

Review of NHESS-2020-429

Assessing short-term climate change impacts on water supply at the Wupper catchment area, Germany.

The paper studies impact of climatic change on the runoff and potential to supply drinking water using short-term climate scenarios. As such it is interesting and relevant for the journal, and it is interesting to see the application of a different approach to providing climate scenarios. I think the results are nicely presented with good figures, and most of my comments and questions to the manuscript is related to the explanations on how and why things are done as they are.

I know this is an outcome from the BINGO project, and in that respect, I do have a couple of issues that came up while reading the paper. The major issue is related to the very frequent citations of BINGO reports. In some cases, this is of course perfectly ok but in other cases some of these results should be in the paper to improve the readability and prevent the reader for having to download a number of reports to find important information. A minor issue is the textual references, e.g. "I the frame of BINGO ..." or similar language. From the intro we know that this is done in the BINGO project so that the text flow and language simplicity can be improved.

I have some questions related to the hydrological modelling:

- Why using both the NASIM and SWAT model in the project? There is a partly explanation of the difference of the model, but what is the benefit of using both in the Wupper catchment? Are they focused on different topics, and if so how do we interpret the results? From the text around line 145 it is not easy to understand.
- What strategy was used for calibration of the models?
- Did you consider uncertainty in the calibration?
- The calibration and validation if totally referred to a project report stored online. You could at least provide a summary of the calibration/validation, goodness of fit measures, difference in calibration results and what this means for the comparison of the simulations of the models. It would be particularly interesting to know how the models handled low flow periods.
- Can a hydrological model be physically and lumped? What do you mean with physically? I do not know NASIM, but isn't SWAT semi distributed and in many process descriptions more conceptual than physical.

You do discuss and use the climate realizations both in the methods and in the results. The difference between these should be explained in the paper (or in supplementary material). This is too important for the understanding of the paper to be left in an online report. I think I understand that realisations 1, 7, 9 are max,min mean (line 159), but how large are the differences?

What is the basis for the claim that the decadal projections provide a more realistic assumption than the ensemble of RCP based scenarios?

I think a clarification of the text around line 75 is in order. So, the reservoir filling depends on spring precipitation and since spring precip now has shifted to summer this is no longer

possible. And in general (meaning in the past?) summer precipitation has not been important (since it previously was in the spring)?

At the start of the Data and methods section there is some basic drought information which should be moved to the introduction.

The text book information in the section on statistics could be more stringent. What method was used, when and the p-level used. All methods are standard so the meaning of H_0 and similar things are not needed.

The start of 4.1 is method material, could be moved. The use of Buchenhofen is also described before in the text.

I think I would have done the test of the overlapping climate predictions before I used them for analysis, but that is not very important.

Table 4: Is the runoff from NASIM or SWAT

Line 350-355: Why do you think the model perform like described here? Does this in any way related to the calibration/validation of the model, e.g. do you see the same pattern when you identify the model parameters? This is a reason to do a more detailed job in describing the model calibration, even if this exist in some online report.

I do also miss a short description of the reservoir model TALSIM-NG and particularly operational rules. A proper reference would also be in place. Any calibration or test of this model? What operational rules was used here, do they change in the future climate due to the changes in runoff observed?

Why was the RCP based models used here? What climate models was used to generate the RCP simulations, is it an ensemble of several?

In my opinion the conclusion is a bit long and repeats some findings presented before, to some extent it now works like a discussion.

What causes the large spread in simulations for the reservoir in figure 9? How does the volume of runoff in the scenarios compare to the observed?

Minor issues:

- Can you provide a proper reference for NASIM?
- On line 57 on the second page, could you say something more on the magnitude of changes in dry periods? Even if it is described before you could say how the rainy season has changed here.
- Line 69: What is meant with the “weather normal distribution” is it the weather normal period and if so 1961 – 1990?
- Line 79: “remained rather natural” could probably be simplified to “remained natural”

- Line 227: “was able to prove” is a bit awkward. The text “The non-parametric” could be removed. Should “less and equal” be “less”? You could just write “Both period show a significant positive trend (Mann-Kendall test, $p < 0.05$)”.
- Line 232: Isn't it quite common to report very small p-values as $p < 0.001$? Above you use $p < 0.05$ and here the detailed decimal number, be consistent.
- Line 300 just repeats that you have used NSIM an SWAT.
- Line 330,
- Some places “less and equal” are used for p-values, shouldn't that be “less” given your explanation in the method section?