1. Summary

This paper runs through a projection of floods for a basin in China, using downscaled GCM data to force a hydrologic model. The paper would be strengthened by increasing the details in some areas, noted below, and by expanding on the results. As the results section reads now, the contribution of this paper is that there is a lot of variability between the models. Is there is some way to take information between the models to have a more robust understanding of how floods are going to change? Or is the overarching conclusion is that these models are not in enough agreement to say anything?

We would like to sincerely thank the anonymous reviewer for her/his valuable comments. These comments have help improve the manuscript and clarify the focus of the paper. It is really true as the Reviewer pointed that the paper would be strengthened by increasing the details in some areas. We followed the suggestions and answered accordingly below.

2. Major points

1) It is unclear how the model calibration and validation was done (Figure 2). How do they have observed runoff? Is it discharge divided by basin area? How was the model routed? No mention of a routing model is made in the text.

Response: We are very sorry for our negligence of information about model calibration and validation and routing model. In our study, the observed runoff data was obtained from the Hengshi hydrological station, which is the discharge station of the study area (located at the outlet of basin, as shown in Figure 1 in the manuscript). Measurement of the discharge of water from the study area was made by the Hengshi hydrological station. Thus, we validated the VIC model by comparing simulated discharges against the observations at the Hengshi hydrological station. In addition, the Dag Lohmann model is used as the routing model, which transports the grid cell surface runoff and baseflow produced by VIC within each grid cell to the outlet of that grid cell and then into the river system. According to the Reviewer's suggestion, we have added the relevant information in Sections 2.1, 2.4 and 3.1 in the manuscript.

2) I found the text vague about the GCM downscaling methodology. The introduction laid out several different possibilities. I believe what was done was a simple interpolation of monthly values followed by bias correction through quantile mapping and then stochastic weather generation. If this is not the case, please modify the text. If it is the case, include some more details on the methodology so that the paper stands on its own. Also, if this is the case, I do not understand Section 3.3 saying it was driving the VIC model through historical resampling. The validation should be of the historical GCM runs downscaled identically to how the future runs are. I'm unsure if the problem here is due to the writing or the methodology.

Response: We thank the Reviewer for the comments. As the Reviewer stated, the GCM downscaling methodology used was an interpolation of monthly values followed by bias correction through quantile mapping and then stochastic weather generation.

The introduction of the GCM downscaling methodology in Section 2.3 is modified as follows:

"To generate local climate conditions (e. g. temperature and precipitation) from GCMs, a simple statistical downscaling method was performed as follows. Firstly, the model data and observed station data were interpolated to 0.25° resolution using bilinear interpolation. Secondly, the bias between the monthly precipitation and temperature of the observed and GCM output data was corrected using a quantile-based mapping method (Li et al. 2010) to reduce system errors in GCM simulations. Finally, a stochastic weather generation method was employed to temporally disaggregate the monthly corrected climate projections into the daily weather forcings required by the hydrological model. To consider the range of variability that this randomness could induce, multiple downscaling simulations were performed for each emissions scenario (Raff et al. 2009). The simulation set size of this study was arbitrarily set to ten simulations (i.e. ten downscaling samples)."

In Section 3.3, we made an evaluation of the flood simulation ability of each GCM. The VIC model was driven by 10 downscaling simulations for each GCM during the period 1970–2000. A comparison of the ECDFs between observed and simulated floods was then analysed (Figure 3c and d). According to the Reviewer's suggestion, we have modified the text in Section 3.3 to ensure better clarity in the understanding of GCM downscaling methodology.

Modified sections in the text: "The VIC model was driven by 10 downscaling simulations for each GCM during the period 1970–2000".

3) There's nothing about the flood regime of this basin as is: : : is it seasonal, etc.? It's hard to know what these changes mean and how much to care about this basin.

Response: We thank the Reviewer for the comments. We are very sorry for our negligence of information about the flood regime of this basin. The study basin is located in the tropical and subtropical climate zones, which have climate conditions favourable for the frequent occurrence of flood disasters in the flood season (e.g. sufficient precipitation and high humidity). Precipitation mostly occurs in the flood season (April–September), which accounts for approximately 70–80% of the annual precipitation. Due to climate warming, extreme rainfall events are recently occurring more frequently in the Feilaixia catchment, which leads to more intense and frequent flooding (e. g. the large floods in June and August 1994, June 1998, June 2005, and July 2006), causing extensive inundations and severe flood damage. It seriously threats to the flood control safety of Guangzhou city (one of the largest cities in South China) and other areas located in the downstream of the Beijiang basin. For example, the study region experienced the worst flood of the twentieth century in 1994, affecting two million people, and leading to the loss of RMB 3.2 billion (Wong and Zhao, 2001). It is therefore imperative to understand the projected changes in flood risk of this basin. According to the Reviewer's suggestion, we have added the relevant information in Section 2.1 in the revised manuscript.

4) Fitting 30 years of data to a 500-year return period (and 200-year) is not scientifically sound, and conclusions should not be made on these statistics. Adding just 5 more years of data to fitting the time series could dramatically alter their results; they are not robust.

Response: We thank the Reviewer for the comments. It is really true as the Reviewer pointed that fitting 30 years of data to a 500-year (200-year) return period is not scientific as these results are not robust. According to the Reviewer's suggestion, we removed the statistical analysis on 500-year and 200-year return periods from the manuscript. Meanwhile, we added a 20-year return period to further discuss the potential changes in the floods. For more detail information please see Table 2 below.

Table 2. Percentage changes (%) in AMX1d and AMX7fv under different scenarios (relative to the baseline period 1970–2000)

Flood index	Return period (a)	GCM	RCP2.6		RCP4.5		RCP8.5	
			T1	T2	T1	T2	T1	T2
AMX1d	100	BCC-CSM1.1	11.4	40.2	37.1	62.9	39.8	49.6
		CanESM2	-8.5	-1.2	91	42.7	25.2	19.7
		CSIRO-Mk3.6.0	32	6	-11.3	18.2	8	74.7
		GISS-E2-R	19.5	25.7	-0.5	12	-8.2	12.8
		MPI-ESM-LR	15	2.1	36.8	41.3	10.3	23
	50	BCC-CSM1.1	8.3	36.6	33.9	59.3	34.1	45.1
		CanESM2	-7.1	-0.4	77.4	39	23.3	18.4
		CSIRO-Mk3.6.0	26.7	8.1	-8.5	20.2	8.2	68.4
		GISS-E2-R	16.8	20.5	-0.5	8.7	-9.4	9.6
		MPI-ESM-LR	11.9	1.3	31.5	36.3	8	20.4
	20	BCC-CSM1.1	3.5	30.9	29.0	53.5	25.5	38.3
		CanESM2	-4.8	0.7	57.3	33.5	20.4	16.3
		CSIRO-Mk3.6.0	17.8	11.4	-4.0	23.3	8.4	57.6
		GISS-E2-R	12.9	12.9	-0.6	3.9	-10.9	4.9
		MPI-ESM-LR	7.5	0.2	23.5	28.8	4.6	16.3
AMX7fv	100	BCC-CSM1.1	8.6	50.3	35.8	58.1	27.6	57.8
		CanESM2	-5.2	-2.9	80.1	49.5	31.6	29.2
		CSIRO-Mk3.6.0	30.2	1.8	-16.8	15.2	9	71.8
		GISS-E2-R	14.5	29.5	0.9	18.7	-6.8	9.8
		MPI-ESM-LR	12.4	3.4	42.9	45.2	12.2	23
	50	BCC-CSM1.1	5.9	44.4	31.5	54.7	23.5	50.3

	CanESM2	-4.2	-1.7	68.5	44.7	28.2	26.1
	CSIRO-Mk3.6.0	25.1	4.8	-13.4	18	9.3	66
	GISS-E2-R	12.3	23.1	0.6	13.9	-8.4	7.3
	MPI-ESM-LR	10.6	2.2	36.1	38.6	9.2	20
	BCC-CSM1.1	1.7	35.5	25.2	49.3	17.3	39.2
	CanESM2	-2.7	0.0	51.4	37.6	23.0	21.4
20	CSIRO-Mk3.6.0	16.7	9.6	-7.9	22.5	9.7	55.9
	GISS-E2-R	9.1	14.0	0.1	7.0	-10.6	3.6
	MPI-ESM-LR	8.0	0.4	26.2	28.9	4.9	15.5

T1, 2020–2050. T2, 2050–2080.

3 Minor points

1) The labels on some of the figures should be larger to be legible.

Response: We thank the Reviewer for the comments. The labels on some of the figures (e.g. Figures 2, 3, and 7) have been enlarged.

2) Did they interpolate the station data to 0.25 resolution? Clarify the methodology (line 5, p9649).

Response: We thank the Reviewer for the comments. Indeed, the observed station data was also interpolated to 0.25° resolution. We have rewritten this sentence as: the model data and observed station data were interpolated to 0.25° resolution using bilinear interpolation.

3) What was the DEM data used for? River network extraction? VIC snowbands?

Response: We thank the Reviewer for the comments. The DEM data was mainly used for river network extraction and flow direction parameter generation used for rout model.

4) p. 9649, line 18: Unless Wang et al. (2012) rewrote the physics of the VIC model, this reference is not appropriate.

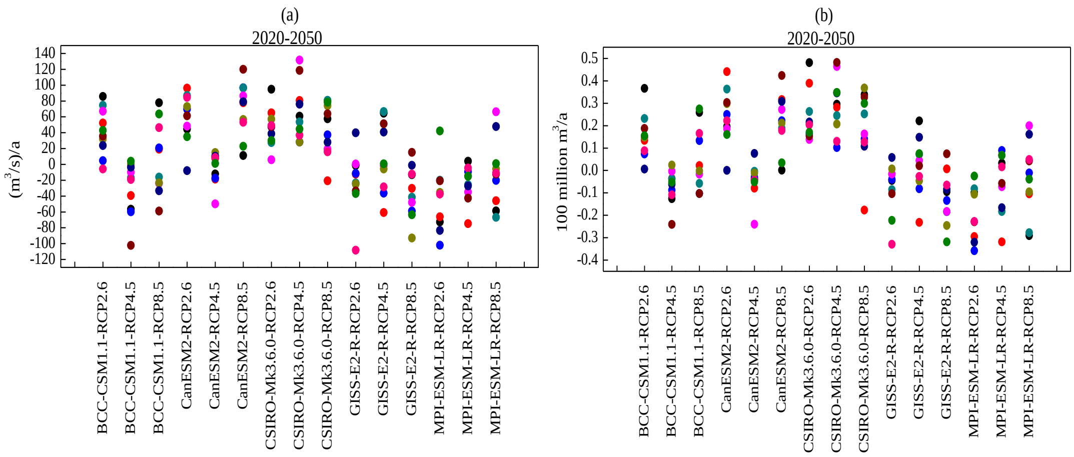
Response: We thank the Reviewer for the comments. We have deleted this reference from Section 2.4 in the revised manuscript.

5) P9649, line 28: Wu et al 2013a citation is not appropriate here unless this reference documented something new about its method and/or its application to hydrometeorological time series.

Response: We thank the Reviewer for the comments. We have deleted this reference from Section 2.4 in the revised manuscript.

6) Figure 4: Showing all trends are misleading, as you could be ending up with a bunch of very minor positive Z statistics for the Mann-Kendall. If you do want to show all trends, somehow incorporating their values, so that the reader knows how larger or small they are, would be useful.

Response: We thank the Reviewer for the comments. To let the readers know how larger or small they (trends) are, we calculated the trends magnitude in the samples of AMX1d and AMX7fv using the nonparametric trend slope estimator developed by Sen. The trends magnitudes (expressed in $(m^3/s)/a$ or 100 million m^3/a) in all the samples for AMX1d and AMX7fv under different scenarios were shown in Figure 4 below.



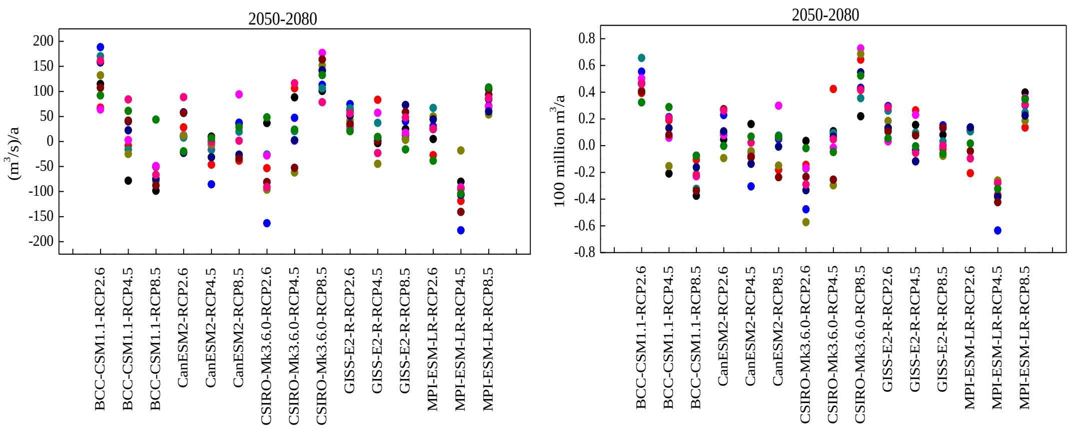


Figure 4. Trends per annum for ten simulated samples of (a) AMX1d and (b) AMX7fv under different emission scenarios.

7) Comment on how good or bad the GCM precipitation was prior to quantile mapping.

Response: We thank the Reviewer for the comments. Before quantile mapping, the GCMs hardly capture the annual distribution characteristics of the observed monthly precipitation during the period 1970–2000, with large relative errors. Overall, simulated and observed precipitation have a poor correlation. Furthermore, strong rainfall features could not be accurately represented for some months, especially in flood season (April-September). For example, most GCMs underestimate the observed in April-June during the study basin. However, quantile mapping method can well solve the problem, as this method has the advantage that it explicitly incorporates changes in the distribution in the future climate.

8) Figure 3 – specify in the legend that these are the downscaled precipitation. Could also make it clear in the text with language: it is only the first half of the first sentence (Section 3.2) that states this.

Response: We thank the Reviewer for the comments. We have rewritten the legend of Figure 3 to further specify that these are the downscaled precipitation. The caption of the Figure 3 is modified as: “Figure 3. ECDFs for precipitation and floods during the period 1970–2000: (a) observed and downscaled AMX1p; (b) observed and downscaled AMX7p; (c) observed and simulated AMX1d; (d) observed and simulated

AMX7fv. Red line represents the observed. Grey lines represent model simulations."

In addition, we have also specified the downscaled precipitation in the text in Section 3.2. Modified sections in the text:

"To assess the performance of the downscaling outputs from GCMs in simulating extreme precipitation, we compared the Empirical Cumulative Distribution Functions (ECDFs) of downscaled maximum 1-day and 7-day precipitation (AMX1p and AMX7p, respectively) against the corresponding observations (Fig. 3a and b)."

9) Why were max 1 day Q and max 7 day volumes used? Are they representative of flooding in this basin?

Response: We thank the Reviewer for the comments. The study basin is located in the tropical and subtropical climate zones, and the basin terrain has a north-to-south inclination. Regional rainstorm often occurs in the downstream catchment area and the upstream high-speed water flows easily form floods. Flood durations of the study area mostly fall within 10 days. Therefore, Max 1 day Q and max 7 day volumes selected are well representative of flooding in the study basin.

10) P 9645, Lines 4-6: Unclear what is being said here. The impacts will exceed economic damage? Isn't economic damage one of the impacts of flooding?

Response: We are very sorry for our unclear sentences. According to the Reviewer's comment, the sentence has been modified as: "the associated impacts will cause probable loss of life and economic damage".

11) The discussion spends a significant amount of time discussing the possibility that the humidity trends are affecting the runoff and could be a source of uncertainty. This can be tested by feeding the model synthetic data. The alternative hypothesis would be that none of that matters for extreme floods driven by extreme precipitation.

Response: We thank the Reviewer for the comments. Our discussion spends a significant amount of time discussing as much as possible sources of uncertainty (including the humidity trends). In the VIC model, daily estimations of evapotranspiration (ET) are achieved by using information on relative humidity, wind speed and long- and short-wave incoming radiation. However, due to the shortage of the observed data (e.g. relative humidity, wind speed and incoming shortwave radiation), daily data of maximum and minimum temperature and precipitation are usually used to estimate the ET. Moreover, this is a common practice in many studies of the VIC model in selected basins worldwide. However, Pierce (et al. 2013) pointed that this approach can result in opposite humidity trends, which are very likely to affect the runoff and could be a source of uncertainty in arid regions (such as much of the western US).

In general, relative humidity (RH) is computed as: RH= $svp(T_{dew})/svp(T_{avg})$, where $svp(T)$ is the saturation vapor pressure at temperature T (Pierce et al. 2013). According to the study of Brown and DeGaetano (2009), the assumption that T_{min} equals T_{dew} (the dew point temperature) is supported by observations in humid regions. The catchment in our study is located in a humid region of southern China, which has sufficient climate conditions (e.g. precipitation and humidity) responsible for the frequently occurred flood disasters. Thus, this approach may have little impact on the results of the future hydrologic scenarios (just inferences). Due to limited space of this paper and large amounts of data which must be processed, it is difficult for us to go into a deeper analysis in this paper. However, in fact, we are now preparing for another article, which performs such analysis on the simulations with humidity and radiative data taken from the GCMs and discuss the potential changes or implications in humid regions.

References:

Brown, P. J. and DeGaetano, A. T.: A Method to Detect Inhomogeneities in Historical Dewpoint Temperature Series, J. Appl. Meteorol. Clim., 48, 2362–2376, 2009.

Pierce, D. W., Westerling, A. L., and Oyler, J.: Future humidity trends over the western United States in the CMIP5 global climate models and variable infiltration capacity hydrological modeling system, Hydrol. Earth Syst. Sci., 17, 1833–1850, doi:10.5194/hess-17-1833-2013, 2013.