„Reconstructing Nineteenth-Century River Water Levels with Transformer-Based Computer Vision“ — Response to the editor‘s letter from 11 February 2026

Dear editors,

Thank you for your clarification and for your work on this!

I had already implemented the changes requested by the two anonymous reviewers in my revised manuscript. They should be visible in the „author‘s tracked changes“ document. To ease verification, I have revised the responses to the reviewers (see below „Reply on RC1“ and „Reply on RC2“, highlighting the changes made in bold font).

——

In addition to this:

1. We noticed that your dataset https://doi.org/10.5281/ZENODO.17296750 is distributed under license CC-BY-SA-4.0. Please note that ESSD requires the data to be distributed under a non-restrictive license such as CC BY 4.0 or CC0. Please adjust the license accordingly.

I have changed the license to CC BY 4.0.

2. Regarding figure 2: OpenStreetMap images must include visible attribution on images/snippers/ or the full credit in the figure caption (e.g. “© OpenStreetMap contributors, https://www.openstreetmap.org/copyright”). Please add it.

Done so.

——

Apologies that I oversaw to respond to your letter from 4 December 2025. Thank you very much for these thoughtful comments!

- Assuming the source data are distributed across several different volumes, would be it possible to explicitly reference them as fully as possible (for example by name of publication, year, publisher etc.)? This information could potentially be included in an additional table somewhere.

Thank you very much for this important hint. **I have added the following in section 2.2:**

„The primary sources are archival river–gauge charts compiled by the Bayerisches Landesamt für Wasserwirtschaft (BLW) and today preserved in the Bayerisches Hauptstaatsarchiv (BayHStA) (Munich) **under the call number (archival signature) BayHStA II 2.11.2.1.3 Landesamt für Wasserwirtschaft 1.2.1 Pegel an der Donau**.“

I have also **added this information to Table 1**:

„Nineteenth-century Bavarian Danube gauges in the archive; case-study gauges highlighted. Six Danube gauges from a later period are not listed here. **Source: Bayerisches Hauptstaatsarchiv II 2.11.2.1.3 Landesamt für Wasserwirtschaft 1.2.1 Pegel an der Donau. The last two digits from**

gauge_id (21-31) are to be added to the call number for accessing the primary sources in the reading room of the archive.“

- *Would it be possible to share all of the scans representing the source data in the Zenodo repository. This could be useful not only in the context of enhancing the reproducibility of this study, but also for testing and comparing possible future / alternative approaches for data extraction. If not possible, then it should be at last mentioned how the original source data could be obtained.*

Yes, this is absolutely desirable and an asset. **The Zenodo repository already contains them** (file: images_months.zip). These are the monthly slices. For the full pictures: I have discussed the issue with the Directorate of the Bavarian State Archives. Their general policy is to put only fully digitized collections online. The original sources are accessible in the archive under the call number given above.

- *L22. Consider briefly describing what LineFormer is here*

Done. This passage now reads:

„A pragmatic, semi-automated workflow (HWLR) that combines light, grid-aware pre-processing, optional dewarping, transformer-based line extraction **(using LineFormer, a deep learning architecture designed to treat line detection as a regression task by predicting polylines directly from image features)**, ...“

- *Section 3.2. This information may be better suited to the introduction section, since it is rather general. The manuscript would then become more focused on your case study as it progresses.*

I moved the whole section. Please note that I did so **after** creating the diff-text, because otherwise the changes within this section would not be visible any more.

——

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”.

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.

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.

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 (in response to questions 1 and 2), I rephrased the paragraph with new references on digitization to (see l. 243):

„Typical workflows comprise creating high quality images from the originals (normally digital photographs with 300ppi resolution as standard, flattened surface, orthogonal alignment, and adequate uniform lighting) (Wilkinson et al., 2019; Initiative, 2023; Rehbein, 2026); image pre-processing (deskewing, dewarping); segmentation; axis calibration to the chart's printed grid; and interactive tracing or point-picking along the plotted curve, followed by export to a regular temporal grid and documentation of datum/unit conversions (Mitchell et al., 2019).“

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 chats of meteorological instruments such as thermographs and barographs (e.g., Sušin and Peer, 2018).

Thank you very much for this. **I have done some more research here and added the references as follows (l. 119):**

„The same applies to subdaily data from charts from meteorological instruments such as pluviographs (rainfall), thermographs, and barographs (Sušin and Peer, 2018; Jaklič et al., 2016; Capozzi et al., 2020).“

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. 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. **I have added these reference as follows (l. 168):**

„In South America, recent efforts have focused on the Paraná River basin, where hydrometric records starting in 1875 were rescued by digitizing paper-format data, initially from hand-drawn figures for the earliest years (Antico et al., 2018). Subsequent discovery of tabulated typewritten sheets allowed for an improved reconstruction of the 1875–1883 period, facilitating more precise corrections for gauge sinking and enabling the capture of extreme events like the 1877–1878 flood (Antico et al., 2020).“

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 project**.

Thank you also for your minor comments which I **have implemented**:

- Title changed
- Brackets added (l. 17, l. 233, now: 238)

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 have fixed that and rephrased the paragraph to:**

„In South and Southeast Asia, twentieth-century river archives from the Mekong and major Indian basins contain daily stage charts; recent syntheses demonstrate how digitizing these historical levels can underpin assessments of infrastructure impacts and regime shifts once station metadata and vertical references are unified. Furthermore, river flow in the Brahmaputra basin has been reconstructed over seven centuries to characterize long-term water-level changes, though this was achieved by utilizing dendrochronology rather than solely through instrumented historical records (Rao et al., 2020). Similar examples of reconstruction using "archives of nature" include studies of the Ganges (Wasson et al., 2013) and the Mekong (Burnhill and Adamson, 2008).“

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. I have added a paragraph to the conclusions as follows (l. 689):**

„Further research may include correlating and aligning these data with information on floods and droughts acquired from other documentary sources, such as historical newspapers. Thanks to modern OCR and LLM-based information extraction, this integration appears highly feasible. However, the availability of local newspapers in digitized archives remains a limiting factor in many regions, especially the one discussed here.“

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. **I have added a reference to it in the conclusion (l. 691): „Examples like (Kendon et al., 2024) make a strong case for an effort to make all these data interconnected.“**

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**.