projection period. Relative counts are also high for GCM choice, though there are pockets of outliers with low detection rates. For flood volume, relative counts are consistently low (20% or less) for RCP and LULC choices, and while GCM counts are low there are outliers with high detection rates. These results are consistent with findings from the single scenario analysis.

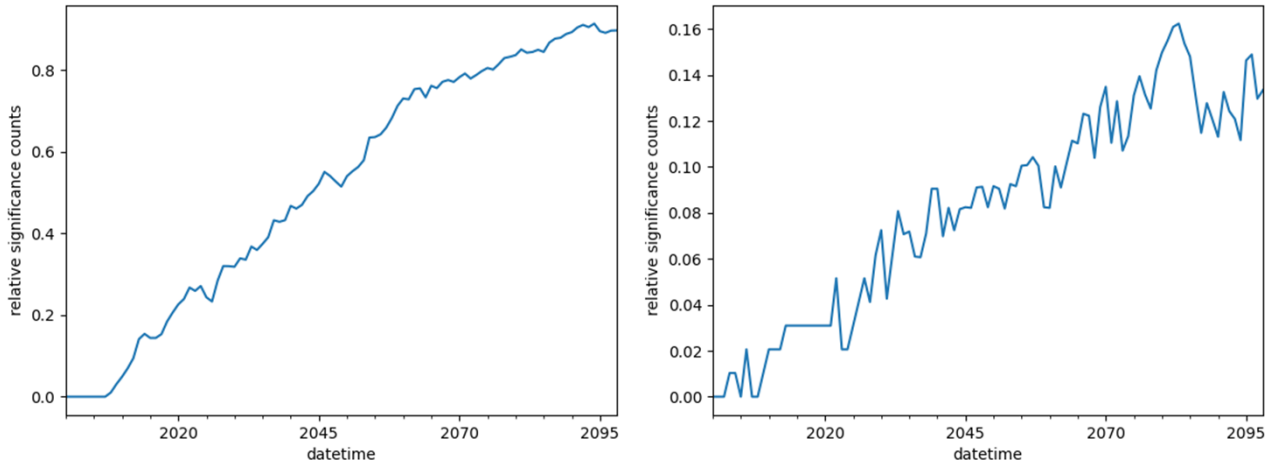

Figure 3. *Annual relative counts of significant detections for the entire ensemble, with the reliability objective shown on the right and flood volume shown on the left*

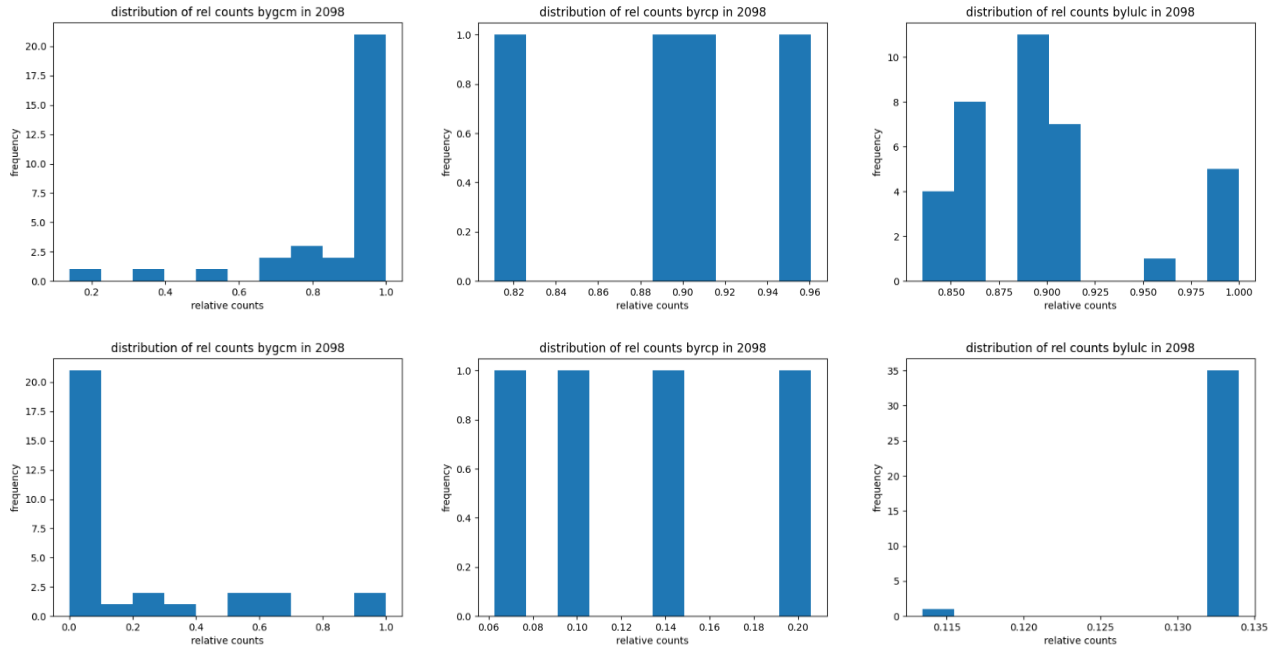

Figure 4. *Distribution of relative detection counts at the end of the projection period (2098) sorted by GCM/RCP/LULC for reliability (top) and flood volume (bottom)*

4.4 Global Sensitivity

The Sobol sensitivity analysis confirms the result from the multiple and single scenario analysis that GCM contributes to the greatest variability in results (Table 3). For total order sensitivity, choice of GCM contributes a Sobol sensitivity index of 1.01 and 1.035 for reliability and flood volume, respectively. RCP, which represents the effects of climate change is next in importance, but only through its interaction with GCM choice. Its total order sensitivity is 0.296 and 0.284, respectively for reliability and flood volume objectives. However, RCP has essentially no first order contribution for reliability and little for flood volume (GCM is notably higher in both), but through its iteration with GCM it carries the highest second order sensitivity index. Bootstrap confidence intervals for the sensitivity indices are reasonably small, at less than 10% except for the second order GCM/LULC interaction for reliability.

Table 3. *Sobol sensitivity analysis*

Rel_SOD_%

	ST	ST_conf	S1	S1_conf		S2	S2_conf
rcp	0.296	0.038	-0.009	0.052	**rcp, gcm**	0.264	0.069
gcm	1.010	0.072	0.706	0.088	**rcp, lulc**	0.018	0.071
lulc	0.052	0.014	-0.003	0.017	**gcm, lulc**	0.042	0.142

Upstream_Flood_Volume_taf

	ST	ST_conf	S1	S1_conf		S2	S2_conf
rcp	0.284	0.036	0.043	0.037	**rcp, gcm**	0.232	0.065
gcm	1.036	0.064	0.767	0.079	**rcp, lulc**	-0.028	0.059
lulc	0.006	0.005	0.002	0.007	**gcm, lulc**	-0.006	0.092

4.5 Objective Severity

It was also asked if there was a relationship between first detection year and the objective projection at the end of the detection for the same scenario (Figure 5). For reliability, earlier detection years have a wide range of possible outcomes, but later detection years occur for predominately less severe scenarios. Thus, severe scenarios are more likely to be detected earlier than later and are not likely to go undetected. This is confirmed by distribution of reliability for no-detect scenarios, which are centered and concentrated on high values. For flood volume, there is little relationship between detection year and severity. However, as shown previously, most scenarios show no detection, and these objectives are concentrated on lower values.

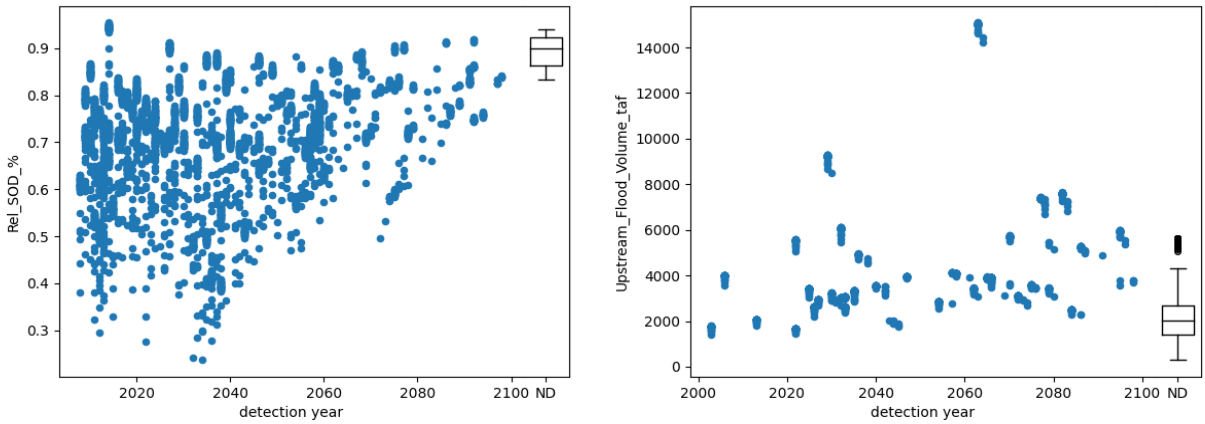

Figure 5. *Detection year plotted against reliability (left) and flood volume (right) at the end of the projection period. "ND" represents the distribution of no-detect scenarios*

4.6 Historical Objectives

Since GCM contributes to the greatest variability in projected outcomes, it was asked if the models predict true historical objectives accurately (Figure 6). This was done by comparing model projections to objectives derived from observed inflows from 1997 to 2020. It appears that the central tendency of historical reliability lines up well with modeled values, though flood volume is severely underpredicted which is also reflected in the predicted vs. observed peak inflow result. Variability for all objectives is high.

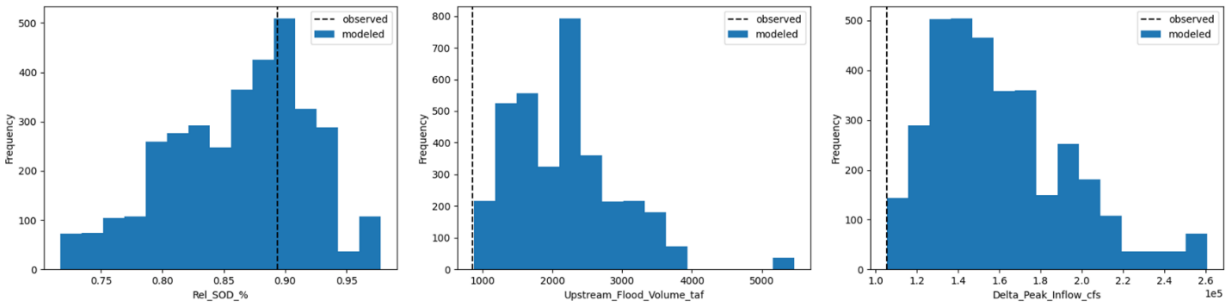

Figure 6. *Mean objectives from historical model projections vs. mean objectives from historical observations in 1997-2020*

5 Discussion and Conclusion

Both the single and the multiple scenario analysis show that the reliability objective is far more likely to exhibit a significant change within the projection period as shown by a super-majority of scenarios in the context of this study. Additionally, the single scenario analysis showed that detections are more likely to occur earlier than later. Flood volume, on the other hand, was dominated by no-detect scenarios. For these scenarios, it is thought that detection would occur after the year 2100 as climate driven impacts grow more severe. These results are useful to consider in the context of adapting water resources planning and management to climate change in the next century and vaguely describes how decisions should be timed. It was also shown that severe scenarios are not likely to go undetected, which may be a useful if not reassuring result for practice.

However, perhaps more important than the implications for engineering practice, are the driving factors behind a "significant detection" in the context of this study. It remains unclear whether or not detections are driven by climate change as reflected in the choice of the RCP scenario, or by model variability. For example, while the median first detection year was generally increasing with more severe RCP pathways, it was also shown that more distant detections correlate with less severe scenarios. If climate change severity were to correlate with more severe scenarios, the opposite could be expected. That is, that median first detection year decreases with increasing climate severity.

One explanation for these seemingly contradictory results is the variability contributed by GCM choice. It was shown that the choice of GCM dominated global sensitivity of first detection year – and contributes to the greatest variability in results either on its own or through its interaction with RCP. Thus, it becomes almost impossible to separate fundamental modeling error and variability from significant detections. A clear example of this is in the flood volume objective, where it was shown that models severely underpredicted historical flood volume – which may be related to the fact that most scenarios did not detect a significant change for flooding. The role of climate change on detection time becomes difficult to see. For reliability, it can be argued that the hydrologic models, which are driven by GCM inputs, predicted central tendency of the mean historical objective. However, these results ranged from 70-100% reliability, which represents a wide variability for a 23-year long comparison. Until modeling error can be reduced and results from different GCM's become more consistent, even statistically significant detections in system objectives will be clouded by variability from other sources, and the true effect of climate change on detection time may become difficult to interpret.

6 References

Chen, J., Brissette, F. P., & Leconte, R. (2011). Uncertainty of downscaling method in quantifying the impact of climate change on hydrology. *Journal of Hydrology*, *401*(3–4), 190–202. https://doi.org/10.1016/j.jhydrol.2011.02.020

Karmalkar, A. V., Thibeault, J. M., Bryan, A. M., & Seth, A. (2019). Identifying credible and diverse GCMs for regional climate change studies—case study: Northeastern United States. *Climatic Change*, *154*(3–4), 367–386. https://doi.org/10.1007/s10584-019-02411-y

Lebel, T., Le Barbé, L., Delclaux, F., & Polcher, J. (2000). From GCM scales to hydrological scales: Rainfall variability in West Africa. *Stochastic Environmental Research and Risk Assessment*, *14*(4–5), 275–295. https://doi.org/10.1007/s004770000050

Mcbean, E., & Motiee, H. (2006). Assessment of impacts of climate change on water resources – a case study of the Great Lakes of North America. *Hydrology and Earth System Sciences Discussions*, *3*(5), 3183–3209. https://doi.org/10.5194/hessd-3-3183-2006

Minville, M., Brissette, F., & Leconte, R. (2008). Uncertainty of the impact of climate change on the hydrology of a nordic watershed. *Journal of Hydrology*, *358*(1–2), 70–83. https://doi.org/10.1016/j.jhydrol.2008.05.033

Shen, M., Chen, J., Zhuan, M., Chen, H., Xu, C. Y., & Xiong, L. (2018). Estimating uncertainty and its temporal variation related to global climate models in quantifying climate change impacts on hydrology. *Journal of Hydrology*, *556*, 10–24. https://doi.org/10.1016/j.jhydrol.2017.11.004

Trinh, T., Ishida, K., Kavvas, M. L., Ercan, A., & Carr, K. (2017). Assessment of 21st century drought conditions at Shasta Dam based on dynamically projected water supply conditions by a regional climate model coupled with a physically-based hydrology model. *Science of the Total Environment*, *586*, 197–205. https://doi.org/10.1016/j.scitotenv.2017.01.202

Xu, Z. X., Chen, Y. N., & Li, J. Y. (2004). Impact of climate change on water resources in the Tarim River basin. *Water Resources Management*, *18*(5), 439–458. https://doi.org/10.1023/B:WARM.0000049142.95583.98