

## Reply on RC1

1. *Could you provide technical information about the image quality required for the proposed DARE methodology? Is there a minimum resolution (DPI) required? I only found some generalities on this matter. For instance, in lines 238-239 you can read: “Typical workflows comprise high-resolution scanning”.*

The State Archives where I got the images from provide scans with a resolution of minimum 300 ppi (measured against the original) which is a standard in cultural heritage digitisation, for instance as defined by the German Research Council (DFG) and many others. I will add this information to the manuscript. However, this is not to say that this resolution is a requirement for the workflow to be effective. I would put more importance on flattening the surface and a uniform and adequate lighting.

2. *In connection to previous comment, I wonder if you worked (or plan to work) with photographs rather than scans. I’m familiar with rescuing meteorological data where it is advised to photograph large amounts of documents because this imaging procedure is many times faster than scanning (see Section 2.4.2. of Wilkinson, 2019). You developed an efficient way of obtaining machine-readable data from graphs and it would also make sense to accompany it with an efficient imaging strategy. Although this imaging issue does not matter much if all the Danube water level charts were already scanned with high resolution.*

You are right, I was too sloppy with the word "scanning." Typically, to my knowledge, the archive is indeed using a system with a one-shot DSLR camera mounted on what we call "Reprostativ" with external lighting for this kind of material. In my own lab, we additionally use scanners with feeders for smaller sizes, also for historical material, which is a huge step forward in efficiency. However, the State Archive would not allow to use such a system due to conservation concerns (which I do not share). We also experimented with a scan tent (with mounted smartphone camera) and with simple smartphone camera shots. Due to the size of the material, both did not deliver sufficient quality for the processing, with the achieved resolution being a lesser problem than inconsistent lighting and warping.

Overall, I suggest to rephrase l. 237 to: "Typical workflows comprise creating high quality images from the originals (normally digital photographs 300ppi resolution as standard, flattened surface, orthogonal alignment, and adequate uniform lighting)."

I do not see a comprehensive imaging strategy within the scope of this paper. There are too many parameters to be discussed systematically, including questions of material, size, cost-benefit, and restriction laid out by the owner of the materials. But I may add a reference to the strategies typically employed in cultural heritage digitisation projects.

3. *In the second paragraph of Section 3.1 you can also mention the methods developed to extract subdaily data from strip charts of meteorological instruments such as thermographs and barographs (e.g., Sušin and Peer, 2018).*

Thank you very much for this. I will do some more research here and add the references.

4. *In section 3.2.1 entitled “Americas” you can cite recent works that rescued long Paraná River hydrometric records, which start in 1875. Indeed Antico et al. (2018) manually digitized daily water levels from a hand-drawn chart similar to the one shown in the upper left panel of Fig. 5 of the revised*

*manuscript. More recently Antico et al. (2020) found the tabulated version of these data and compared tabulated values with those digitized from the chart.*

Thanks a lot also for these that I missed. I will add them. Tabulated data and its alignment with charts is also on my to-do-list for some further work, and it is good to have a near reference now. Tabulated historical data is currently widely discussed thanks to the progress made in automated reading of those by multimodal LLMs.

*5. Did you consider using documentary sources (e.g., newspapers) or metadata provided by the charts to correct time misalignments of positive and negative peaks (floods and drought)? This could be a useful correction, as knowing the exact date of these peaks is important for many studies.*

*6. Similarly, documentary sources may inform the exact river levels attained during these peaks. That is, these sources may serve to correct these levels.*

This is a great suggestion! I would like to think about it but as a follow-up project. Thanks to modern OCR and LLMs and progresses in Natural Language Processing, obtaining such information from newspapers on large-scale appears feasible. However, at least in Germany, the state of digitisation of local newspapers from the 19th century is not what it could and should be (incomplete, some kind low quality, de-centralised and difficult to access). I will investigate into Passau (gauge 30) which is my home town with easy access to the municipal archive, but rather leave it for a follow-up.

Thank you also for your minor comments which I will implement!

## **Reply on RC2**

Thank you very much for your valuable, competent comments and kind words. I appreciate the time you invested in helping to improve this research.

With regards to the typos: I cannot spot one in line 35. Line 182, however, sounds indeed awkward -- I will fix that.

To your general questions and comments: Another reviewer also suggests to correlate floods and arid periods with other historical sources. This a great suggestion which I would like to follow up. I am currently in dialogue with some archives that might have additional sources, and I am also looking into mining historical newspapers. As a response to RC1, I wrote: Thanks to modern OCR and LLMs and progresses in Natural Language Processing, obtaining such information from newspapers on large-scale appears feasible. However, at least in Germany, the state of digitisation of local newspapers from the 19th century is not what it could and should be (incomplete, some kind low quality, de-centralised and difficult to access). I will investigate into Passau (gauge 30) which is my home town with easy access to the municipal archive." It is a new research project on its own, and I am looking forward to it.

Thank you for pointing me to the Kendon et al. report. I think this is a good case for making all this historical data openly available in a connected form.

About "the magnitude of the floods relative to more recent events": That would be interesting, too. The data I present here are "self-contained" in a way that they have all parameters required to make the data in this series comparable in themselves. Aligning them with modern data requires some more research

to deal with gauge-zero settings or possibilities that the gauges could have been moved or replaced. Here, too, I would say that it is a great suggestions that should be followed-up but that it is beyond the scope of this paper.