Below is an itemized list of all comments in plain text and our responses in blue.

Anonymous Referee #1

The manuscript simulates End of Century (EOC) extremes and their effects on the water-energy balance in the Cosumnes river basin, using cutting-edge global climate and integrated hydrologic models (ParFlow-CLM). I really like the way the authors used to analyze the hydroclimatic changes by median WY, dry WY and wet WY (e.g., Figures 3-5). The manuscript is overall clearly written, and the results are well discussed.

We thank the reviewers for their positive comments and feedback and for acknowledging the quality and the significance of our work.

My first concern is the insufficient validation of the models' simulations in the historical period. Besides temperature and precipitation outputs, other watershed-integrated fluxes, and storages (e.g., ET, soil moisture, TWS and streamflow) should also be validated as much as possible using the observations, remote sensing data and reanalysis, to ensure the models' simulations reasonable. Only then will we believe the further analysis between future and historical periods is valid. In my opinion, the historical simulation of VR-CESM is not so good because the simulated dry, median, and wet water years are distinct from the PRISM (Figure A2).

The developed hydrologic model was previously compared to measurements: simulated ET was compared to remotely sensed ET derived from METRIC, soil moisture was compared to SMAP, snow water equivalent to SNODAS and a reanalysis by Bair et al., streamflow and groundwater levels variations were compared to ground measurements (4 stations were used to compare streamflow and 3 wells for groundwater levels comparisons). Comparisons with GRACE TWS are not meaningful given the size of this watershed (~7000 km^2) which is far smaller than the footprint of GRACE TWS (200,000 km^2). An appendix containing these comparisons will be added to the revised manuscript.
Because the hydrologic model was only run for a certain period of time and specific years, the comparisons were only performed for these years on the contrary to the climate model which has been compared throughout the entire historical period.

Validation of the hydrologic model

We compared temporal variations of streamflow at 3 stations located in the Sierra (uplands), the intersection between the Sierra and the Central Valley, and the outskirts of Sacramento (see Figure R1). Four wells in the watershed (see Figure R1a) have reasonable, publicly-available records of groundwater levels and were used to check the ability of the model to reproduce water table depth variations.

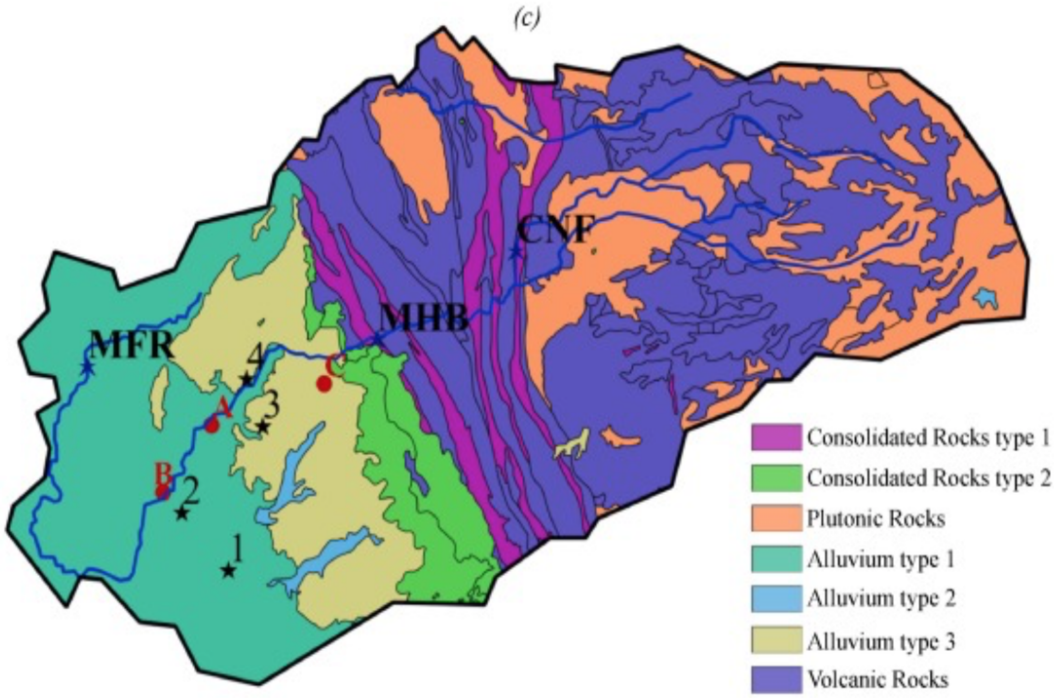

Figure R1a: The Cosumnes watershed geology and the locations of the 3 streamflow gauges (CNF, MHB, and MFR) and 4 groundwater wells (stars).

Figure R1b depicts the comparisons between simulated and measured river stages at the 3 stations indicated in figure R1a. Absolute errors (L1) in m and relative errors (L2) are shown in Table R1a. Differences between simulated and measured streamflow vary between 0.4 and 0.8 m (Table R1a) indicating that the model is able to reproduce the river dynamics.

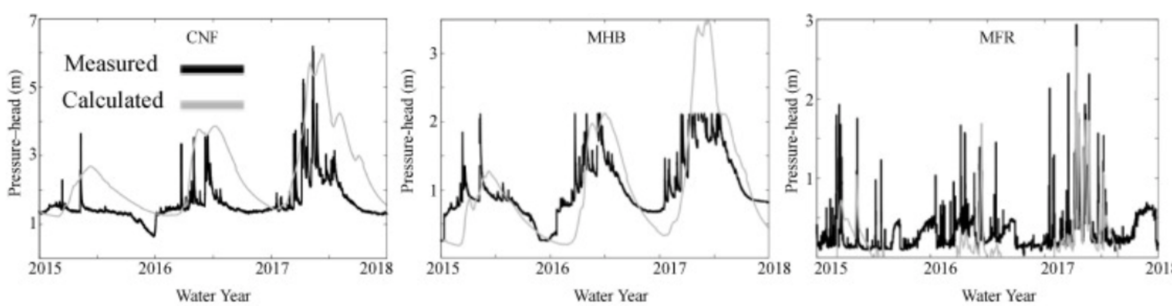

Figure R1b: Comparisons between measured and calculated river stages (i.e., pressure-heads simulated by ParFlow-CLM). Measurements locations are indicated in Figure R1a.

Measurements		
	L_1**(m)**	L_2
River stages (CNF)	0.8	0.5
River stages (MHB)	0.4	0.36
River stages (MFR)	0.57	1.09
Groundwater Levels (Well 1)	3.73	0.0553
Groundwater Levels (Well 2)	1.63	0.025
Groundwater Levels (Well 3)	0.476	0.0077
Groundwater Levels (Well 4)	1.08	0.016

Table R1a: Differences between measured and calculated surface and groundwater levels. L1 is the absolute error and R2 the relative error.

Comparisons between simulated and calculated groundwater levels (here referred to as the pressure-heads at the bottom of the domain) shown in figure R1c indicate that the model has reasonable agreements with measurements. As shown in table R1a, the error varies between 0.47 to 3.73 m depending on the station. Mismatches between simulated and observed groundwater levels at wells 1 and 2 are likely due to an inaccurate estimation of pumping in these areas. The temporal variations of the groundwater levels show an impact of withdrawals but because these withdrawals are hard to estimate the model isn't correctly reproducing these trends.

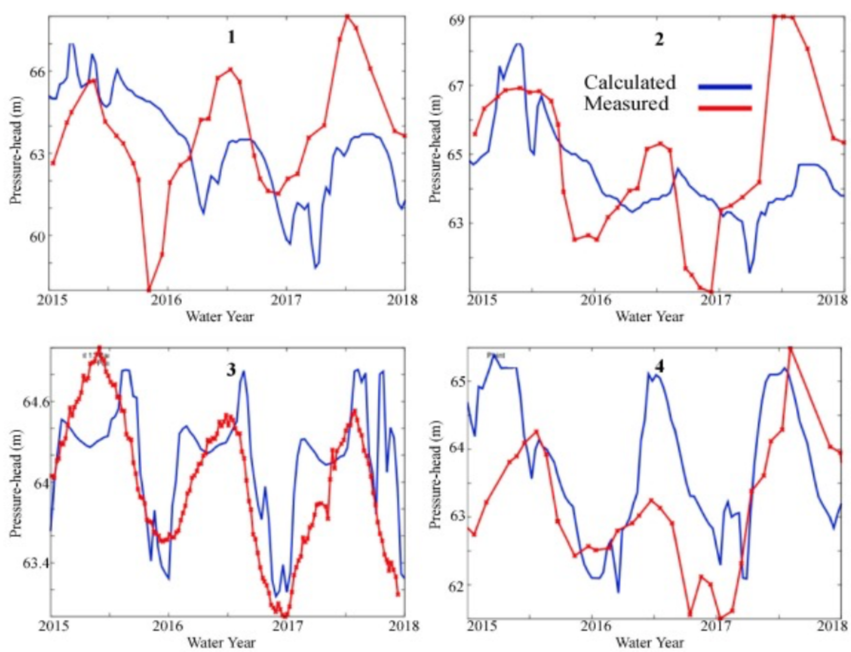

Figure R1c: Comparisons between measured and calculated pressure-heads at the bottom of the domain. Measurements locations are indicated in Fig. R1.

ParFlow-CLM also solves the key land surface processes governing the transfer of water and energy at the land-atmosphere-soil interface: evapotranspiration, snow dynamics, and soil *moisture. In Maina et al., (2020a), rigorous comparisons between the ParFlow-CLM simulated land surface processes and remotely sensed estimates of these variables was conducted. Table R1b shows the correlation coefficient between ParFlow-CLM results and the various datasets compared.*

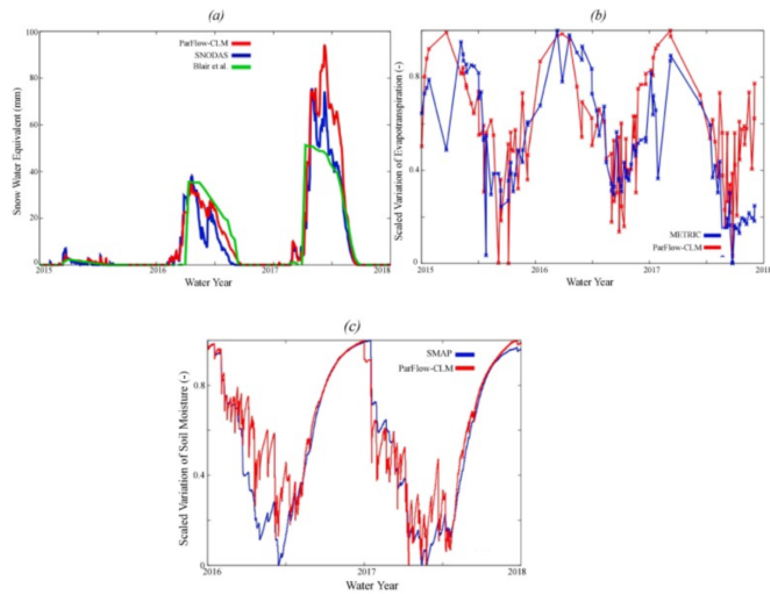

Figure R1d: (a) Comparisons between domain-averaged total snow water equivalent obtained with ParFlow-CLM, SNODAS and Bair et al., reconstruction, (b) Comparisons between actual evapotranspiration obtained with ParFlow-CLM and METRIC (c) Relative variation of soil moisture obtained with ParFlow-CLM and SMAP. Note that the x-axis of (c) is shorter because of the availability of SMAP data

Satellites-based products

	L_1	L_2	Pearson correlation coefficient
SNODAS (mm)	3.09	3.77	0.97
Bair et al., 2016 (mm)	3.80	2.69	0.84
SMAP (-)	0.217	3.07	0.94
METRIC (mm/s)	0.0367	1.40	0.6

Table R1b: differences between measured and remotely sensed evapotranspiration (METRIC), soil moisture (SMAP), and snow water equivalent (SNODAS and Bair et al., 2016)

VR-CESM is simulated under AMIP-protocols (bounded by monthly observed sea-surface temperatures and sea-ice extents), and therefore we do not expect VR-CESM to exactly recreate past historical water years. However, we do expect that our 30-year simulation can reasonably recreate the range of water year types over California and the Cosumnes, which is why we utilize the broader range of PRISM water years that are available. While the water years are different, the magnitudes of the precipitation are similar.

The authors may argue the historical simulations are acceptable, because a global climate and integrated hydrologic models are used (more complex and larger simulation domain). However, one can use a finer-resolution hydrological model (e.g., VIC, SWAT, and many others) driven by statistically or dynamically downscaled regional climate model outputs to obtain more reasonable (maybe more accurate from the perspective of validation) simulations in this river basin (7000 km2), and to do further analysis like the authors did in this study. Please explain why the global climate and integrated hydrologic models are more suitable for this case study?

We set up our modeling framework by taking into account the:

- Californian atmospheric dynamics.
- Impacts of groundwater dynamics and lateral flow on the hydrology and the land surface processes of the region.
- Dependance of the groundwater dynamics in the valley to the snow dynamics in the Sierra Nevada

These considerations are critical for a better understanding of the impacts of a changing climate on Californian hydrology.
ParFlow-CLM is an integrated hydrologic model that solves the transfer of water and energy from the bedrock to the canopy. Parflow uses the Richards equation a physics-based equation that solves the subsurface flow in three dimensions and therefore accounts for deeper and lateral flow. Previous studies have demonstrated that the lateral flow is very important to the surface and land energy dynamics (Maxwell and Condon, 2016). On the contrary to VIC, ParFlow accounts for the lateral flow. It also employs a series of physics-based equations contrary to SWAT. In addition to the subsurface flow, ParFlow also solves the overland (i.e., surface) flow by using the kinematic wave equation contrary to VIC.
When simulating the evolution of climate in California, the interaction between dynamical and thermodynamical responses has important, and sometimes, offsetting effects on features such as atmospheric rivers. Payne et al. (2020) show that the thermodynamic response to climate change enhances atmospheric river characteristics (e.g., Clausius-Clapeyron relationship), whereas the dynamical response diminishes atmospheric river characteristics (e.g., changes in the jet stream and storm track landfall location). Therefore, it is important to employ a modeling framework that accounts for the dynamical and thermodynamical effects of climate change such as VR-CESM.
We also did not perform statistical downscaling as this is one of the "selling points" of leveraging variable-resolution Earth system model capabilities, namely that it enables dynamical downscaling internally within an Earth system model which has the benefits of limiting multiple model bias propagation, allows for more consistent teleconnection responses, enables upscale/downscale effects to influence the broader climate, etc. As a result, we think this study adds a "unique" data point to the literature regarding changes in end-century hydrology in California given that it is a slightly different methodology compared with traditional regional climate model based dynamical downscaling efforts and/or bias-corrected statistically downscaled global climate model efforts.
To capture both the particularities of Californian climate and its interactions with the hydrology from bedrock to the canopy, the approach we used is more adequate than the aforementioned approaches.

Below is a table with the most used hydrologic models and their advantages and limitations when simulating the hydrology of California. Only Hydrogeosphere and ATS have similar advantages as ParFlow-CLM and are suitable to model the Californian hydrology. Because the equations and the coupling approaches used by these models are similar, we expect their results to be the same. Moreover, these models are also computationally expensive hence they also have to limitation of resolution.

Hydrologic Model	**Land Surface**	**Surface**	**Subsurface**	**Limitations when simulating Californian hydrology**
MODFLOW (Harbaugh, 2005)/FELFOW (Trefry and Muffels, 2007)	No	No	Yes (diffusivity equation)	These models do not integrate land surface processes (such as snow dynamics) and their interactions with the subsurface critical to the Californian hydrology.
SWAT (Soil and Water Assessment Tool) (Neitsch et al., 2000)	Yes	Yes	Yes	The model is based on HRU (hydrologic response units). The model isn't physics-based, therefore, it doesn't account for the two-way interaction between the land surface and the subsurface processes.
SAC-MA (Sacramento Soil Moisture Accounting Model)	No	Yes (Rainfall-Runoff)	Yes (Water Budget)	The model doesn't simulate snow dynamics and evapotranspiration. A water budget equation is used to simulate the groundwater dynamics which doesn't account for the lateral flow and unsaturated zone flow.
Noah-MP (Niu et al., 2011)	Yes (water and energy balance)	Yes (a routing scheme can be used to derive surface flow)	Yes (percolation)	Although this model physically solves the land surface processes including evapotranspiration and snow dynamics, it doesn't account for the two-way interaction between the land surface processes and the subsurface. Lateral and unsaturated zone flows are not represented.
VIC (Variable Infiltration Capacity Model Macroscale	Yes	Yes (Rainfall-Runoff)	Yes (percolation and water budget)	Although this model physically solves the land surface processes including evapotranspiration and snow dynamics, it doesn't account

Hydrologic Model) (Liang et al., 1994)				for the two-way interaction between the land surface processes and the subsurface. Lateral and unsaturated zone flows are not represented.
Hydrogeosphere (Aquanty, 2015)	Yes (water and energy balance)	Yes (2D diffusive wave equation)	Yes (3D Richards equation)	This model has similar advantages as ParFlow-CLM and could be used to model the hydrology of California.
CATHY (Catchment Hydrology) (Bixio et al., 2002)	Yes (there is a version coupled to Noah-MP)	Yes (1D Saint Venant Equation)	Yes (Mass balance equation)	The mass balance equation is not as robust as the Richards equation for describing the variably saturated flow in the subsurface and recharge processes. In addition, the original model doesn't solve land surface processes.
MIKE-SHE (Abbott et al., 1986)	No	Yes (diffusivity equation)	Yes (Darcy equation and a 1D Richards equation)	The main limitation of this model is the lack of land surface processes and the Darcy equation used to describe subsurface flow doesn't account for the unsaturated flow.
ATS (Advanced Terrestrial Simulator) (Coon et al., 2016)	Yes (water and energy balance)	Yes (2D diffusivity equation)	Yes (3D Richards equation)	This model has similar advantages as ParFlow-CLM and could be used to model the hydrology of California.
ParFlow-CLM (Kollet and Maxwell, 2006)	Yes (water and energy balance)	Yes (2D diffusivity equation)	Yes (3D Richards equation)	

Table R1c: Advantages and limitations of the most used hydrological models

Additional references

Maina, Fadji Z., Siirila-Woodburn, E. R., Newcomer, M., Xu, Z., & Steefel, C. (2020a). Determining the impact of a severe dry to wet transition on watershed hydrodynamics in California, USA with an integrated hydrologic model. Journal of Hydrology, 580, 124358. https://doi.org/10.1016/j.jhydrol.2019.124358

Maina, F. Z., Siirila-Woodburn, E. R., & Vahmani, P. (2020b). Sensitivity of meteorological-forcing resolution on hydrologic variables. Hydrology and Earth System Sciences, 24(7), 3451–3474. https://doi.org/10.5194/hess-24-3451-2020
Maina, Fadji Zaouna, & Siirila-Woodburn, E. R. (2020c). Watersheds dynamics following wildfires: Nonlinear feedbacks and implications on hydrologic responses. Hydrological Processes, 34(1), 33–50. https://doi.org/10.1002/hyp.13568
Harbaugh AW (2005) MODFLOW-2005, The U.S. Geological Survey modular ground-water model: the ground-water flow process. US Geol Surv Tech Methods 6-A16. http://pubs.usgs.gov/tm/2005/tm6A16/.
Trefry, M.G.; Muffels, C. (2007). "FEFLOW: a finite-element ground water flow and transport modeling tool". Ground Water. 45 (5): 525–528. doi:10.1111/j.1745-6584.2007.00358.x.
Neitsch, S. L., Arnold, J. G., Kiniry, J. R., & Williams, J. R. (2001). Soil and Water Assessment tool (SWAT) user's manual version 2000. Grassland Soil and Water Research Laboratory. Temple, TX: ARS.
Niu, G.-Y., et al. (2011), The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. J. Geophys. Res., 116, D12109, doi: 10.1029/2010JD015139.
Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges (1994), A simple hydrologically based model of land surface water and energy fluxes for general circulation models, J. Geophys. Res., 99(D7), 14415–14428, doi:10.1029/94JD00483.
Aquanty, I. (2015), HydroGeoSphere User Manual, 435 pp., Waterloo, Ont.
Bixio, A. C., G. Gambolati, C. Paniconi, M. Putti, V. M. Shestopalov, V. N. Bublias, A. S. Bohuslavsky, N. B. Kasteltseva, and Y. F. Rudenko (2002), Modeling groundwater-surface water interactions including effects of morphogenetic depressions in the Chernobyl exclusion zone, Environ. Geol., 42(2-3) 162-177
Abbott, M. B., J. C. Bathurst, J. A. Cunge, P. E. Oconnell, and J. Rasmussen (1986), An introduction to the european hydrological system: Sys- teme hydrologique Europeen, She .2. Structure of a physically-based, distributed modeling system, J. Hydrol., 87(1–2), 61–77.
Coon, E. T., J. D. Moulton, and S. L. Painter (2016), Managing complexity in simulations of land surface and near-surface processes, Environ. Modell Software, 78, 134-149
Kollet, S. J., & Maxwell, R. M. (2006). Integrated surface–groundwater flow modeling: A free surface overland flow boundary condition in a parallel groundwater flow model. Advances in Water Resources, 29(7), 945–958. https://doi.org/10.1016/j.advwatres.2005.08.006
Maxwell, R. M., & Condon, L. E. (2016). Connections between groundwater flow and transpiration partitioning. Science, 353(6297), 377–380. https://doi.org/10.1126/science.aaf7891