This paper was co-reviewed by Katherine Reece and Wouter Knoben.

Summary of the Paper:

- This paper aims to quantify equifinality's impact on post-fire streamflow predictions. The authors propose a novel method to reduce uncertainty by filtering parameter sets based on their bias stationarity. The authors use the 2020 Creek Fire wildfire as a case study for post-disturbance hydrological modelling, where changes in bias before and after a disturbance can compound uncertainty.
- Using the Distributed Hydrology Soil Vegetation Model (DHSVM), the authors simulate streamflow in the Upper San Joaquin River Basin. A 14-parameter space encompassing vegetation, subsurface, snow, and meteorological factors is calibrated using multi-objective Bayesian optimization, which produced a 30-member behavioural ensemble. The authors introduce the "bias shift" metric to compare the accuracy of pre- and post-fire simulations, and they develop a Bayesian statistical metamodel that utilizes this bias shift to predict changes in streamflow. They compare these process-based results to those from an empirical regression model based on a simplified water balance.
- The raw DHSVM ensemble predicted post-fire streamflow increases ranging from +2% to +17%, revealing substantial uncertainty. However, parameter sets with stationary bias across the fire event produced more consistent results. Using these "stationary" parameter sets, the authors reduced uncertainty in key variables such as leaf area index and transmissivity. The Bayesian metamodel, which filters for zero bias shift, estimated an average post-fire streamflow increase of +11% ±1%, representing an 80–82% reduction in uncertainty compared to either the raw DHSVM ensemble or the empirical model alone.

Strengths:

- The introduction of a bias shift metric to screen for equifinal parameter sets is novel and can be applied to other disturbance studies. By linking bias shift with hydrologic response variability, the authors present a method for identifying parameter sets that are robust to bias shift.
- The paper combines process-based simulations (DHSVM) with a Bayesian statistical metamodel to reduce uncertainty and enhance the insights that can be gained from the model. This hybrid approach can also be used as a template for future disturbance impact assessments.

Detailed Comments:

- The authors should define counterfactual. On page 2, paragraphs 2 and 3, the authors use this term, but the meaning is not clear. It may need to be rephrased to increase the manuscript's accessibility.
- On page 3, in the first paragraph, the word "since" should be replaced by "because" due to the authors' intent.
- On page 3, paragraph 1, the logical connection between the lack of landscape-scale observation of many environmental properties and model properties is that we cannot infer the parameters from data and therefore need to calibrate. We suggest making this explicit to increase the accessibility of the text.
- On page 3, line 71, "since" is used instead of because.
- On page 3, line 72, the authors hypothesize that equifinal parameter sets may produce divergent predictions. Could work be cited here to state this as a fact rather than formulating this as a hypothesis?

- Figure 1 on page 4 could be improved by including more details. This would reduce the number of assumptions readers will have to make in order to understand the image.

  -"Initial water balance": for the sake of argument, we'll assume that all models perfectly fit current streamflow, even if given different inputs (P), the simulated ET_tree and ET_other components are such that Qsim = Qobs = 1. "Disturbance response": some disturbance reduces ET_tree (line 79-80).

  The image could better explain what the up and down arrows in this column show. Readers can speculate that the downward ET_tree arrows mean that (to match reality) the models simulate no ET_tree anymore. In model 1, this means Q_sim goes up by 2, but why would ET_other in models 2 and 3 go up by either the negative change in ET_tree or half that? If these are meant as examples of what could happen with a given model (rather than what will happen), it would be good to state this in the text explicitly. E.g. "[..] increased soil water availability. The three examples show cases where ET_other does not change (model 1), where ET_other fully compensates the reduction in ET_tree (model 2), and where ET_other only aprtly compensates the reduction in ET_tree".

  -"Streamflow bias": here seems to have an underlying assumption that Qobs increases by 1 after the disturbance. This is key to understanding why the models are assumed to underpredict/overpredict/no change and should be explicitly stated somewhere.

Given that this example is intended to give the reader an easy intro into the concepts used in the paper, it's worthwhile to make sure all assumptions are clearly stated. Without this figure, it may raise more questions than it answers. A possible solution is to move the paragraph with lines 92 to 100 before the figure.

- Figure 2 should have labels in a better font. The current font choice looks out of place. The caption also needs to define the acronyms RCMAP and UTM.
- On page 7, lines 139 to 141, the description of the methods would be strengthened if the resampling technique used to aggregate the 30 m resolution data to the 90 m resolution model were explicitly stated. Was it averaging, weighted aggregation or majority rule? An explanation of how this mapping works in cases where there are differences in burned area at the original 30m resolution would help reproducibility.
- On page 7, lines 143 to 144, the authors should provide a quick justification for why October 1 was chosen to update the vegetation maps. A quick sentence saying it corresponds to the water year, or the availability of vegetation maps at the end of the fire season, or any other justification, should be stated. An October 1st update could introduce artificial bias, so it would be good to justify why this approach was followed.
- On page 7, lines 147 to 148, the authors should explain why estimation is preferable to using LAI observations from something like. Are the empirical estimates close to satellite observations?
- On page 8, it is important to speak about calibrating correction factors for meteorological inputs, which has the potential to compensate for any deficiencies in the model itself. For a study trying to investigate how calibrated models predict process changes after a disturbance, calibrating bias-correction parameters for the forcing could introduce a lot of complexity (in other words, substantially enhance equifinality) that will complicate later analysis. A broad statement like "gridded meteorological data can have considerable uncertainty" is insufficient to justify this. Is there any concrete evidence that the specific meteorological data chosen are biased in this particular watershed? (Even) If so, the correct approach would be to bias-correct the forcing before calibration starts, so that every model in the ensemble uses the same inputs. This seems the only way to get a clean comparison between models later.
- On page 8, lines 194 to 195, some extra explanation about why these three objectives are based on reconstructed daily streamflow is needed and how it works

would be good to add. A reference to a more in-depth explanation would be enough. How does the reconstruction approach deal with the disturbance if this is reconstructed? Is that accounted for somehow?

- On page 9, the caption for Table 2 states that NSE is identical to R^2 for statistical models. It would be better if it stated that NSE is analogous to R^2 and not identical, and if a citation from the original Nash and Sutcliffe paper from 1970 were provided.
- On page 10, line 217, 600 samples for a 14-dimensional space does not seem a lot, and a large number of these might already be Pareto-optimal just because of the low number of samples. A table that shows how many of each sample fulfills each individual criterion would be good.
- On page 10, line 234, given that streamflow is reconstructed, the authors should use the term "reconstructed streamflow" instead of calling this "observed" to help the reader understand that we're comparing results from one model to another.
- On page 11, line 242, the authors should state the importance of the bias shift metric. The co-reviewers came to differing conclusions about its importance. Is it a metric to inform people about models, or is the bias shift something that needs to be corrected, and how is the correction applied?
- The reasoning on page 11, lines 249 to 250, is a bit difficult to follow. Are there any tests that can be done to see if this assumption is valid, or can the authors speculate how the results would change if it didn't hold?
- Figure 3 shows some variability in the simulations, particularly around the middle flow magnitudes. Can it be clarified why these simulations are called "satisfactory"? Presumably, there's an implicit middle step that says "these NSE scores are [something], therefore these parameter sets all satisfactorily reproduce observed streamflow". I encourage the authors to add their reasoning for thinking that these simulations are good enough for this study. Are these simulations of typical quality for this particular watershed (i.e., comparable to other modelling efforts)? Are these ranges of scores particularly high for this specific watershed?
- Figure 4 needs more explanation. For example, for both x-axes, are we looking at different parts in space? In other words, does this figure show how different parts of the watershed respond? The reviewers are asking because the caption says that the values on both x-axes are derived from data and not calibrated, but if this is so, the reviewers are unsure how this figure shows parameter uncertainty. By counting we can summarize that each symbol stands for a behavioural parameter set, but this should be stated somewhere.
- The data and code availability statement on page 22 is incomplete. See the guidelines at https://www.hydrology-and-earth-system-

sciences.net/submission.html. The co-reviewers would like to take a look at the streamflow series for this basin but cannot easily find the location of this data either in this section or through the California DWR reference mentioned in the text.

- In the references, line 586, this link does not go to the dataset but to the landing page. A DOI, an accurate link, and the access date are needed.