



Reconstructing nineteenth-century Danube river water levels with transformer-based computer vision

Malte Rehbein^{1,2}

¹Chair of Computational Humanities, University of Passau, Passau, Germany

²Max Planck Institute of Geoanthropology, Jena, Germany

Correspondence: Malte Rehbein (malte.rehbein@uni-passau.de)

Received: 19 October 2025 – Discussion started: 5 December 2025

Revised: 13 February 2026 – Accepted: 23 February 2026 – Published: 10 March 2026

Abstract. We convert nineteenth-century Bavarian Danube gauge charts (1826–1894) into daily water-level series referenced to gauge zero through a novel semi-automated workflow combining light document pre-processing, dewarping, transformer-based line extraction, pixel-to-curve calibration, and targeted human checks. A curated ground-truth sample supported benchmarking and uncertainty quantification. Across three representative gauges (Neu-Ulm, Vilshofen, Passau), the pipeline attains high series-level accuracy (mean composite score 0.979) while reducing manual effort by roughly an order of magnitude relative to full manual digitisation. Outputs include versioned datasets with page-level provenance, confidence scores, and methodological descriptors to ensure transparency and reuse. The approach offers a replicable template for rescuing analogue hydrometric records and enabling long-term analyses of extremes, regulation impacts, and ecological context. Data are openly available under CC BY 4.0 (<https://doi.org/10.5281/zenodo.17296750>, Rehbein, 2025).

1 Introduction

Long hydrological series allowing among others flood frequency analysis are essential for putting contemporary extremes, regulation impacts, and ecological responses into context (Brázdil et al., 2006; Neppel et al., 2010; Lucas et al., 2024; Wetter, 2017). Yet large volumes of nineteenth-century hydrometric information remain locked in historical archives, limiting their contribution to present analyses. This paper addresses that gap for the Bavarian Danube by converting hand-drawn annual gauge charts for the years 1826–1894 into daily, machine-readable water-level (stage) series with transparent uncertainties and page-level provenance.

Our sources are printed quarterly templates with a regular day-level grid and handwritten hydrographs (see Fig. 1 for an example). They exhibit typical historical heterogeneity: paper warp and book curvature, stroke shadowing and show-through, inconsistent drawing styles, unit transitions, and occasional changes to gauge zero. Any practical reconstruction must therefore be robust to variable image quality while keeping human effort modest and the process auditable.

We make the following contributions. (i) 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), and a pixel-to-curve calibration tied to documented gauge zero, with targeted human checks at clearly defined checkpoints. (ii) A curated ground-truth sample and an evaluation protocol that report standard errors (RMSE/MAE, Pearson's r) alongside a peak-aware composite score that better reflects the historical charts' intended use. (iii) A case study on three representative gauges (Neu-Ulm, Vilshofen, Passau) demonstrating high series-level fidelity while substantially reducing manual digitisation effort. (iv) An open, versioned release of the reconstructed data with month-level provenance, uncertainty flags, and code pointers to enable verification and reuse (CC BY 4.0; <https://doi.org/10.5281/zenodo.17296750>, Rehbein, 2025).

The remainder of the paper introduces the sources and their structure; details the reconstruction method and imple-

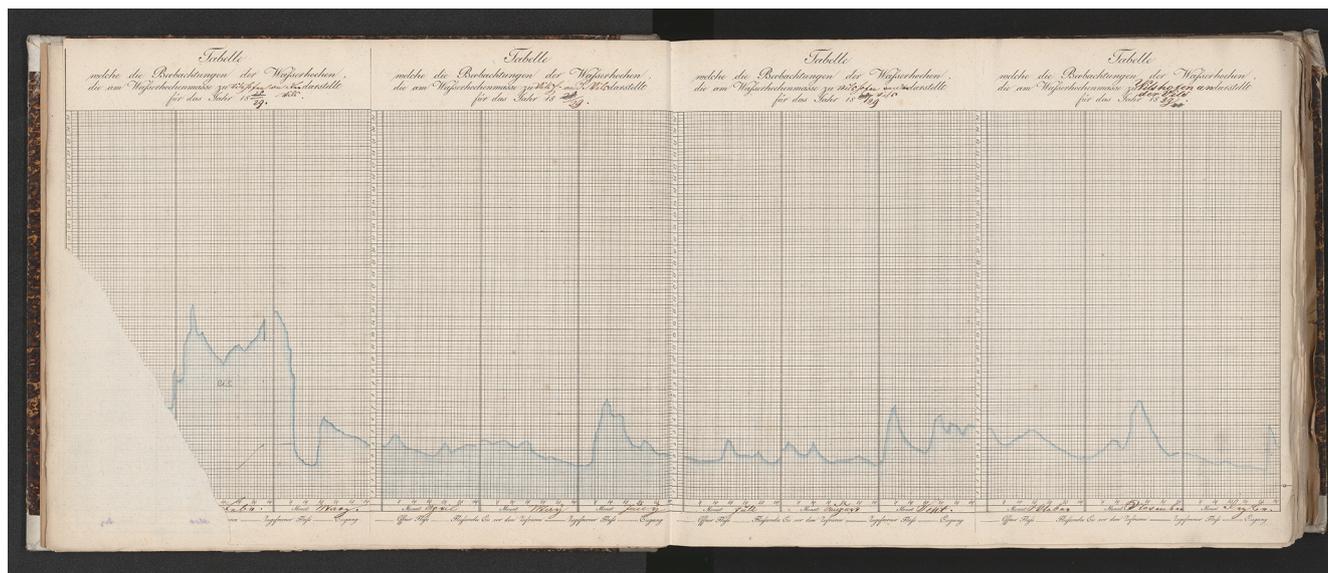


Figure 1. Sample from the Vilshofen gauge chart (year 1829). It also illustrates one of the (rare) cases of paper blight. Source: BayHStA BLW 29, page 7.

mentation choices; presents accuracy and efficiency results, including an analysis of systematic errors and mitigations; and discusses limitations, generalisability, and maintenance as a living dataset. Together, these elements offer a replicable template for mobilising historical hydrometric charts at scale.

1.1 Regional surveys of historical water-level data rescue

1.1.1 Americas

Across the Americas, tide- and river-level reconstructions combine archival discovery, datum control, and digitisation. NOAA's Joint Archive for Sea Level (JASL) provides a backbone for nineteenth-century tide data, with scanned marigrams and homogenised hourly records (Caldwell et al., 2001). Reviews identified large reserves of marigrams and registers and established workflows for recovery (Talke and Jay, 2013); NOAA's CDMP scanned thousands of tsunami charts. River data show similar promise: for the Lower Columbia, nineteenth-century stage records were digitised and merged with modern observations to assess long-term estuarine change (Helaire et al., 2019). These cases highlight archive scanning plus benchmark control as key to producing robust, machine-readable series. 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).

1.1.2 Africa

African archives contain large volumes of undigitised meteorological logs, many dating back to the late nineteenth and early twentieth century. Recent work in the Democratic Republic of the Congo has scanned thousands of such records from 37 stations and applied machine-learning methods (*MeteoSaver*) to transcribe handwritten tables with high accuracy (Castelvecchi, 2025). These efforts illustrate both the scale of Africa's pre-digital heritage and the emerging role of AI in converting fragile paper series into usable long-term climate data.

1.1.3 Asia

Asia holds extensive analogue tide and river records with strong rescue potential. Around the Indian Ocean, recovery of marigrams has enabled re-analysis of early twentieth-century tsunami and tide events (Matsumoto et al., 2009). In India, long tide-gauge series such as Mumbai/Colaba, beginning in the late nineteenth century, have been re-examined with modern methods to assess long-term sea-level change (Unnikrishnan and Shankar, 2007). 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).

1.1.4 Oceania

Australia and New Zealand have progressively rescued tide and river records by scanning, digitising, and tying them to geodetic networks. At Williamstown (Port Phillip Bay, Victoria), digitised registers and marigrams extended the local sea-level record by almost a century and were accompanied by open workflows (McInnes et al., 2024). Fremantle (Western Australia) illustrates the need for careful land-motion corrections, as geodetic studies revealed non-linear subsidence at the tide-gauge site (Featherstone et al., 2015). At Port Arthur (Tasmania), mid-nineteenth-century observations (1841–1842), tied to a historic benchmark on the “Isle of the Dead,” were recovered and merged with modern records to produce a continuous reference series (Hunter et al., 2003). In New Zealand, long-running digitisation efforts have combined historical gauges (Auckland, Wellington, Lyttelton, Dunedin) with geodetic control to quantify regional sea-level change (Hannah and Bell, 2012).

Beyond tide gauges, Australia also hosts notable long-term hydrological series. Reconstructions of Murray River streamflow extend variability estimates back to the late eighteenth century (Gallant and Gergis, 2011), while a two-century compilation of Lake George (NSW) water levels integrates diverse archival observations into a consistent reference series (Short et al., 2021). In New Zealand, archival staff-gauge readings, some extending to the late nineteenth century, document daily river levels prior to the advent of automated instrumentation (<https://archivescentral.org.nz/horizons-regional-council/recordset/hydrology-stage-%28water-level%29-daily-staff-gauge-readings>, last access: 5 March 2025). Collectively, these tide- and river-level initiatives demonstrate a mature template for historical water-level reconstruction: scan and catalogue legacy sources, digitise with quantified uncertainties, unify vertical references through benchmark histories, and publish harmonised series with documented provenance.

1.1.5 Western Europe

European work combines data archaeology, datum control, and image-based digitisation. Frameworks emphasise marigram/register rescue and harmonisation (Woodworth et al., 2016; Latapy et al., 2023; Khan et al., 2023). Case studies show accurate chart-to-series conversion: Dún Laoghaire marigrams in Ireland (McLoughlin et al., 2024), Williamstown (McInnes et al., 2024), River Ouse (Macdonald and Black, 2010), and Thames ledgers (Inayatillah et al.,

2022). Longer records (e.g., Liverpool, British Isles) highlight the need for careful quality control (Woodworth and Blackman, 2002; Bradshaw et al., 2016). For rivers, scanning of mechanical charts plus metadata curation extends sparse digital records; Irish catchment work offers replicable methods (De Smeth et al., 2024). Document restoration, warping correction, and datum management are essential to produce robust series.

1.1.6 Central Europe

Central Europe’s long tide- and river-gauge traditions have yielded many successful rescues. Along the German Bight and Baltic, analogue records (e.g., Cuxhaven, Wismar) were scanned, datum-checked, and homogenised for trend analysis (Dangendorf et al., 2013, 2014; Kelln et al., 2019). Documentary and instrumental evidence for the Elbe and Oder provide multi-century chronologies (Becker and Grünewald, 2003). Curated datasets consolidate metadata for reproducible re-engineering (Kelln et al., 2019). Datum unification via PSMSL RLR and GNSS benchmarks remains central (Holgate et al., 2012). River archives likewise benefit from careful metadata curation and image-based line extraction (Becker and Grünewald, 2003). Complementary European frameworks document end-to-end procedures for digitisation and harmonisation (Latapy et al., 2023; Inayatillah et al., 2022).

1.1.7 Danube

The Danube combines documentary sources (flood marks, accounts, newspapers) with institutionalised hydrometry. Austria’s *Hydrographisches Jahrbuch* (from 1893) provides daily to sub-daily summaries now digitised and online, anchoring older records to station datums. In Vienna, nineteenth-century regulation and dense documentation contextualise gauge data, as engineering works introduced datum shifts (Winiwarter et al., 2013; Hohensinner et al., 2013a, b). For the Middle Danube, compilations of medieval-modern floods situate nineteenth-century gauge data in a long-term context (Kiss and Laszlovszky, 2013).

2 Description of Extant Records (Materials)

2.1 Historical Context

Hydrometric data, measurements of river stage and flow, have long been foundational to human settlement, agriculture, and infrastructure planning. In ancient Egypt, water level monitoring along the Nile dates back to at least 3000 BCE, forming the basis for predicting seasonal floods and organising agricultural cycles (Willems and Dahms, 2017; Spanoudi et al., 2021). Parallel advancements in ancient Greece and China established foundational elements for water-level monitoring (Needham et al., 1971; Berking

et al., 2019; Koutsoyiannis et al., 2008). The systematic use of hydrometry in Western Europe was largely forgotten until the Renaissance, however (Biswas, 1970). It was not until the 18th and 19th centuries that sustained, systematic approaches to river monitoring re-emerged, driven by growing urban populations, expanding infrastructure, and repeated flood disasters. By the early 1800s, a network of river gauges and water level charts had been established across Europe, including in the Kingdom of Bavaria – one of the largest German states with then 75 865 km² and access to both central European large streams: Rhine and Danube. This period also marked the rise of large-scale land surveys and environmental data collection for purposes ranging from state administration to nation-building.

Water-level observation underpinned navigation, flood protection, and water-resources management throughout the nineteenth century. In Bavaria-mirroring broader European developments-hydrometric practice became increasingly institutionalised as river training, urban growth, and industrial land-use change accelerated. This institutionalisation is visible in administrative reforms, technical rulebooks, and the emergence of printed hydrological yearbooks and standardised gauge documentation (Bayerisches Landesamt für Umwelt, 2020); see also the historical overview provided by the Bavarian environmental administration (https://www.lfu.bayern.de/wasser/wasserstand_abfluss/entwicklung_pegelwesen/index.htm, last access: 5 March 2025).

A concise synthesis of the hydraulic transformation of Bavarian rivers is provided by Scheurmann (1981), who traces nineteenth- and early twentieth-century interventions motivated by flood control, land reclamation, and navigation. Major rivers, including the Isar, Vils, Amper, and the Danube, were straightened, confined, and stabilised using fascines, levees, groynes, and weirs. One illustrative case is the Danube between Ingolstadt and Großmehring, where channel realignment and embankment works (ca. 1826–1830) fundamentally altered local hydraulics and floodplain connectivity (City Museum Ingolstadt: <https://www.ingolstadt.de/stadtmuseum/stadtmuseum/scheurerer/museum/r-01-003.htm>, last access: 5 March 2025). Archival base maps from the Bavarian land survey (*Uraufnahme*) already display projected realignments, offering a cartographic window into early planning and design choices (Landesamt für Digitalisierung, Breitband und Vermessung: https://www.ldbv.bayern.de/mam/ldbv/dateien/faltblatt_historischekarten.pdf, last access: 5 March 2025).

From ca. 1700 to the late nineteenth century, Danube water-level information survives in documentary registers, municipal accounts, newspapers, flood marks, and later hydrographic ledgers. Taken together, these sources enable reconstructions of flood chronologies, relative levels, and, where cross-sections or benchmarks are known, approximate stages. With the establishment of dedicated hydrographic services and printed year-

books in the late nineteenth century, qualitative accounts increasingly dovetail with systematic instrumental observations (https://www.lfu.bayern.de/wasser/wasserstand_abfluss/entwicklung_pegelwesen/index.htm, last access: 5 March 2025).

Administrative and legal frameworks for hydraulic engineering and hydrometric observation were codified in progressively more detailed technical guidelines. Notably, the *Technische Vorschriften für den Wasserbau an den öffentlichen Flüssen in Bayern* (21 November 1878) standardised procedures for surveying and documenting works on public rivers; programmatic summaries followed in 1888 and 1909. This lineage culminated in the *Vorschrift und Vollzugsanweisung für Flussausstattung, Flussaufnahmen und deren Verarbeitung* (1930), which remained legally binding until 1958 (Bayerisches Landesamt für Umwelt, 2020; Weigmann, 1935; Faber, 1903). These regulations unified practices for river surveying, gauge-station documentation, and the graphical representation of water levels across Bavaria.

In terms of technical terminology, German nineteenth-century hydrometric literature may refer to water levels as *hydrometrische Coten* – positive or negative water heights relative to a fixed gauge zero (*Pegelnull*) – a convention mirrored in the layout of historical charts analysed in this study (Schrader, 1884). As state hydrologic services matured, guidance converged on consistent gauge datums and reporting standards; in Bavaria, the first *Gewässerkundliches Jahrbuch* with water levels from Bavarian gauges appeared in 1895, and a unified *Pegelvorschrift* was introduced across Germany in 1935 (https://www.lfu.bayern.de/wasser/wasserstand_abfluss/entwicklung_pegelwesen/index.htm, last access: 5 March 2025).

2.2 Records

- *Provenance and format.* 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*. The record extends back to the early nineteenth century and consists of a continuous, regionally extensive series of hand-drawn hydrographs on pre-printed forms. Each standard form (approx. 20 × 34 cm) covers *one quarter* (three months) with a daily grid, a vertical scale and unit annotations; a hand-drawn curve traces the water level at daily resolution. Four quarterly forms were glued edge-to-edge to create a single annual sheet, typically annotated with units, vertical ticks, and occasional later corrections. Series of annual sheets for a given gauge were then bound into volumes; some sheets carry additional notes on the reverse.

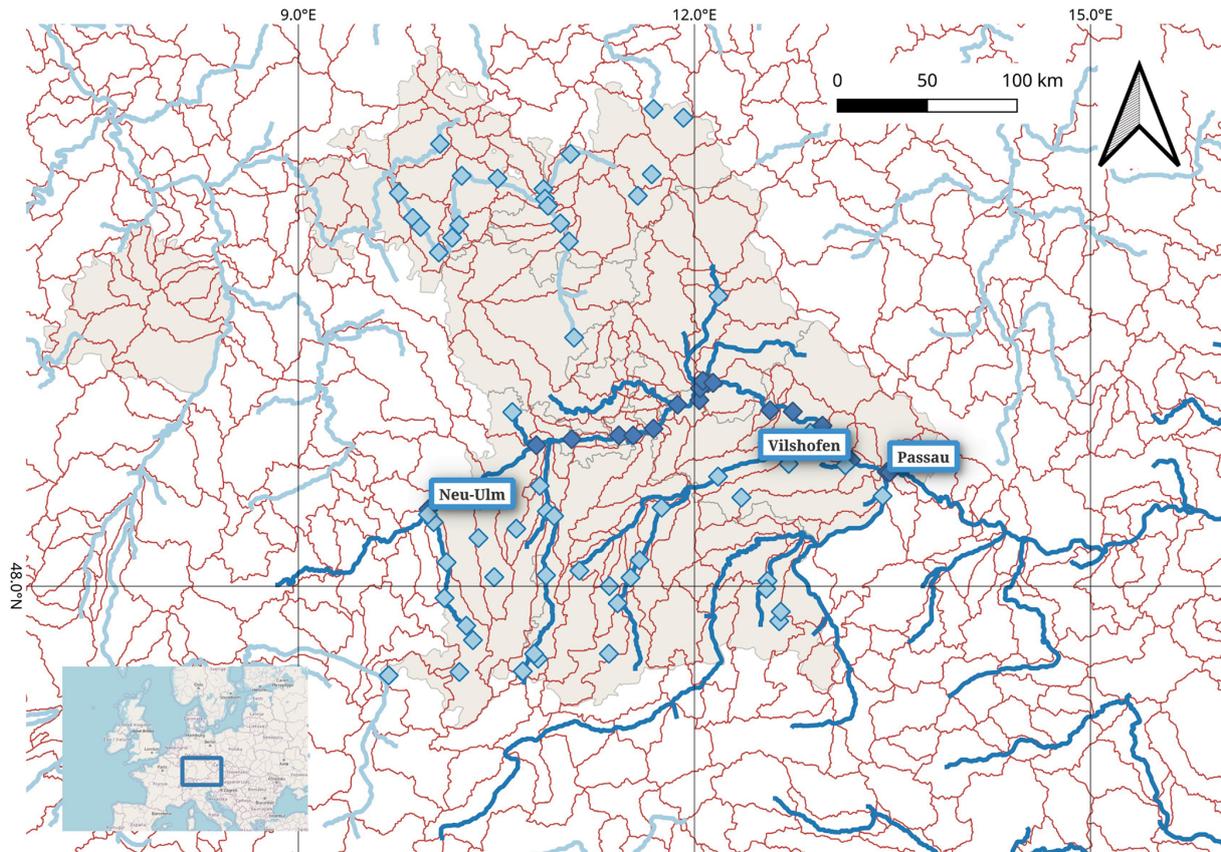


Figure 2. Kingdom of Bavaria in its administrative borders from 1845 (grey underlay). Danube and its tributaries (dark blue), other main rivers (light blue). HydroBasins Level 7 (red). Gauge stations for which level recordings are reported: along the Danube (dark blue) and other rivers (light blue). Gauges stations for this study (highlighted). Sources: Bayerisches Hauptstaatsarchiv, HGIS Germany, OpenStreetMap (© OpenStreetMap contributors, HydroBasins, AQUASTAT (FAO)).

- *Compilation practice.* Visual inspection shows that the annual sheets were not updated day-by-day but compiled retrospectively, possibly on a quarterly cadence. This implies the existence of primary daybooks or tabular ledgers from which the quarterly charts were prepared. For the Bavarian Danube, such underlying tabular sources have not been located to date; consequently, the bound charts constitute the most complete surviving record for the nineteenth century.
- *Archival indexing and coverage.* The archival finding aid (*Findbuch*) maps call numbers to gauges and calendar periods, which we use to construct a station-time index for sampling and comparison. For the Bavarian Danube in the nineteenth century, our survey identifies 20 gauge stations with observation spans from 9 to 113 years (Fig. 2); Regensburg is the longest continuous series. 14 of these stations exceed 50 years of coverage, offering strong potential for long-term hydrological analysis (Table 1).

2.3 Case study

The charts function both as hydrological records and as administrative artefacts. Although layouts and production practices are not fully uniform across years and stations, their broadly standardised format, dense daily annotation, and long temporal coverage make them well suited to a semi-automated reconstruction workflow.

We analyse three representative gauges along the Bavarian Danube: Neu-Ulm, Vilshofen, and Passau, spanning the upper to lower reaches of the stream within the country (Fig. 3). Together they offer long temporal coverage and markedly different visual characteristics (paper stock, handwriting, template variants), providing a robust and diverse test bed for semi-automated extraction.

3 Water-Level Reconstruction

3.1 Data Rescue from Historical Hydrographs

Despite their scientific value, much of this pre-digital record remains underused. Yet, climate and environmental data-

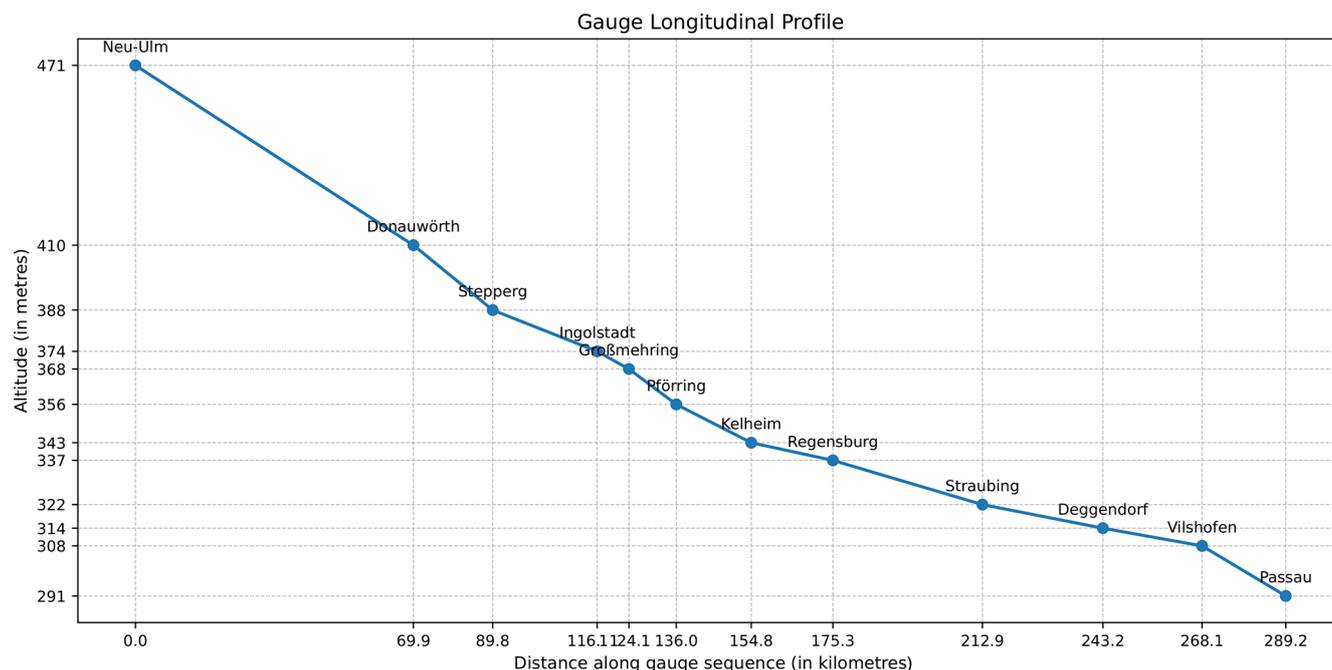


Figure 3. River longitudinal profile of the 19th century Bavarian Danube gauges; distances between gauge station as the crow flies (standard great-circle (Haversine) distance).

Table 1. Nineteenth-century Bavarian Danube gauges in the archive; case-study gauges highlighted in bold font. 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.

| gauge_id | gauge_name | longitude | latitude | altitude | reporting period |
|---------------------|--------------------|---------------|---------------|------------|------------------|
| DE_BY_DAN_21 | Neu-Ulm | 9.987 | 48.390 | 471 | 1826–1894 |
| DE_BY_DAN_22.1 | Donauwörth | 10.803 | 48.711 | 410 | 1848–1889 |
| DE_BY_DAN_22.2 | Pförring | 11.691 | 48.794 | 356 | 1848–1889 |
| DE_BY_DAN_23 | Ingolstadt | 11.428 | 48.760 | 374 | 1826–1894 |
| DE_BY_DAN_24.1 | Großmehring | 11.537 | 48.760 | 368 | 1848–1894 |
| DE_BY_DAN_24.2 | Stepperg | 11.071 | 48.741 | 388 | 1848–1894 |
| DE_BY_DAN_25 | Kelheim | 11.873 | 48.914 | 343 | 1826–1894 |
| DE_BY_DAN_26 | Regensburg | 12.102 | 49.021 | 337 | 1826–1894 |
| DE_BY_DAN_27 | Straubing | 12.574 | 48.886 | 322 | 1826–1894 |
| DE_BY_DAN_28 | Deggendorf | 12.971 | 48.809 | 314 | 1826–1894 |
| DE_BY_DAN_29 | Vilshofen | 13.187 | 48.636 | 308 | 1826–1894 |
| DE_BY_DAN_30 | Passau | 13.459 | 48.577 | 291 | 1826–1894 |
| DE_BY_DAN_31.1 | Passau (Ilzstadt) | 13.479 | 48.575 | 295 | 1826–1894 |
| DE_BY_DAN_31.2 | Passau (Ortspitze) | 13.477 | 48.574 | 290 | 1826–1894 |

rescue efforts emphasise that such sources are essential for extending baselines, quantifying long-term change, and improving extremes assessment (Brönnimann et al., 2018; Hawkins et al., 2023a). Long historical series enable retrospective flood chronology, attribution of anthropogenic impacts (land-use change, river training, canalisation), and independent validation of hydrological and hydraulic models

used for planning and adaptation. They also support public communication of changing hazards by placing recent events in a longer context (<https://climatelabbook.substack.com/p/visualising-climate-impacts-on-uk>, last access: 3 March 2025). Recent reports underpin the importance of acquiring historical data (World Meteorological Organization, 2025). All these considerations are in line with single and coor-

minated efforts of data rescuing on all scales and in all areas of Earth System and Anthropocene Studies (e.g., *Copernicus*, “the European Union’s Earth Observation Programme”: <https://datarescue.climate.copernicus.eu/>, last access: 3 March 2025, Huffman et al., 2023). Especially in the context of water as an essential good, we hope to provide to a historical perspective (WBGU, 2025a) to create a climate resilient strategic water management (WBGU, 2025b).

Recovering structured data from historical scientific graphics is variously termed *datafication of historical sources* (Donig and Rehbein, 2022; Rehbein, 2024; Navarro et al., 2025), *reverse engineering*, *data archaeology*, or *data rescue* (Poco and Heer, 2017; Wilkinson et al., 2019; WMO, 2024). The aim is to reconstruct machine-readable series (with associated metadata) from pre-digital figures when the original tabular sources are lost, incomplete, or inaccessible. In hydrology, this typically means converting hand-drawn hydrographs into daily or sub-daily level records with explicit uncertainty and datum control. For sea-level based data rescue, Latapy et al. (2023) give a comprehensive survey. Also seismology provides a closely related precedent: analogue seismograms on smoked paper or film have been digitised by tracing instrument curves into time series, with both manual and automated workflows now in use (Wang et al., 2019; Bogiatzis and Ishii, 2016; Paulescu et al., 2016; Aghajani et al., 2023; Wang et al., 2025). 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).

In other fields such as medicine and chemistry, chart digitisation has advanced through shared benchmarks and competitions. ICDAR, for example, has stimulated progress in chart recognition and curve vectorisation by releasing annotated datasets and standardised evaluation tasks (Davila et al., 2019; Yepes et al., 2021). Extending similar frameworks to hydrology and geoscience would enable reproducible pipelines, shared ground-truth collections, and benchmarked performance. Yet these initiatives mainly address clean, modern prints and leave aside the distinctive problems of historical sources. Unlike modern, digitally rendered plots with uniform backgrounds and crisp axes, the archival charts present heterogeneous artefacts that complicate line detection and coordinate recovery. The most relevant categories and their effects are (see Fig. 5 for examples):

- *Page/scan geometry*. Spine-induced curvature, local warp, minor tears, and quarter-join misalignment. These introduce slow baseline drift and small discontinuities that can bias the pixel \rightarrow level conversion and day assignment;
- *Background and ink*. Paper ageing, show-through, and ink bleed produce textured backgrounds; pigment loss (e.g. crayon rubbed thin) reduces stroke contrast. Both

lower the signal-to-noise ratio and increase false positives;

- *Grid and layout variability*. Changes in grid alignment, vertical scaling, axis labelling, and template versions cause domain shifts that a model must accommodate; slight horizontal spacing inconsistencies can nudge date bins;
- *Scribe/style effects*. Variations in pen colour, line thickness, and drawing style across years and hands yield multi-modal stroke appearances across stations, within a single station’s series, and even within a station year;
- *Editorial marks and anomalies*. Sloppy drawing, later corrections, marginal annotations, occasional discontinuities of the curve, erroneous dates (e.g. 30 February), extra paper added for floods, and imperfectly glued quarters create distractors and small step offsets.

Despite these challenges, the charts are broadly regular in layout and densely annotated at daily resolution, which makes them well suited to *structured* extraction with transformer-based models when paired with light, grid-aware calibration and simple plausibility checks (details in Sect. 4). Figure 4 illustrates a common failure mode: book-spine-induced curvature that must be accounted for during calibration.

3.1.1 Problem structure

Across disciplines, the essential steps, whether conducted consecutively or in one go, are similar: (1) figure parsing (panel segmentation, axes and grid localisation), (2) coordinate system recovery (mapping pixels to physical units under a declared datum/zero), (3) trace extraction (identifying and vectorising the plotted curve), and (4) transform vectors into data using the coordinate system. Each step interacts with documentary conventions: changing templates, variable grid spacing, marginal notes, and later corrections. For hydrometric charts, the additional requirement is *datum consistency* across months/years (gauge zero, units), so that reconstructed series can be compared across time. Community practices in sea-level work (e.g., revised local reference schemes) illustrate the importance of such cross-record unification.

3.1.2 River and tide hydrographs

River charts and tide charts present different signal and layout regimes. River hydrographs emphasise seasonal cycles and episodic floods, often with long panels and coarser vertical ticks; tide charts exhibit strong semidiurnal/diurnal periodicity and denser vertical scaling (Latapy et al., 2023; Talke et al., 2020). In both cases, page artefacts (paper warp, spine curvature, quarter-join offsets), heterogeneous inks, and template revisions complicate automated extraction. The common core is a single, continuous hand-drawn line that must

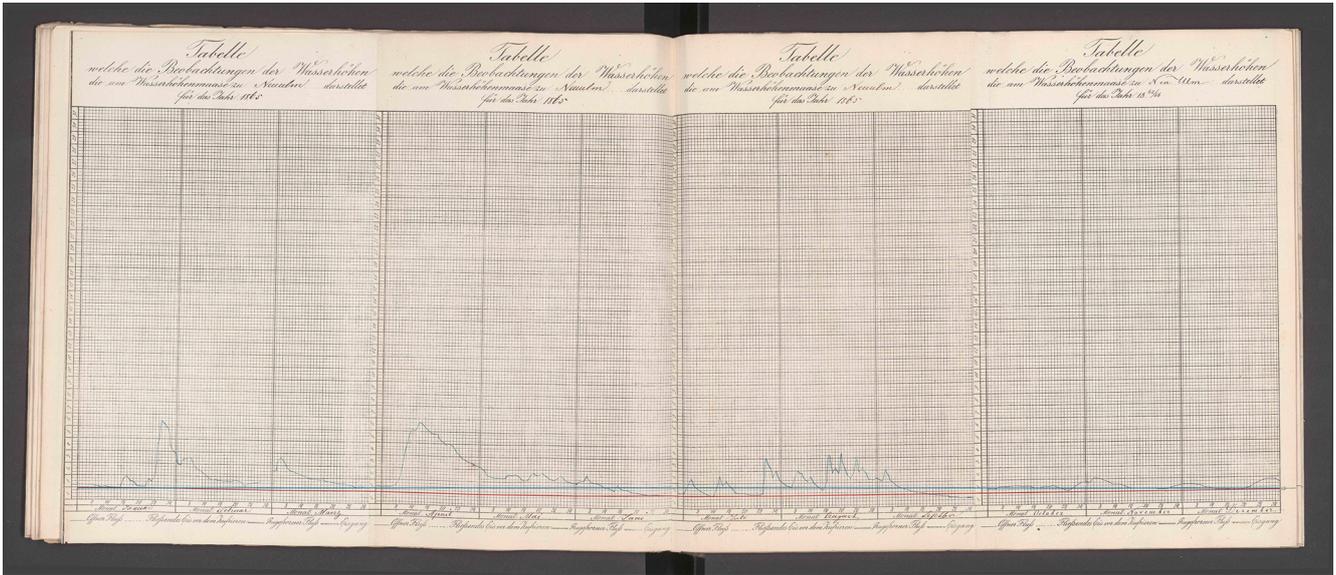


Figure 4. Book-spine-induced page curvature persists despite an orthogonally aligned scan. The blue line marks a constant y-pixel across the image; the red line indicates the form's zero-foot baseline. Around mid-year, near the spine bulge, the offset reaches about 21.9 cm in gauge level and must be corrected during calibration (Passau, 1865). Source: BayHStA BLW 30.

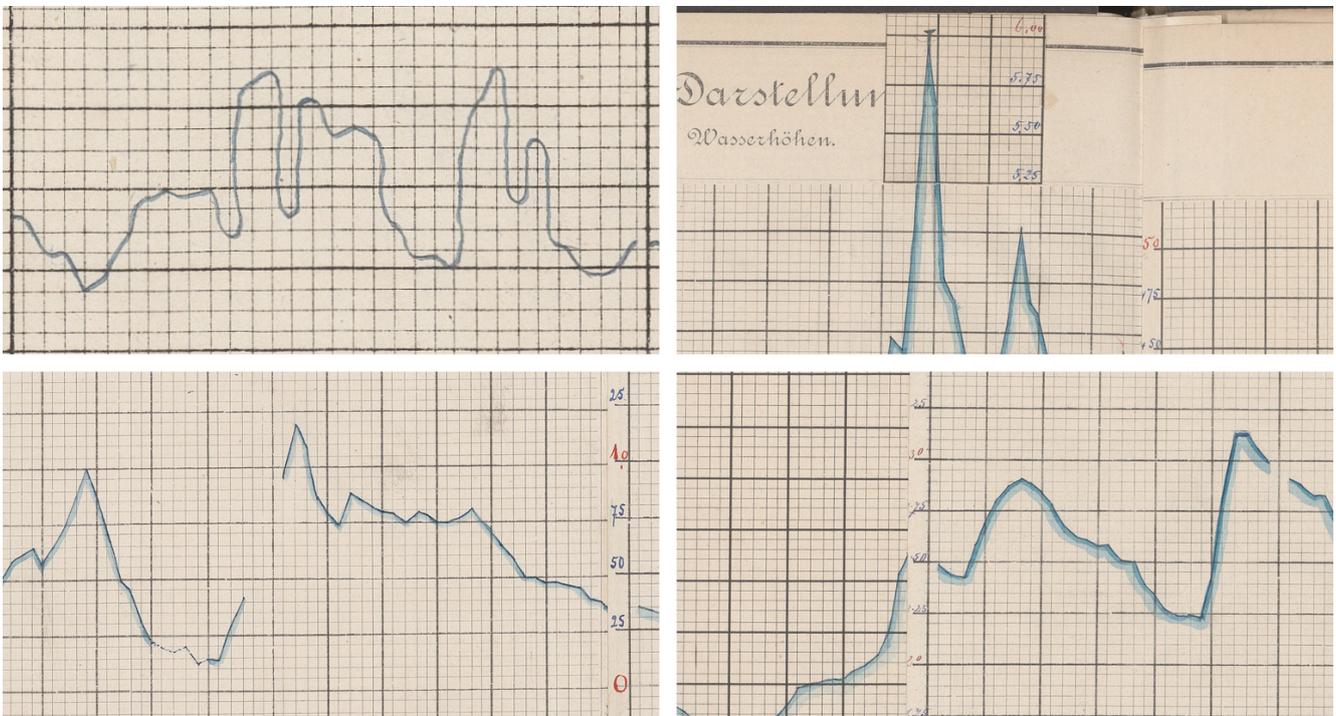


Figure 5. Samples illustrating some types of anomalies in line chart drawing often found in the sources. The selection also illustrates some of the different drawing styles. Top left: some creativity employed in this sample with the curve even going backwards (Neu-Ulm, 1826); top right: extra paper added for covering high flood (Passau, 1892); bottom left: non-continuous line between months February and March (Vilshofen, 1894); bottom-right: misaligned gluing of quarterly forms (Passau, 1892). Image sources: BayHStA BLW 21, 29, 30.

be followed through noise, gaps, and overprinted, written, or drawn annotations, but especially the panels/plateaus as visualised here caused additional challenges in differentiating them from the underlying grid (Fig. 6).

3.2 Data Extraction: Methods and Practice

Several strategies have emerged to extract data from analog scientific figures, each with trade-offs in scalability, accuracy, and labour intensity. In the following, we briefly discuss manual annotation, citizen science, computer vision and machine learning, and transformer-based approaches.

3.2.1 Manual Annotation

The most direct approach of graphical charts, is manual annotation, using specialised software tools such as TIITBA (Corona-Fernandez and Santoyo, 2023), which provides interfaces for tracing, vectorising, and correcting seismogram curves. While reliable, this process is time-consuming and difficult to scale. For large archives such as river gauge records spanning decades, fully manual workflows are prohibitive.

Manual (and semi-manual) digitisation remains a baseline strategy for re-engineering analogue charts and ledgers when automation is unreliable or when provenance requires human verification. Typical workflows comprise creating high quality images from the originals (normally digital photographs with 300 ppi resolution as standard, flattened surface, orthogonal alignment, and adequate uniform lighting) (Wilkinson et al., 2019; Federal Agencies Digital Guidelines 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).

In sea-level work, manual and semi-automated digitisation have recovered long, high-frequency series directly from paper marigrams – for example Dún Laoghaire (Ireland, 1925–1931), where careful geometric correction and uncertainty quantification were integral to the workflow; Belfast Harbour (UK, 1901–2010), which used scanning plus line-seeking software to reconstruct a century-scale record; and Socoa/Saint-Jean-de-Luz (France), where a data-archaeology programme combined scanning, metadata curation and quality control to extend the series back to 1875 (McLoughlin et al., 2024; Murdy et al., 2015; Khan et al., 2023). For ledger-based archives (e.g. Thames high/low waters), manual transcription yields long homogeneous series when coupled with rigorous checks and clear provenance (Inayatillah et al., 2022). Reviews of sea-level data archaeology emphasise that analogue-to-digital recovery should document benchmark ties, instrument changes, and vertical da-

tums alongside the digitised series (Bradshaw et al., 2015; Latapy et al., 2023).

Manual digitisation offers transparency and expert control at page level and can achieve high accuracy on clean charts, but it is labour-intensive and sensitive to operator drift. Best practices therefore include (i) double-keying or replicate tracing for a sample of pages and (ii) inter-coder checks to quantify bias (Drevon et al., 2017). Where feasible, semi-automated toolkits tailored to marigrams (e.g. NUNIEAU) can accelerate tracing while preserving quality control (Ullmann et al., 2005).

3.2.2 Citizen Science Initiatives

Citizen science (CS) has become a practical accelerant for mobilizing historical sources in general (Rehbein and Ernst, 2023) and for re-engineering analogue environmental records at scale in particular (Navarro et al., 2025). Successful campaigns decompose digitisation into micro-tasks (e.g. transcription of register entries, tracing of plotted lines, identification of unit changes), replicate inputs across multiple volunteers, and fuse results by consensus with expert review. In this way, CS delivers high-value datasets while preserving provenance and audit trails.

Large programmes have mobilised volunteers to transcribe historical meteorological records, including the UK&Ireland Rainfall Rescue (Hawkins et al., 2023b; Hawkins, 2025), *Unearthing Australia's climate history* (<https://climatehistory.com.au/>, last access: 3 March 2025), Canada's *DRAW* project (<https://draw.geog.mcgill.ca/>, last access: 3 March 2025), and *Jungle Weather* (<https://www.zooniverse.org/projects/khufkens/jungle-weather>, last access: 3 March 2025). Long-running efforts initiated by *Old Weather* (<https://www.oldweather.org/>, last access: 3 March 2025) continue to recover ship log observations of pressure, temperature, wind and sea state (Brohan et al., 2009; Hawkins et al., 2019; Burr et al., 2021; Westcott et al., 2021; Teleti et al., 2024), while regional campaigns (e.g. Southern Weather Discovery, Antarctic logbooks) show how CS can target specific archives (Lorrey et al., 2022; Lakkis et al., 2023). For a broader overview, see <https://datarescue.climate.copernicus.eu/citizen-science> (last access: 3 March 2025).

In Germany, the NFDI4Earth incubator pilots explicitly invite volunteers to datify historical river gauge records, turning scanned hydrographic charts into structured level series (<https://git.rwth-aachen.de/nfdi4earth/pilotsincubatorlab/incubator/do-it>, last access: 3 March 2025). Here, CS tasks extend beyond transcription: participants help identify axes and baselines, flag warped or damaged pages, and trace short line segments where automated extraction fails. Such experiments demonstrate how CS can be integrated directly into hydrographic data rescue, providing training and validation data for computer-vision models while ensuring quality control through human oversight.

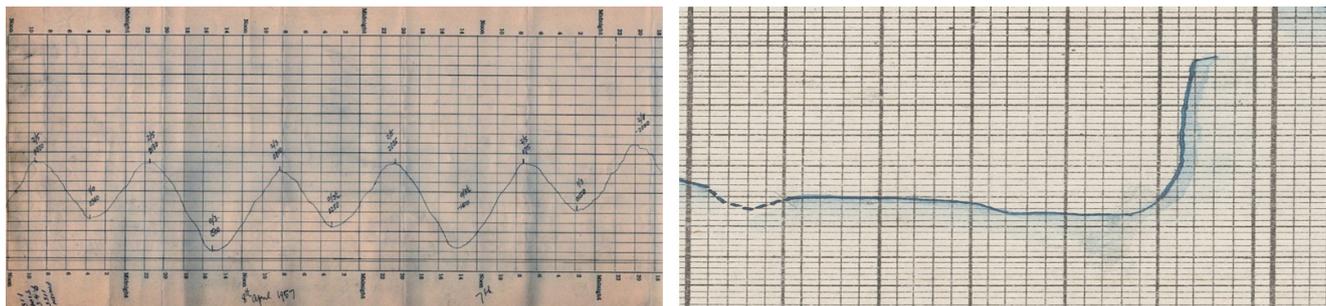


Figure 6. Juxta-position of a typical tidal sea level record (marigram) and a particularly flat period of a river water level. Left: Williamstown (Australia) tide data (1875–1965, here: 1957) illustrating typical periodicity. Source: Mcrae (2018); right: sample of a panel in river gauge level (Neu-Ulm, February 1873). Image source: BayHStA BLW 21.

Effective CS re-engineering projects typically (i) define unambiguous micro-tasks with visual examples; (ii) use n -fold replication with majority/weighted consensus; (iii) seed “gold-standard” items to measure volunteer accuracy and provide training feedback; (iv) surface low-agreement items for expert decision; and (v) maintain open, versioned data releases. For graphical sources (like gauge charts), CS tasks can include locating axes and baselines, flagging months with warping or template changes, and tracing short segments where automatic line-following fails.

CS scales human attention and can rapidly unlock otherwise inaccessible archives, but throughput depends on participant engagement and careful task design. Although automation is the ultimate goal in data rescue, CS may complement the process when manual attention is still required: for (a) generating training and validation data for machine-learning models (e.g. pixel-level line annotations, baseline marks), although synthetic data generation might be a viable option, too; (b) metadata enrichment (e.g. unit changes, marginal notes, year-page mapping); and (c) targeted quality control of months flagged as ambiguous by the predictor. This hybrid strategy balances speed with reliability and reduces expert time spent on routine corrections.

3.2.3 Computer Vision and Machine Learning

Automated extraction from analogue charts has evolved from classical computer-vision pipelines to transformer-based systems that can detect chart components, reconstruct data tables, and reason about values. Early pipelines relied on edge detection, Hough transformation, and active-contour (snake) models to trace plotted lines and grid structures. While effective on clean, modern figures, they are brittle on historical scans with ageing, noise, and warp (Kass et al., 1988). Interactive hybrids combine chart-type classification with semiautomated mark extraction and user corrections, improving robustness on heterogeneous material (Savva et al., 2011; Jung et al., 2017).

Recent deep-learning systems target “chart-to-table” derendering and chart understanding. *ChartOCR* introduced

a hybrid framework that detects chart type and structural elements (axes, legends, lines) to reconstruct underlying values, and remains a strong baseline for line-chart data extraction (Luo et al., 2021). *ChartReader* unifies derendering and comprehension with transformer components, reducing heuristic rules and improving cross-task generalisation (Cheng et al., 2023). Question&Answer-style benchmarks such as *PlotQA* and *ChartQA* have driven progress on detection and reasoning over plots and provide large synthetic/curated corpora for pretraining and evaluation (Methani et al., 2020; Masry et al., 2022). Vision-language approaches like *DePlot* translate chart images into linearised tables that a language model can then query; while domain adaptation is needed for historical sources, the modality-conversion step is directly relevant to time-series recovery (Liu et al., 2023).

For historical gauge charts in general (though not in our case study), accurate recovery hinges on robust polyline extraction under noise, bleed-through, and curvature. Line-segment detectors and wireframe parsers (e.g. Holistically-Attracted Wireframe Parsing, HAWP) offer vectorised outputs with junction consistency and have been adapted to noisy, low-contrast imagery (Zhao et al., 2020; Xue et al., 2023).

Paper bulging from book spines and scanning skew introduce systematic geometric error that propagates to level estimates. Learning-based document dewarping – *DocUNet* and *DewarpNet* – predicts pixel-wise forward mappings from warped to flat coordinates, offering a principled pre-processing step before line extraction (Das et al., 2019). In our context, mild, grid-aware dewarping combined with per-month anchor calibration (Sect. 4.5.3) reduces day-shift and vertical bias.

Because annotated historical river charts are scarce, most systems rely on synthetic training corpora with realistic degradations (ink bleed, paper texture, yellowing, local warp) and self-supervised pretraining. Public chart datasets (e.g. *PlotQA*, *ChartQA*) and chart-derendering corpora (*ChartOCR/ChartReader*) provide pretraining targets; fine-tuning

on even small, well-curated hydrological samples markedly improves stability.

3.2.4 Transformer-Based Models

Transformers have become the workhorse for chart rendering and curve extraction because self-attention captures global context (axes, legends, gridlines) while retaining fine-grained detail along thin strokes.

End-to-end detectors in the *DETR* (DEtection TRansformers) lineage treat chart components as a set prediction problem, reducing post-processing and improving robustness to clutter (Carion et al., 2020). For line charts, *LineFormer*, which we build on in the following, reframes data extraction as instance segmentation of polylines and reports strong performance on synthetic and scanned material; its decoder naturally separates overlapping traces and simplifies downstream coordinate mapping (Carion et al., 2020; Lal et al., 2023).

Vision-language transformers *Donut* (Document Understanding Transformer) and *Pix2Struct* parse documents and UI-like artefacts directly to token sequences, while *DePlot* converts plots to linearised tables for subsequent reasoning by an LLM. *UniChart* pretrains on diverse chart corpora to support QA and summarisation. These models have been developed for modern graphics and not yet tested for adaptation to historical hydrological samples (Kim et al., 2022; Lee et al., 2023; Liu et al., 2023; Fields and Kennington, 2023; Masry et al., 2023).

Overall, transformer-based models, such as *LineFormer* which we chose for our study, offer a high-leverage starting point for nineteenth-century gauge charts.

4 Method of Reconstruction

4.1 Goals and design

The primary objective of this study is to convert nineteenth-century Bavarian Danube water-level charts into daily gauge-level series expressed in physical units (millimetres), referenced to the respective gauge zero. Beyond the technical reconstruction itself, the approach is guided by three overarching goals: (1) efficiency, i.e. minimising the need for manual intervention; (2) effectiveness, i.e. ensuring completeness and accuracy, and (3) reliability of the reconstructed data, i.e. providing a transparent, evaluated, reproducible and well-documented workflow.

The research design follows a stepwise structure aligned with these goals. First, evaluation criteria are defined to assess accuracy and practicability. Second, a representative sample is selected and a ground-truth dataset established. Third, an algorithm for pixel-to-curve transformation is formed, followed by an experimental pilot study to test the workflow. This includes image pre-processing, warping adjustments, and the optimisation of *LineFormer* hyperparam-

eters. The results are benchmarked against the ground-truth data. Fourth, systematic errors are analysed and the workflow iteratively adjusted. Finally, the optimised workflow and parameters are subjected to final benchmarking and validation, forming the basis for subsequent large-scale production.

4.2 Evaluation protocol

This section specifies how we quantify the accuracy of reconstructed daily gauge levels. Our choices are shaped by two constraints of the corpus: (i) evaluation is conducted *per month* (the operational unit for extraction and quality control), and (ii) we do not rely on pixel-level ground truth polylines for accuracy; instead, we evaluate against ground truth in original physical units. Consequently, curve-space distances (e.g., Hausdorff/Fréchet) are not used as headline measures. We do, however, report a panel of series-level metrics – RMSE, MAE (average deviation), maximum absolute deviation, `Pearson_r`, and an adopted *peak-aware composite score* as our primary objective for decisions in the pilot study (parameter selection) and subsequent evaluations.

4.2.1 Unit of analysis and dataset separation

Evaluation is conducted on a month-by-month basis; we never split within a month or use partial page fragments. Sampling is stratified across gauges and years to balance coverage; no model training is performed during these evaluations.

Hydrological use cases may care about physical levels and timing at daily resolution (flood peaks, seasonal lows). Series-level metrics in original physical units directly reflect those targets and are robust to purely graphical idiosyncrasies (ink density, micro-warp) that do not affect the numerical series. Pixel-space curve distances would require an additional layer of ground truth and can over-emphasise small geometric deviations that have negligible effect after conversion to levels. For this reason we (i) evaluate in level space, and (ii) reserve pixel-space diagnostics (continuity/fragmentation, edge hits) for internal quality control (by visual inspection, especially in candidate selection) rather than accuracy reporting.

4.2.2 Base metrics (series-level)

Let (y_1, \dots, y_N) be the ground-truth daily levels (in mm) for a given month m , $(\hat{y}_1, \dots, \hat{y}_N)$ the predicted levels after conversion from pixels in the scanned images, and $N \in \{28, 29, 30, 31\}$ the number of days of m . We then define the difference, i.e. the error between a data point in the prediction series and the ground truth, for a given date (day) t as $e_t = \hat{y}_t - y_t$.

– *Root mean squared error (RMSE)*. We consider

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{t=1}^N e_t^2} \quad (1)$$

a standard overall accuracy in data series comparison; squaring emphasises larger mistakes (e.g., missed peaks) and expresses error in physical units. While it is sensitive to outliers, it can be dominated by a few large errors; less interpretable when errors are highly skewed.

– *Mean absolute error (MAE)/average deviation (AD)*.

$$\text{MAE} = \frac{1}{N} \sum_{t=1}^N |e_t| \quad (2)$$

reports the “typical miss” in actual units (mm), it is the average deviation (AD) between two series. While it is robust and easily comparable across the corpus as it is in the same unit as the gauge levels itself, it is less sensitive to rare but important extreme errors.

– *Maximum absolute deviation (MaxAE)*.

$$\text{MaxAE} = \max_{1 \leq t \leq N} |e_t| \quad (3)$$

reports the worst-case point error over the month. It is useful because it explicitly guards against single catastrophic misses, but it ignores overall performance and can be noisy month-to-month. Both, MAE and MaxAE indicate the potential impact of errors as they are given in gauge level unit.

– *Pearson correlation coefficient (Pearson r)*. Pearson r measures co-variation of shape: r is invariant to affine changes of either series (additive offsets and positive rescalings), so it reports agreement in the pattern of rises and falls independently of datum or overall amplitude. It is also dimensionless and bounded in $[-1, 1]$, hence directly comparable across stations, years, and unit systems. Pearson r is widely used in time-series validation as a complementary view to magnitude-based errors (RMSE/MAE), helping to separate shape fidelity from bias and scale issues.

$$r = \frac{\sum_{t=1}^N (y_t - \bar{y})(\hat{y}_t - \bar{\hat{y}})}{\sqrt{\sum_{t=1}^N (y_t - \bar{y})^2} \sqrt{\sum_{t=1}^N (\hat{y}_t - \bar{\hat{y}})^2}}, \quad (4)$$

$$\bar{y} = \frac{1}{N} \sum_t y_t, \quad \bar{\hat{y}} = \frac{1}{N} \sum_t \hat{y}_t.$$

A high r confirms that peaks, recessions, and seasonal phases co-vary correctly even when an overall offset or gain error exists. This makes r particularly useful during pre-processing and datum control: combinations that preserve r while harming RMSE/MAE typically introduce bias/scale errors rather than distorting the hydrograph.

4.2.3 Peak-aware composite metric

We judge the hydrograph reconstruction by how well they reproduce both the overall shape and the seasonal/flood peaks in value and timing of the original. To capture these aspects in a single indicator, we adopt a custom peak-aware metric that combines and weighs three components:

1. *Pearson correlation*. r , as above. If either series is constant/invalid, we set $r := 0$ for stability.
2. *Peak value similarity*. Let $y_{\max} = \max_i y_i$ and $\hat{y}_{\max} = \max_i \hat{y}_i$. We define *peak value similarity* as:

$$s_{\text{peak-val}} = \begin{cases} 1 - \frac{|y_{\max} - \hat{y}_{\max}|}{y_{\max}}, & y_{\max} \neq 0, \\ 0, & y_{\max} = 0, \end{cases} \quad (5)$$

which equals 1 for an exact peak match and decreases linearly with relative peak error. Note: $s_{\text{peak-val}}$ is not clamped above or below $[0, 1]$; large over-prediction can make it negative. We implemented this behaviour to preserve sensitivity to extreme overshoots.

3. *Peak timing alignment*. Let $i^* = \arg \max_i y_i$, $j^* = \arg \max_i \hat{y}_i$, $d = |t_{i^*} - t_{j^*}|$ (days). We define *peak timing alignment* as:

$$s_{\text{peak-time}} = \max\left(0, 1 - \frac{d}{N}\right). \quad (6)$$

This lies in $[0, 1]$ by construction (1 for same-day peaks, 0 when the shift spans the month length). It penalizes delayed prediction: the closer to 0, the later the predicted peak was from the actual one.

The peak-aware composite metric is computed on aligned daily series after an inner join on dates. With user-chosen $\alpha, \beta \in [0, 1]$ and $\gamma = 1 - \alpha - \beta$, we define our custom peak-aware composite metric as:

$$\text{score} = \alpha r + \beta s_{\text{peak-val}} + \gamma s_{\text{peak-time}} \quad (7)$$

The upper bound of score is 1; the lower bound can be < 0 if $r = -1$ and/or $s_{\text{peak-val}} < 0$. When $\alpha + \beta \leq 1$, the terms form a convex combination; when $\alpha + \beta > 1$ (thus $\gamma < 0$), the timing term reduces the score – useful if timing errors must be penalised more heavily than value errors. The default choice in our reference implementation is $\alpha = 0.4$, $\beta = 0.4$ (thus $\gamma = 0.2$), which is a peak-sensitive setting emphasising overall shape and value of the curve without disregarding timing.

RMSE and MAE capture magnitude accuracy; Pearson r captures shape; none alone reflects event significance (peaks). The composite explicitly encodes the three desiderata – shape, peak amplitude, and peak timing – in a single, tunable quantity. This reduces multi-criterion ambiguity when selecting parameters in the pilot study (e.g., stretch factor, baseline aid, resampling rule): we can optimise a single objective that aligns with hydrological priorities, then audit trade-offs with the base metrics.

4.3 Source Preparation and Metadata Acquisition

- *Volumes and page-year mapping.* The original gauge level documentation is preserved by the Central Bavarian State Archive (*Bayerisches Hauptstaatsarchiv, BayHStA*) in bound volumes, typically one per gauge, covering long multi-decadal periods (1826–1894 in our corpus). Each year is assembled from four quarterly forms that were glued edge-to-edge into a single long slip before binding. Non-chart pages (notes, memoranda, corrections) are interleaved irregularly. Consequently, scan/page indices are not monotonic in calendar time. As a first step we compiled a book-level table of contents that maps each document identifier and page number to a unique (gauge, year) pair and flags pages without charts. This register underpins sampling and all subsequent tasks.
- *File organisation and provenance.* Scans are used by the archival identifier schema: ProvenanceID_GaugeIDPageID (e.g. Bay_Landesamt_fuer_Wasserwirtschaft_210085 for the records from Neu-Ulm, 1891). For each image we create a metadata record with: (i) gauge identifier, name and image size; (ii) nominal year; (iii) unit system noted on the form (Bavarian foot/mm); and (iv) page-level notes (annotations, overwriting, template changes). This ensures page-level provenance and allows exact reproduction of any result.
- *Metadata Extraction.* Each scan was linked to its calendar year and measurement unit by decoding the archival identifier and inspecting marginal notes. Interleaved occasional commentary pages required manual checks to ensure that the image–year mapping is unambiguous. Axis units change from Bavarian feet to millimetres on 1 April 1872 (Fig. 7); this transition is flagged in the pipeline and propagated to the coordinate-to-level conversion routines.

To complement the coordinate data, we manually transcribed the commentary pages for the entire document series, extending beyond the sample itself. This enriches the human contextual understanding of each chart and provides additional semantic layers for future modelling.

4.4 Ground-truth dataset construction

Reliable ground-truth data are essential for assessing the accuracy of predicted water-level curves and for calibrating the pixel-to-gauge transformation. We compiled two complementary ground-truth sets: (i) pixel-level annotations of the chart line and (ii) manually keyed gauge-level values. Together they form the basis for evaluation and pipeline optimisation.

4.4.1 Sample selection

Fifteen annual charts were randomly sampled from three Danube gauges – Neu-Ulm, Vilshofen and Passau – yielding five charts per site (Table 2). This stratified selection balances spatial coverage with manageable annotation effort and captures variation in handwriting, paper quality and measurement units.

4.4.2 Pixel-based ground truth

- *Annotation workflow.* Using an in-house tool `point_annotator` we traced the gauge curve at daily resolution (1 January–31 December). Each vertex stores pixel coordinates (x, y) relative to the full-page scan; an experienced annotator required ≈ 12 –15 min per annual chart (see Fig. 9 for a sample outcome).
- *Temporal mapping.* The ordered list of vertices is mapped to calendar days under the assumption of uniform horizontal spacing within each month; leap years are detected automatically.
- *Vertical scaling.* Already during the annotation for the monthly slicing, two anchor points (c_{low} and c_{high}) on the y axis were recorded per chart in pixels and the according gauge level values manually entered in the chart’s unit (foot or mm: v_{low} and v_{high}). Assuming linear scaling, physical gauge levels v_{test} are recovered from pixel y -coordinates c_{test} via

$$v_{\text{test}} = \frac{(c_{\text{test}} - c_{\text{low}})(v_{\text{high}} - v_{\text{low}})}{c_{\text{high}} - c_{\text{low}}} + v_{\text{low}} \quad (8)$$

4.4.3 Level-Based Ground Truth

For 4 of the same 15 charts we independently keyed the gauge level for each day directly from the grid (≈ 70 min per chart). The two ground-truth sets agree to 0.996 on Custom peak-aware score (see below) with a maximum absolute deviation of 2.614 mm, demonstrating that pixel-based annotation is both reliable and an order of magnitude faster.

4.4.4 Cross-Validation and Quality Control

A random subset of transformed curves was visually overlaid on the original scans to detect residual misalignments caused by scanning artefacts (warp, skew, page curvature). None exceeded one grid interval. These checks, together with the high agreement between both annotation modes, give confidence that the ground-truth data are fit for purpose. All remaining distortions are systematic and addressed later by the affine calibration described in Sect. 4.4.5.

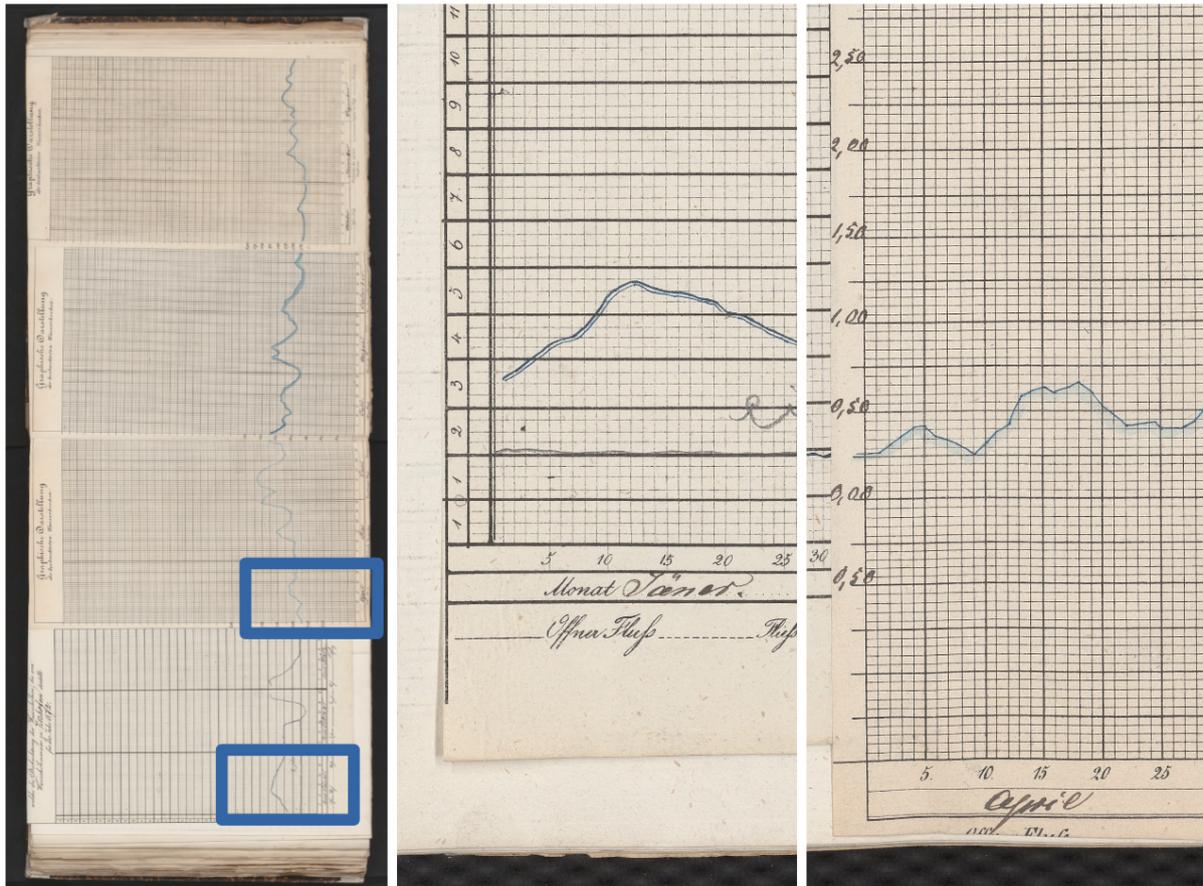


Figure 7. The Vilshofen 1872 chart illustrates the unit change from Bavarian foot to millimeters effective 1 April. Note that at the transition from March to April, the chart aligns. Images source: BayHStA BLW 21.

Table 2. Sample random selection.

| Gauge location | DocID* | Pages | Years |
|-----------------|--------|--------------------|------------------------------|
| Pegel Neu-Ulm | 21 | 18, 45, 51, 56, 85 | 1839, 1865, 1871, 1876, 1891 |
| Pegel Vilshofen | 29 | 22, 24, 45, 53, 69 | 1844, 1846, 1866, 1872, 1883 |
| Pegel Passau | 30 | 26, 46, 49, 60, 70 | 1848, 1865, 1868, 1878, 1888 |

* DocID and page number form the full document identifier as follows:
 Bay_Landesamt_fuer_Wasserwirtschaft_{DocID}{Page:04}, e.g.,
 Bay_Landesamt_fuer_Wasserwirtschaft_210018.

As a third method for creating ground truth data (and also as an alternative benchmark for our semi-automated workflow), re-drawing the chart line on the screen using a stylus or graphic tablet with a post translation of coordinates as above is thinkable. This approach was not tested by us, but we may well assume a slight gain in efficiency without loss of quality.

4.4.5 Curve-to-Series Transformation

Since LineFormer predictions are denser than daily sampling (typically 2000+ points per month), the selected line data is resampled to one point per day. We partition each

month into 28–31 (depending on calendar length) equal slices and retain the last coordinate in each slice, which outperformed median, mean and medoid aggregations in our evaluations. X -coordinates are mapped to ordinal day numbers; y -coordinates are converted to physical gauge levels utilizing the per-tile anchor pair defined during the monthly slicing process using Eq. (8).

4.5 Experimental pilot study

The pilot phase determined workable inputs for transformer-based extraction on historical scans and locked key pre-

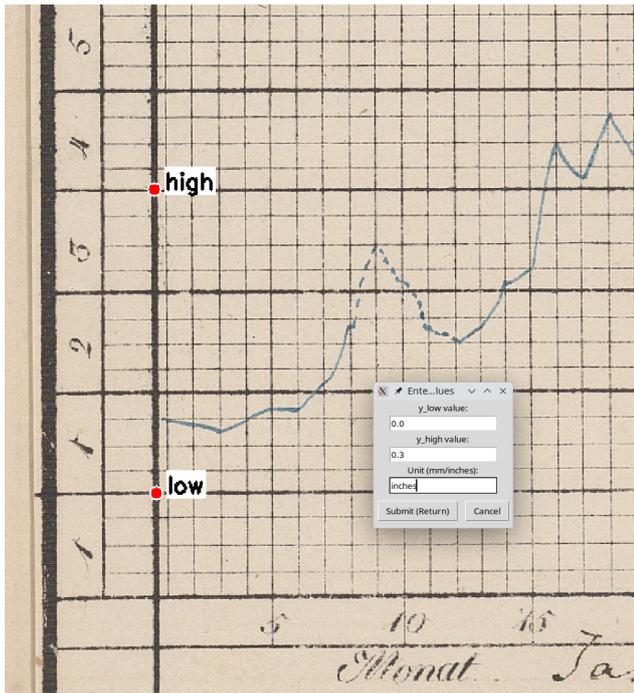


Figure 8. Annotation of y axis anchor points during monthly annotation using the `month_annotator` tool. Image source: BayH-StA BLW 21.

processing choices. This phase was conducted with the sample (ground truth) data described in Sect. 4.4, i.e. 15 gauge years from 3 different gauges with an overall of 180 monthly data.

4.5.1 Pre-processing: slicing, cropping and y axis calibration

Following the promising result on a single test image, we adopted a pre-processing strategy based on monthly horizontal slicing combined with vertical cropping. As previous experiments suggested that horizontal stretching might further enhance prediction quality, we proceeded to systematically evaluate its impact.

We evaluated full-page, semi-annual, quarterly and monthly tiling. Monthly tiles reduced local warp and date-assignment error and were adopted for the pipeline. As automating this step is reserved for future advances of the pipeline, full-page scans were manually annotated with the `month_annotator` tool to identify the chart area as bounding boxes for one month each. Precise placement proved critical: a small x -overshoot can shift the date assignment by a whole day (Fig. 10).

From the annotation, twelve horizontal tiles were automatically produced. The human annotation workload for this averaged 160 s per gauge year which includes the annotation of vertical anchor points required for y axis transformation from pixels to gauge levels (see above, Sect. 4.4.2).

This monthly cropping had another effect. While its primary goal was to slice the yearly full image horizontally into manageable and processable units, it also allowed to reduce the vertical space of the image to process to what was actually needed by the graph. Eliminating an often high percentage of the space, made the task easier for the predictor.

4.5.2 Preprocessing: horizontal stretching

Monthly tiles are visually compressed along the horizontal (time) axis: steep strokes and tight meanders can collapse into only a few pixels per day. This increases local tangling and small gaps in the predicted polyline. A purely horizontal anisotropic stretch increases the effective sampling along time without altering vertical geometry or units. In practice, stretching reduces steep slopes and widens close strokes, making stroke-following easier for the transformer while leaving the y axis calibration (anchors, units) unchanged.

We evaluated different stretch factors $s \in \{0.5, 0.75, 1.0, 1.25, 1.5, 2.0, 2.5, 3.0, 4.0, 5.0\}$ applied to the x axis only using bicubic interpolation (y -coordinates and vertical anchors unchanged). For each s , we re-ran inference on the development set (complete months) under identical post-processing: candidate limit set to 1, same day-slicing (last-in-bin), and the same pixel \rightarrow level conversion. We then computed evaluation scores, especially *peak-aware Custom* against ground truth (Fig. 11).

The quality improved monotonically from $s = 0.5$ to $s = 2.0$, with the largest gains between $s = 1.0 \rightarrow 1.5$ and diminishing returns thereafter. Beyond $s = 2.0$ the curves plateaued: average gains became negligible and several months showed small regressions. We therefore fix $s = 2.0$ for the production pipeline. For rare pages with atypical layout or memory constraints, s may be changed, but all data produced here use $s = 2.0$. Vertical scaling and date binning is preserved, i.e. no change to y -units is performed and no month bounds affected.

4.5.3 Baseline adjustment

To mitigate spine-induced page curvature (relevant especially for the “inner” months of June and July, see Fig. 4 for an example), we apply a light, grid-aware baseline adjustment using the printed zero line we manually annotated as a polyline with the tool `baseline_annotator.py`. The polyline serves as a normalization for the y -value-conversion, i.e. we subtract locally estimated baseline bias before applying the linear y -scale. Aligned with the horizontal image boundary, the distortion is calculated and added to adjust the gauge level while converting from pixel values as follows.

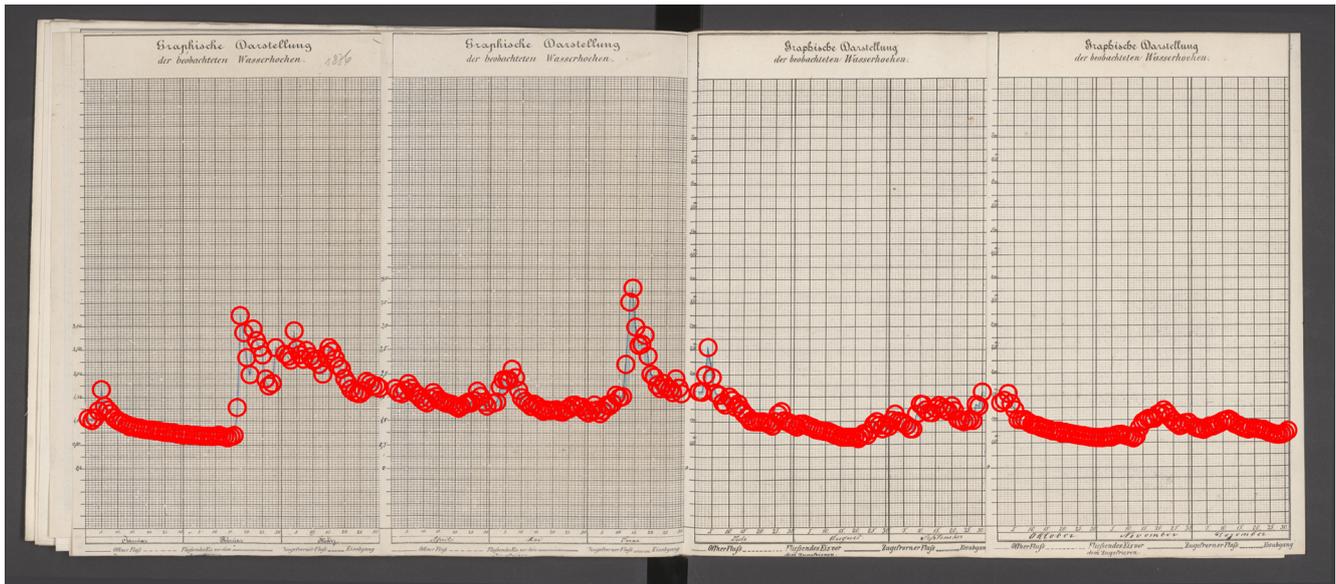


Figure 9. Sample of a pixel-based ground-truth annotation (Neu-Ulm, 1876). Image source: BayHStA BLW 21.

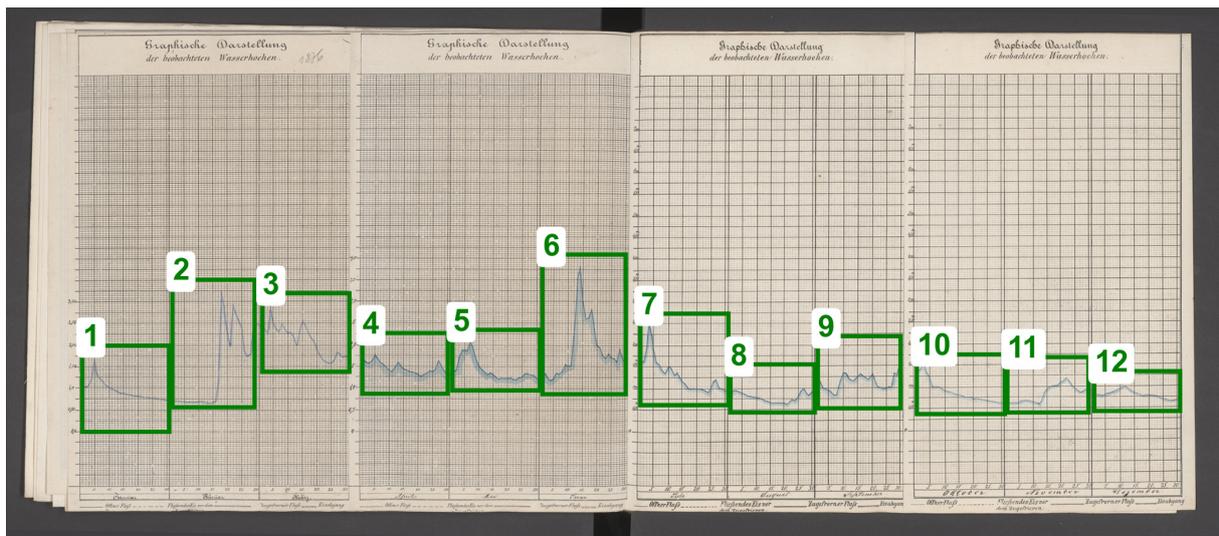


Figure 10. Annotation of monthly segments for slicing (Neu-Ulm, 1876). Image source: BayHStA BLW 21.

Based on the calibration values from Sect. 4.4.2, let β be the linear box correction as

$$\beta = c_{\text{low}} - \frac{v_{\text{low}}}{v_{\text{high}} - v_{\text{low}}} (c_{\text{high}} - c_{\text{low}}) \quad (9)$$

Let $((x_1, y_1), \dots, (x_n, y_n))$ be an (x) -ordered set of pixel pairs to describe the annotated baseline polyline. For each segment with endpoints (x_i, y_i) and (x_{i+1}, y_{i+1}) and each x_q with $x_i < x_q < x_{i+1}$, let

$$\hat{y}(x_q) = y_i + (y_{i+1} - y_i) \frac{x_q - x_i}{x_{i+1} - x_i} \quad (10)$$

Then, the adjusted y -value $y(x_q)$ is calculated as:

$$y(x_q) = \hat{y}(x_q) - \beta \quad (11)$$

Table 3 summarises the recognition accuracy comparing predictions with and without baseline adjustment. We observe a general improvement in mean Custom peak-aware score which is (only) slightly higher in the critical central months. In terms of efficiency, this additional manual pre-processing averages 26.89 s per page (gauge year) for annotating the polyline base. Overall, we assess manual baseline adjustment as valuable and hence leave it in the pipeline.

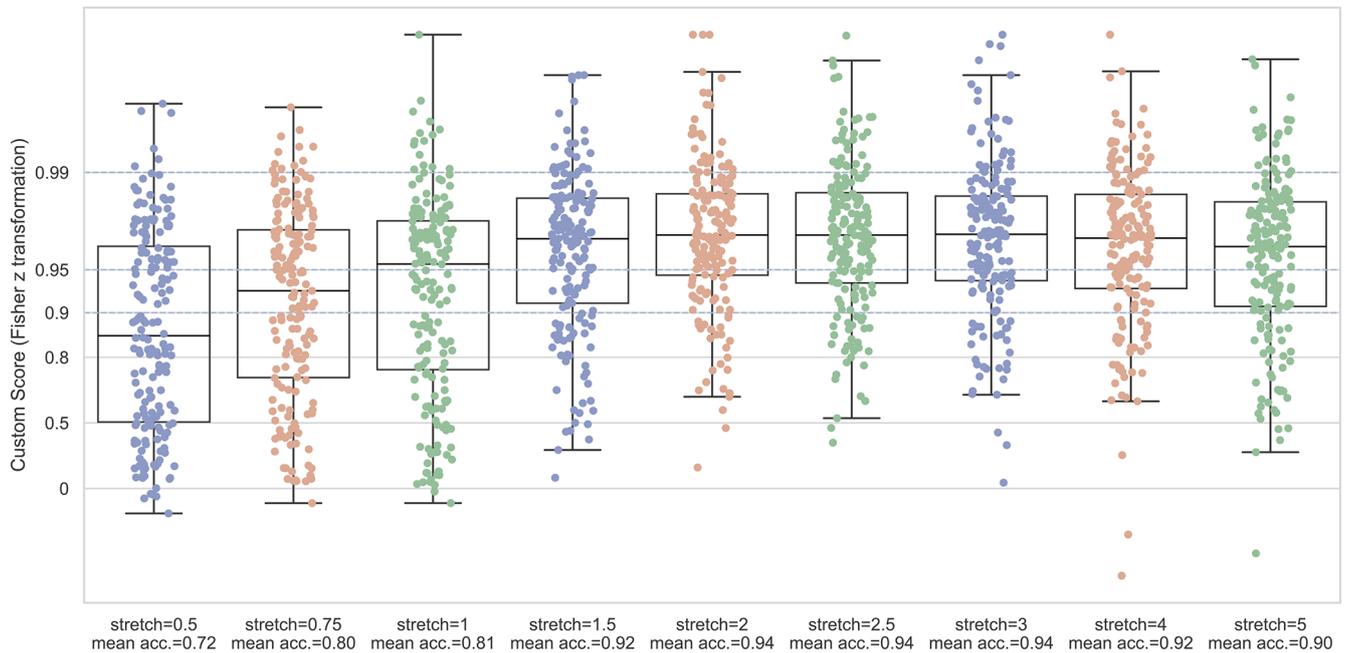


Figure 11. Evaluation of x axis stretching factors against ground truth with peak-aware Custom metric.

Table 3. Accuracy (Custom peak-aware score) of sample predictions with and without baseline adjustment.

| | count | mean | std | min | 25 % | 50 % | 75 % | max |
|------------------------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|
| without baseline adjustment: all | 180 | 0.958 | 0.067 | 0.467 | 0.957 | 0.978 | 0.988 | 1.000 |
| without baseline adjustment: June/July only | 30 | 0.943 | 0.070 | 0.691 | 0.934 | 0.968 | 0.985 | 0.998 |
| without baseline adjustment: all but June/July | 150 | 0.961 | 0.067 | 0.467 | 0.961 | 0.980 | 0.989 | 1.000 |
| with baseline adjustment: all | 180 | 0.968 | 0.055 | 0.474 | 0.967 | 0.980 | 0.990 | 0.999 |
| with baseline adjustment: June/July only | 30 | 0.956 | 0.059 | 0.695 | 0.957 | 0.974 | 0.984 | 0.996 |
| with baseline adjustment: all but June/July | 150 | 0.970 | 0.054 | 0.474 | 0.968 | 0.981 | 0.990 | 0.999 |

4.5.4 LineFormer hyperparameters and candidate selection

Polyline prediction is based on the *LineFormer* implementation by Lal et al. (2023) which in turn builds on the open *MMDetection* framework (Chen et al., 2019). We made only minor adjustments to *LineFormer* to expose and analyse two scores, *confidence* and *coverage*. Here, *coverage* denotes the fraction of the image's x axis for which a data point is predicted. Because images were sliced to span the expected chart range (first to last day of the month), *coverage* should ideally approach 100%. We treat $\text{coverage} > 0.985$, as viable, corresponding to a potential time bias smaller than half a day (i.e. less than one vertical grid interval), consistent with the manual visual inspection described below.

Among *LineFormer*'s hyperparameters, only maxperimage – the maximum number of polylines predicted per image (gauge-month) – had a material impact on outcomes. The default is $\text{maxperimage} = 100$, which in our sample produces between 1 and 12 polylines per image (see Fig. 12 for an example). As our sources indicate that

each image contains exactly one chart, $\text{maxperimage} = 1$ would be ideal, as it obviates candidate selection. However, this setting does not always yield the most accurate or most complete prediction.

We evaluated $\text{maxperimage} \in \{1, 2, 3, 100\}$ and compared accuracies. Whenever *LineFormer* returned more than one prediction for an image (94.4% of cases), a selection step was required. For $\text{maxperimage} > 1$ we used the model's *confidence* score to select a single candidate. Values $\text{maxperimage} > 1$ did not improve accuracy relative to $\text{maxperimage} = 1$, making $\text{maxperimage} = 1$ the only viable option for fully automated processing at this stage (see Fig. 13 and Table 4).

As an alternative to automatic selection by confidence, we conducted manual review with the `inspector.py` tool at $\text{maxperimage} = 100$. A human evaluator inspects all predictions per image, selects the visually best curve, and flags it as requiring a major correction (deviation $>$ one grid tick on either axis), a minor correction (within one tick), or no material correction. This process takes approximately 3 s per

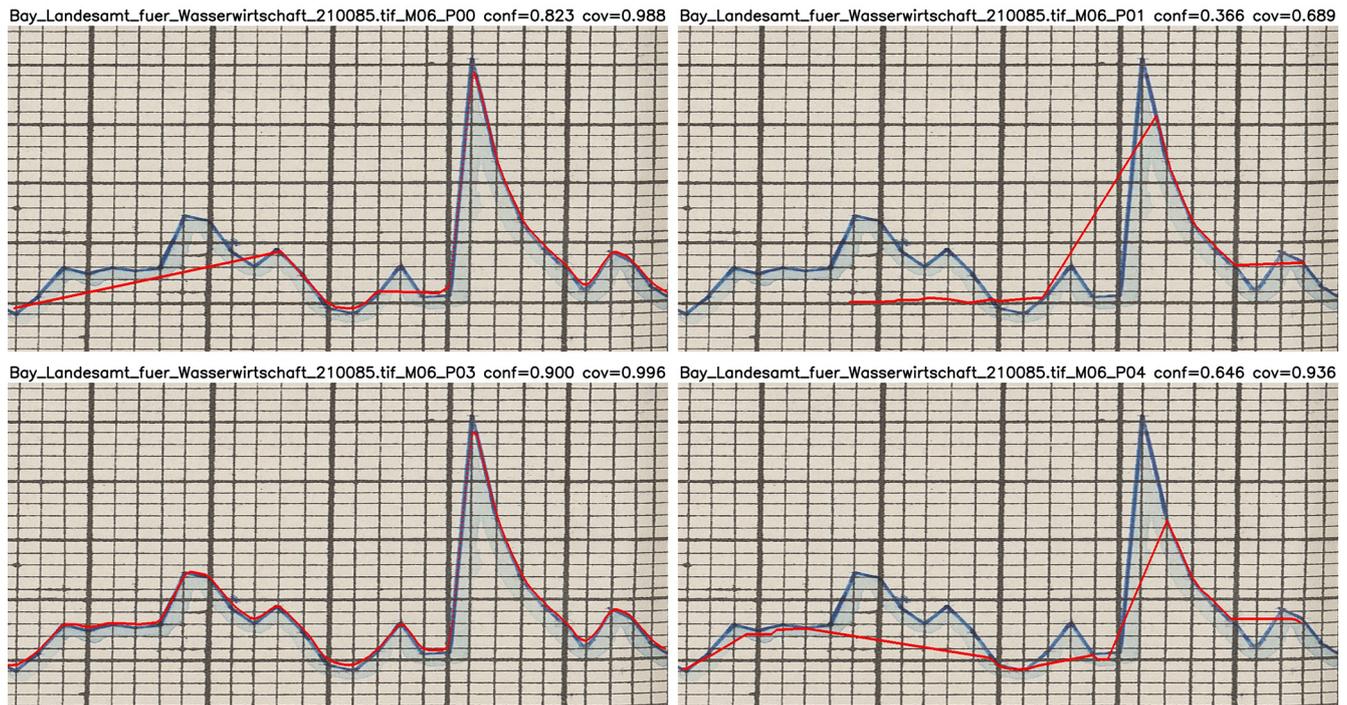


Figure 12. Samples illustrating different candidates returned by *LineFormer* during the same prediction run (Neu-Ulm, June 1891). Image source: BayHStA BLW 21.

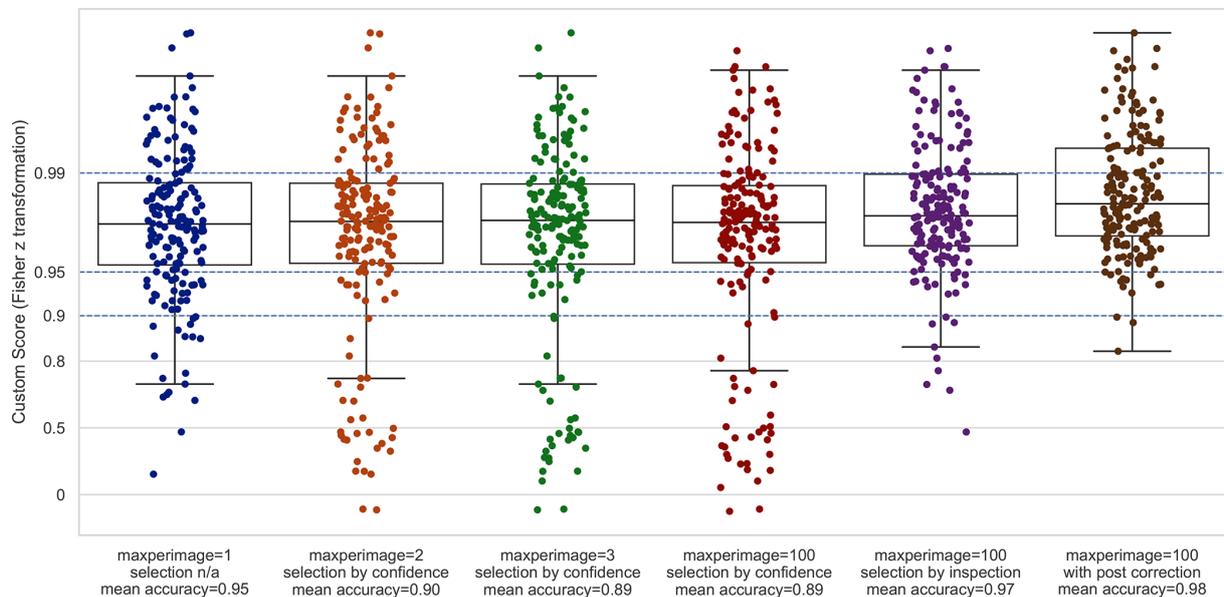


Figure 13. Evaluation of sample predictions using different *maxperimage* settings for *LineFormer* with the peak-aware *Custom* metric.

gauge-month. In our 180-month sample, 88.9 % of months were accepted without correction, 3.9 % were flagged for minor edits, and 7.2 % for major edits. Manual selection outperformed automatic confidence-based selection (overall accuracy 0.968 vs. 0.886).

Within the *maxperimage* = 100 setting, we asked whether manual selection could be approximated by a simple combination of confidence and coverage. Using confidence alone recalled 78.2 % of the visually best predictions. GroupKFold cross-validation identified $\alpha = 0.69$ as the opti-

Table 4. Accuracy (Custom peak-aware score) of sample predictions with different settings of `maxperimage` *LineFormer* hyperparameter, sorted by mean average accuracy. n/a: not applicable.

| <code>maxperimage</code> | selector | count | mean | std | min | 25 % | 50 % | 75 % | max |
|--------------------------|----------------------|-------|-------|-------|--------|-------|-------|-------|-------|
| 100 | by confidence | 180 | 0.884 | 0.233 | −0.137 | 0.957 | 0.978 | 0.988 | 0.999 |
| 1 | n/a | 180 | 0.951 | 0.092 | 0.167 | 0.955 | 0.977 | 0.988 | 0.999 |
| 2 | by confidence | 180 | 0.899 | 0.210 | −0.125 | 0.957 | 0.978 | 0.988 | 0.999 |
| 3 | by confidence | 180 | 0.892 | 0.221 | −0.125 | 0.956 | 0.978 | 0.988 | 0.999 |
| 100 | by visual inspection | 180 | 0.968 | 0.055 | 0.474 | 0.967 | 0.980 | 0.990 | 0.999 |
| 100 | with post correction | 180 | 0.979 | 0.021 | 0.827 | 0.972 | 0.983 | 0.993 | 1.000 |

mal weight in the convex combination

$$s_{\alpha} = (1 - \alpha) \cdot \text{confidence} + \alpha \cdot \text{coverage} \quad (12)$$

for automated candidate selection. This increased recall to 87.7 % and improved overall accuracy from 0.886 to 0.937.

In summary, fully automated selection still underperforms manual choice by a meaningful margin (0.937 vs. 0.968). Manual inspection additionally helps to identify cases that benefit from post-prediction correction. We therefore retain manual visual inspection in the processing pipeline. As a pragmatic alternative, setting `maxperimage` = 1 and adding a lightweight human flag for predictions requiring post-correction appears sound.

4.5.5 Post correction

Post correction was conducted and evaluated in those ca. 11 % cases for which the human inspector decided that minor or major deviations exist. As seen above, it improved the mean accuracy significantly from 0.968 to 0.979. In terms of efficiency, however, this manual correction with tool `corrector.py`, which allows movements of single points on a graphical display, comes at a cost of averaging 60 s per gauge month.

In summary, the experimental pilot study revealed that high-resolution scans of full historical pages are unsuitable for direct use with *LineFormer*. Instead, a tailored pre-processing approach combining monthly cropping with *x* axis stretching and introducing human visual inspection for candidate selection as well as manual post corrections, was required to unlock highly usable predictions from the model. These insights were instrumental in shaping our production workflow.

4.5.6 Evaluation

The evaluation for the complete sample set and the finally chosen setup with `maxperimage` = 100, candidate selection by human visual inspection and optional post correction results were already given (Table 4). Figure 14 gives an impression of what an average prediction of 0.968 accuracy can mean; it shows the gauge month prediction closest to the

mean average according to custom peak-aware score and features a time bias.

Figure 15 shows the worst prediction of the sample with a custom peak-aware accuracy of 0.359 and a negative $r = -0.239$. At the current stage of research, no setting could be developed to improve this prediction significantly; it was hence flagged “M” for major manual correction required.

4.5.7 Systematic error analysis

We examined recurring failure modes across gauges, months, and candidate settings to identify systematic (i.e. non-random) error sources (cf. Fig. 16). Below we summarise the main patterns, in addition to those described in Sect. 3.1:

- *Grid-parallel strokes and steep segments.* When the hydrograph runs almost parallel to the printed grid or coincides with it (especially near vertical segments at floods), local contrast with gridlines decreases and *LineFormer* may fragment or hop tracks or detection confuses ink with print.
- *Extreme inclines.* Near-vertical segment runs parallel to the grid; day bins cut through a very steep slope causing peak assigned to two consecutive days.
- *Shadow hiding data.* Stroke shadowing below the curve plus extreme slope leads to under-prediction or missing the true apex.
- *Shadow pulls prediction below line.* Local paper darkening/ink bleed forms a halo under the stroke; predicted track sits a few pixels below the true curve.
- *Undecidable data source.* The way in which the chart is drawn does not allow a precise dating.
- *Missing or superfluous data.* The drawn chart is either prolonged over the first or last day of month, or days are missing, both leading to inaccuracy and a time bias.

Systematic errors primarily affect extremes (peak timing/height) and month-transition continuity. The adopted

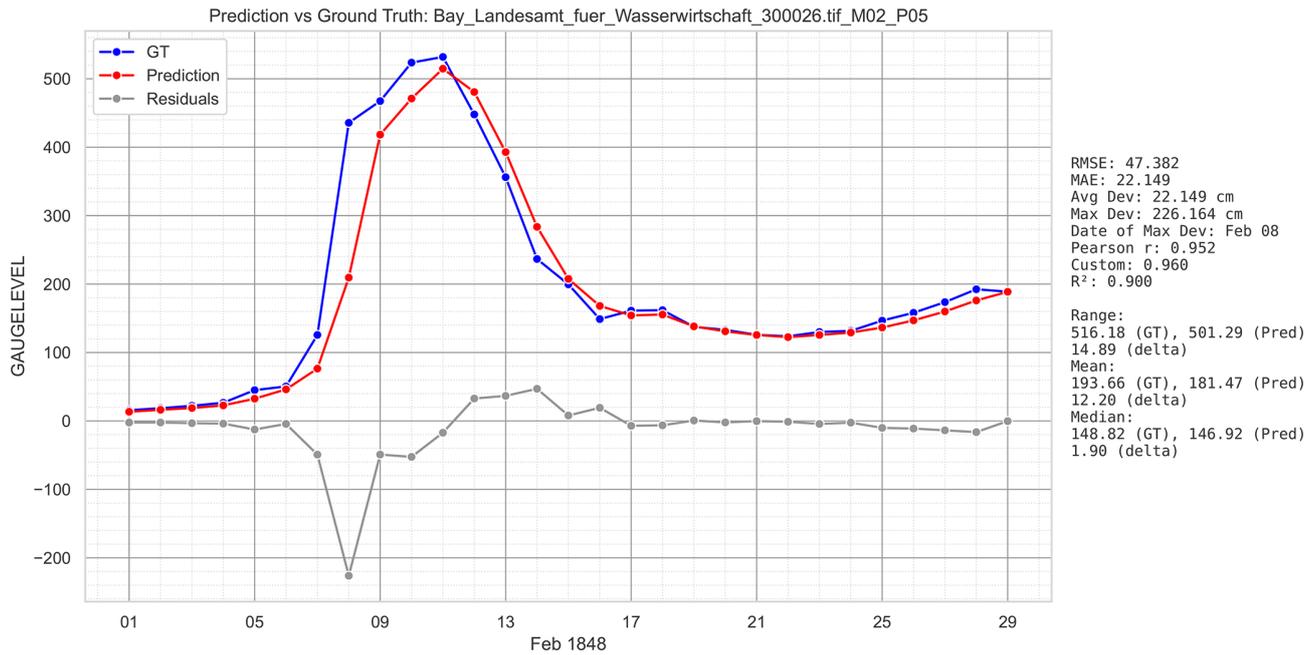


Figure 14. Average prediction (0.960) in the sample set according Custom peak-aware metric (Passau, February 1848). It shows a time bias of one day delay and consequently, due to the steep incline, a maximum deviation of 2.26 m on that day (8 February). Despite the delay, the overall shape and the peak is quite accurate, hence a high Pearson r of 0.952.

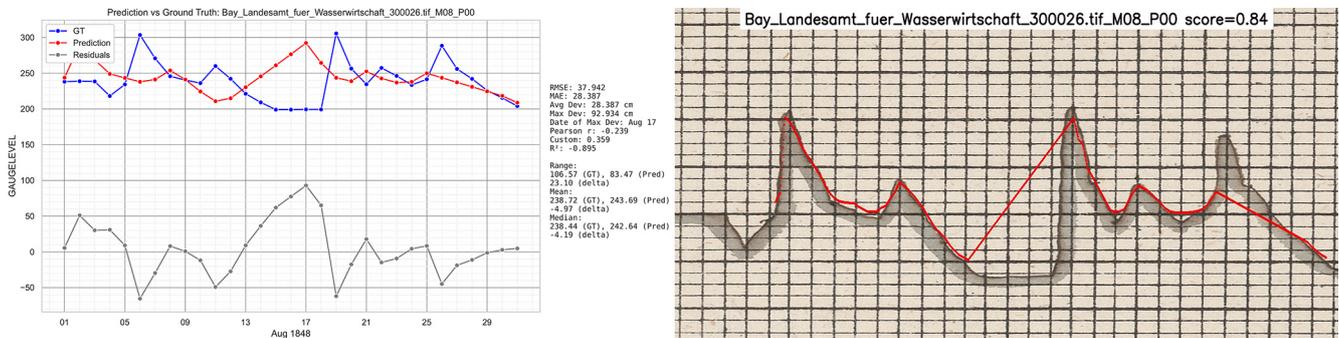


Figure 15. Worst prediction in the sample set according Custom peak-aware metric (0.359, Pearson $r = -0.239$). The right part of the image shows the results directly from the *LineFormer* prediction. While there are parts of the charts predicted nicely, there are three issues: (1) the incompleteness of the predictions, leaving out the first four days; (2) not recognizing correctly the transition between 13 and 18 April and at the end of the months; and (3) the time shift due to the missing data during curve-to-series transformation (Passau, August 1848). Image source: BayHStA BLW 30.

mitigations improve robustness without increasing manual load substantially; remaining edge cases motivate targeted fine-tuning on historical artefacts, automated panel/axis detection with confidence-gated fallbacks, and tighter dewarping-calibration integration in future work.

4.5.8 Benchmarks

Table 5 shows a benchmarking of different approaches evaluated in order to discuss effect (accuracy) and efficiency (time cost). As discussed above (Sect. 4.4.3), the two manual approaches have a strong agreement (0.996). With the manual

pixel-based annotation serving as ground truth, either manual approach makes the gold standard for data rescuing historical gauge level charts in terms of accuracy.

In terms of efficiency, the semi-automated approaches decrease costs in form of manual labour by factors 9.62 to 6.10 respectively. For the production process, we chose to include manual post correction with is an overall trade-off of manual labour saving of 1505 s per gaugeyear and an accuracy drop to 0.977. We also assess that refraining from manual post correction (accuracy drop to 0.964 at 1577 s savings per gauge year) is also a viable way.

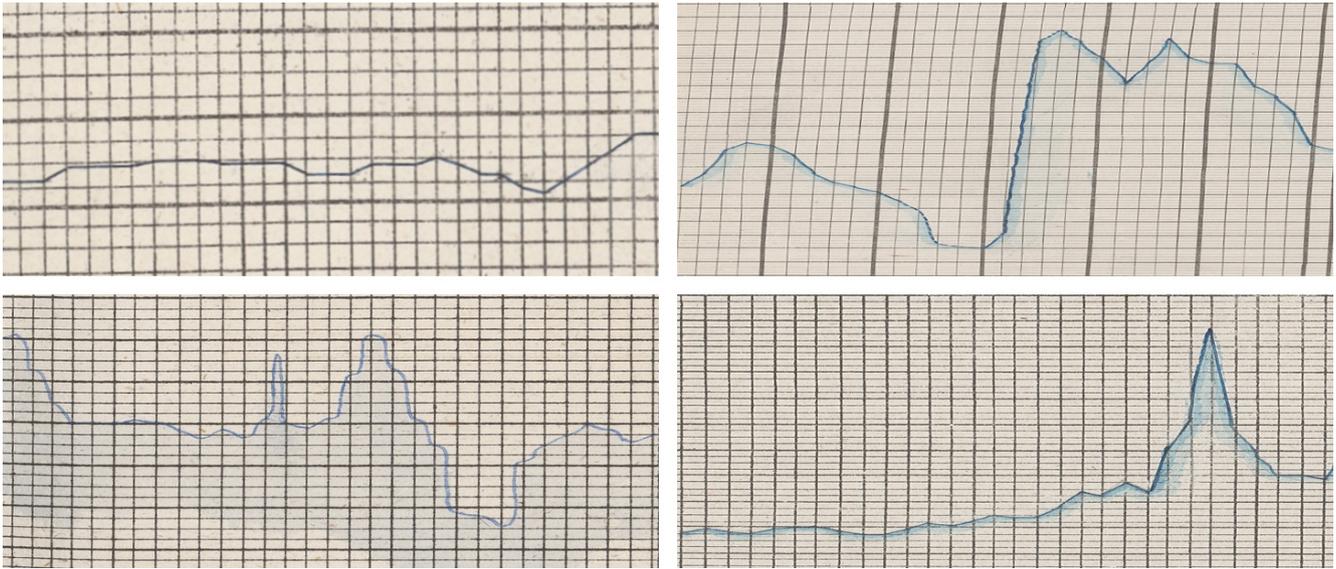


Figure 16. Samples illustrating different systematic sources of errors. Top left: Horizontal overlap of grid and graph (Neu-Ulm, October 1865); top right: near vertical stroke in mid of month (Passau, January 1878); bottom left: undecidable day of peak around 14 July (Passau, July 1845); bottom right: month of April displaying 31 instead of 28 values (Neu-Ulm, April 1871). Image sources: BayHStA BLW 21, 30.

Table 5. Efficiency/effectiveness comparison of different evaluated approaches. Compare with Fig. 17.

| approach | N | mean accuracy | mean time | time cost ratio |
|------------------------------------|-----|---------------|-----------|-----------------|
| manual keying | 48 | 0.996 | 8400 | 4.67 |
| manual pixel-based annotation (GT) | 180 | 1.000 | 1800 | 1.00 |
| by confidence | 180 | 0.884 | 187 | 0.104 |
| by visual inspection | 180 | 0.968 | 223 | 0.124 |
| with post correction | 180 | 0.979 | 295 | 0.164 |

Manual pixel-based annotation serves as benchmark. Efficiency measured in seconds per gauge year; time cost ratio = (seconds per gauge-year)/1800 (GT baseline); effectiveness (accuracy) in Custom peak-aware score. Manual processes include double-keying but no post-correction. Other estimates as described above: baseline annotation: 27 s gaugeyear⁻¹, visual inspection: 36 s gaugeyear⁻¹, post correction: 60 s gaugemonth⁻¹ with 10 % required.

4.6 Validation

For validation, we created another test set by randomly selecting five gauge years that were not part of the tuning sample: Neu-Ulm 1833 (ID 210011), Vilshofen 1849 (ID 290027), Passau 1845 (ID 300023), 1861 (ID 300042), and 1871 (ID 300052). Running the complete pipeline with the settings as described above but without manual post correction, we achieved a mean accuracy of 0.954 on Custom peak-aware score against a manual pixel-based annotation. This is within the range of the sample set which yielded 0.870 and confirms the set-up.

4.7 Final workflow

Taking these consideration and pre-experiments into account, we introduce in the following our pipeline *Historical Water Level Reconstruction (HWLR)* as summarised in Fig. 17.

For each gauge year the pipeline processes (a) a high-resolution annual chart scan; (b) station metadata (identifier, year, unit system), including any known unit transitions; (c) twelve monthly bounding boxes (annotated or detected); (d) two vertical anchors per month (e.g. lowest and highest labelled grid marks with pixel and gauge level values); and (e) a horizontal baseline to detect possible image wrapping.

For each gauge month the pipeline produces: a set of (x, y) -pixel coordinates with confidence score as the presumably best prediction of the actual drawn line chart (usually around 2000 per gauge month) and a daily time series in physical units (millimetres) matched from interpolated (x, y) -pixel coordinates and based on the individual gauge zero mark. Pixel coordinates are measured on the full-resolution scan with origin at the top-left; y increases downward using YOLO-style percentages of the total image size.

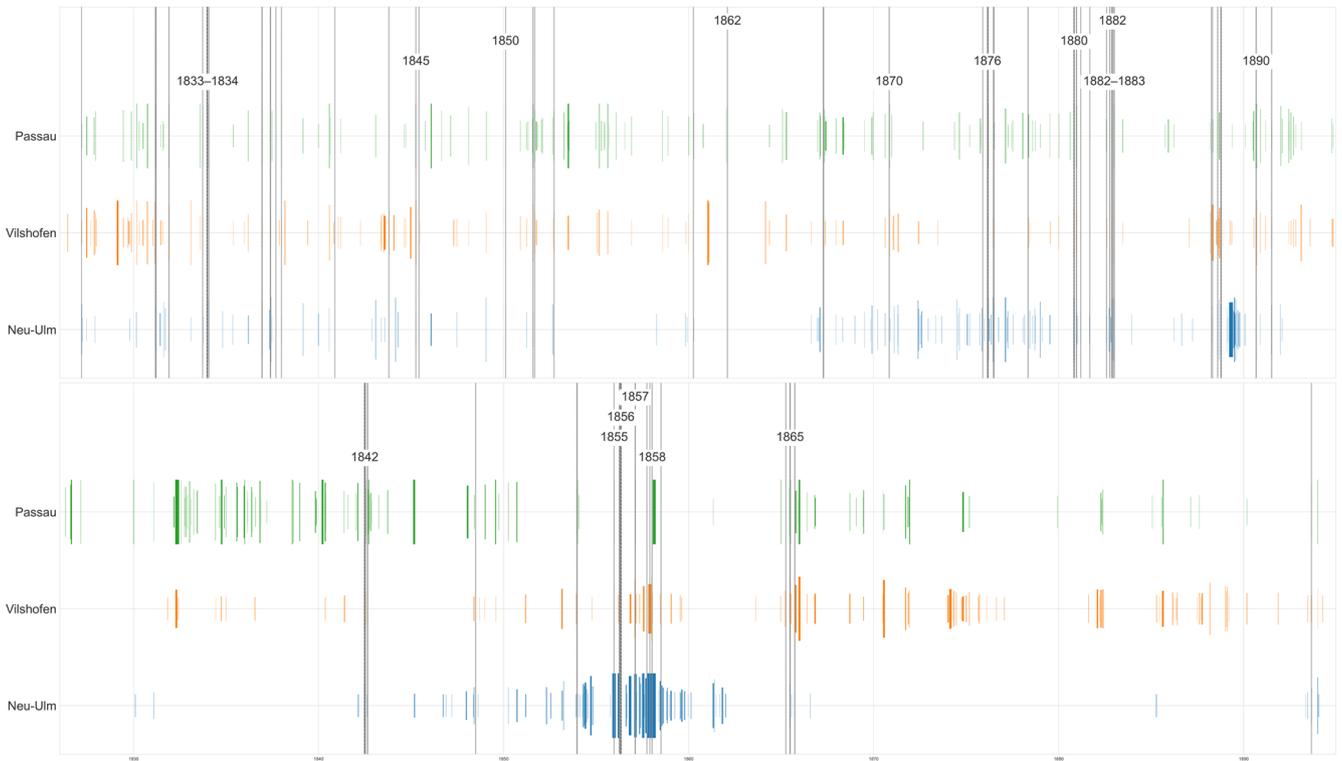


Figure 18. Flood (top) and arid periods along the Danube gauges Neu-Ulm (DE_BY_DAN_21), Vilshofen (DE_BY_DAN_29), and Passau (DE_BY_DAN_30) for the years 1826–1894.

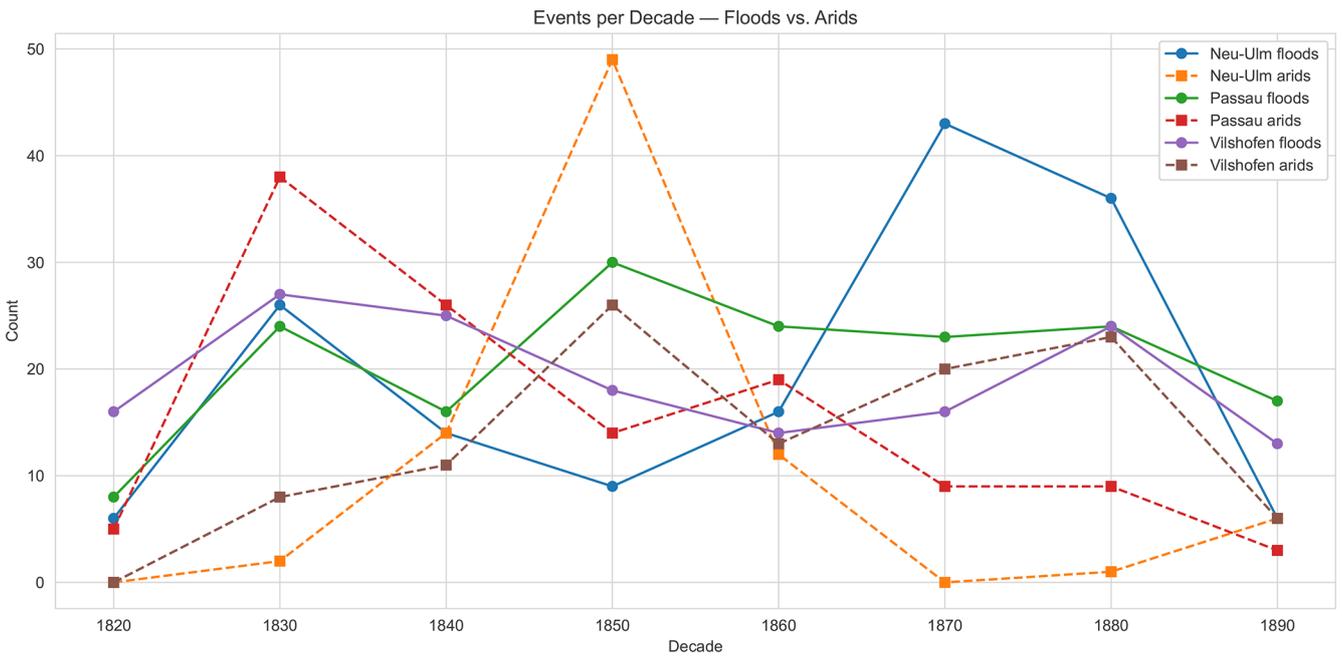


Figure 19. Count of flood and arid events along the Danube gauges Neu-Ulm (DE_BY_DAN_21), Vilshofen (DE_BY_DAN_29), and Passau (DE_BY_DAN_30) for the years 1826–1894.

(water-level) series. Thresholds are computed *per calendar month* to respect intra-annual variability.

We define flood days as those on which daily stage exceeds the monthly 95th non-exceedance percentile, with a minimum duration of at least three consecutive days to avoid one-day spikes. This follows ecological-hydrology practice that emphasises frequency, duration, timing, and rates of change of high stages.

Arid (low-water) periods are defined analogously using the monthly 5th non-exceedance percentile of daily stage, again requiring at least three consecutive days. This mirrors standard low-flow statistics (often discussed via “Q95” low-flow metrics in European practice) while preserving ecological meaning by accounting for seasonal minima.

Without further interpretation, Figs. 18 and 19 report flood and arid periods for the Danube gauges Neu-Ulm, Vilshofen, and Passau from 1826 to 1894.

5 Code and data availability

The data for this particular study has been published in an open online repository under the Creative Commons CC BY 4.0 International licence <https://doi.org/10.5281/zenodo.17296750> (Rehbein, 2025). LineFormer is available at <https://github.com/TheJaeLal/LineFormer> (Lal et al., 2023).

6 Conclusions

This study demonstrates that nineteenth-century, hand-drawn river gauge charts can be systematically transformed into machine-readable daily water-level series with transparent uncertainties and page-level provenance. We present a pragmatic, semi-automated pipeline (HWLR) that combines light, grid-aware pre-processing, transformer-based line extraction, and targeted human oversight. On our Bavarian Danube case study (1826–1894; three gauges), the workflow achieves high accuracy at series level while reducing manual effort by an order of magnitude relative to fully manual digitisation. The resulting datasets, code pointers, and processing artefacts are openly released to support audit, reuse, and extension.

Beyond the immediate case, the approach provides a replicable methodological and practical template for mobilising large analogue hydrometric holdings. By tying reconstruction explicitly to gauge datums and documenting unit transitions, the series produced here can underpin long-term hydrological analyses, before/after assessments of nineteenth-century river works, and multi-decadal context for extremes. The open release also invites community validation and iterative improvement.

While effective on heterogeneous archival material, several limitations remain. First, specific drawing styles and shadowed strokes occasionally degrade line following, in-

dicating value in fine-tuning or training model variants on curated historical samples. Secondly, two manual steps – monthly panel detection and y axis anchor extraction – are currently the main bottlenecks; robust automation of both would further reduce effort and variance. Thirdly, automatic candidate selection still underperforms rapid visual inspection in edge cases; incorporating coverage-aware scoring, better uncertainty estimates, and modest domain adaptation should narrow this gap. Finally, mild residual page warp can bias day assignment where grid geometry departs from ideal; tighter integration of document dewarping with coordinate recovery is a promising avenue.

Scaling to the full archival series is a priority; the present release should be seen as a foundation rather than an endpoint. Reconstruction at scale requires not only techniques but also sustained curation. We align this work with ongoing international data-rescue efforts and propose maintaining the Bavarian Danube series as a “living data” resource with versioned updates, expanded station coverage, and clear provenance metadata. Examples like (Kendon et al., 2024) make a strong case for an effort to make all these data interconnected. The current dataset is openly available, providing an immediate basis for hydrological, geomorphological, and ecological studies. 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.

Competing interests. The author has declared that there are no competing interests.

Disclaimer. Publisher’s note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. The authors bear the ultimate responsibility for providing appropriate place names. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Acknowledgements. This work benefited greatly from the support and access provided by the Generaldirektion der Staatlichen Archive Bayerns and the Bayerisches Hauptstaatsarchiv München. I am grateful to the Computational Historical Ecology research group at the University of Passau for many valuable discussions. The technical workflow builds on *LineFormer* and its foundations; I thank the developers and maintainers for their outstanding contributions. ChatGPT 5 was used as a coding assistant. DeepL and ChatGPT5 supported translations and English stylistics. I am particularly grate-

ful to the two expert reviewers for their knowledgeable, constructive, and very helpful comments.

Financial support. The article processing charges for this open-access publication were covered by the Max Planck Society.

Review statement. This paper was edited by James Thornton and reviewed by two anonymous referees.

References

- Aghajani, M., Ahmadi, H., Mirzaei, N., and Batlo, J.: Digitizing Analog Seismograms of Iranian Seismic Stations – Preserving Records for Strong Earthquakes in Iran 1960–1990, *J. Seismol.*, 27, 693–706, <https://doi.org/10.1007/s10950-023-10158-4>, 2023.
- Antico, A., Aguiar, R. O., and Amsler, M. L.: Hydrometric Data Rescue in the Paraná River Basin, *Water Resour. Res.*, 54, 1368–1381, <https://doi.org/10.1002/2017WR020897>, 2018.
- Antico, A., Mendizabal, S., Ferreira, L. J., Aguiar, R. O., and Amsler, M. L.: Addendum to “Hydrometric Data Rescue in the Paraná River Basin” by Andrés Antico, Ricardo O. Aguiar, and Mario L. Amsler, *Water Resour. Res.*, 56, e2019WR026654, <https://doi.org/10.1029/2019WR026654>, 2020.
- Bayerisches Landesamt für Umwelt: Grundlagen Zu Flussaufnahmen Und Deren Dokumentation, http://www.bestellen.bayern.de/shoplink/lfu_was_00182.htm (last access: 18 August 2025), 2020.
- Becker, A. and Grünewald, U.: Flood Risk in Central Europe, *Science*, 300, 1099–1099, <https://doi.org/10.1126/science.1083624>, 2003.
- Berking, J., Oleson, J. P., Sürmelihindi, G., Nenci, E., Maganzani, L., Ronin, M., Schrakamp, I., Gerrard, C., Gutiérrez, A., Trümper, M., Bouffier, S., Dumas, V., Lenhardt, P., Paillet, J.-L., and Schomberg, A.: Water Management in Ancient Civilizations, Humboldt-Universität zu Berlin, <https://doi.org/10.18452/19758>, 2019.
- Biswas, A. K.: History of Hydrology, North-Holland Publ., Amsterdam, ISBN 978-0-444-10025-2, 1970.
- Bogiatzis, P. and Ishii, M.: DigitSeis: A New Digitization Software for Analog Seismograms, *Seismol. Res. Lett.*, 87, 726–736, <https://doi.org/10.1785/0220150246>, 2016.
- Bradshaw, E., Rickards, L., and Aarup, T.: Sea Level Data Archaeology and the Global Sea Level Observing System (GLOSS), *GeoResJ*, 6, 9–16, <https://doi.org/10.1016/j.grj.2015.02.005>, 2015.
- Bradshaw, E., Woodworth, P., Hibbert, A., Bradley, L., Pugh, D., Fane, C., and Bingley, R.: A Century of Sea Level Measurements at Newlyn, Southwest England, *Mar. Geod.*, 39, 115–140, <https://doi.org/10.1080/01490419.2015.1121175>, 2016.
- Brázdil, R., Kundzewicz, Z. W., and Benito, G.: Historical Hydrology for Studying Flood Risk in Europe, *Hydrolog. Sci. J.*, 51, 739–764, <https://doi.org/10.1623/hysj.51.5.739>, 2006.
- Brohan, P., Allan, R., Freeman, J. E., Waple, A. M., Wheeler, D., Wilkinson, C., and Woodruff, S.: Marine Observations of Old Weather, *B. Am. Meteorol. Soc.*, 90, 219–230, <https://doi.org/10.1175/2008BAMS2522.1>, 2009.
- Brönnimann, S., Brugnara, Y., Allan, R. J., Brunet, M., Compo, G. P., Crouthamel, R. I., Jones, P. D., Jourdain, S., Luterbacher, J., Siegmund, P., Valente, M. A., and Wilkinson, C. W.: A Roadmap to Climate Data Rescue Services, *Geosci. Data J.*, 5, 28–39, <https://doi.org/10.1002/gdj3.56>, 2018.
- Burnhill, T. and Adamson, P.: Floods and the Mekong River System (Part 3), Catch and Culture, Fisheries Research and Development in the Mekong Region, 14, 25–30, 2008.
- Burr, G., Podolsky, L., and Lapointe, Y. A.: Redrawing historical weather data and participatory archives for the future, *Comma*, 297–315, <https://doi.org/10.3828/coma.2021.28>, 2021.
- Caldwell, P. C., Merrifield, M. A., and Thompson, P. R.: Sea Level Measured by Tide Gauges from Global Oceans as Part of the Joint Archive for Sea Level (JASL) since 1846, NOAA [data set], <https://doi.org/10.7289/V5V40S7W>, 2001.
- Capozzi, V., Cotroneo, Y., Castagno, P., De Vivo, C., and Budillon, G.: Rescue and quality control of sub-daily meteorological data collected at Montevergine Observatory (Southern Apennines), 1884–1963, *Earth Syst. Sci. Data*, 12, 1467–1487, <https://doi.org/10.5194/essd-12-1467-2020>, 2020.
- Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., and Zagoruyko, S.: End-to-End Object Detection with Transformers, in: *Computer Vision – ECCV 2020*, edited by: Vedaldi, A., Bischof, H., Brox, T., and Frahm, J.-M., vol. 12346, Springer International Publishing, Cham, 213–229, ISBN 978-3-030-58451-1, https://doi.org/10.1007/978-3-030-58452-8_13, 2020.
- Castelvecchi, D.: “Revolutionary” AI Tools Rescue Old Weather Data to Improve Climate Models, *Nature*, <https://doi.org/10.1038/d41586-025-02798-y>, 2025.
- Chen, K., Wang, J., Pang, J., Cao, Y., Xiong, Y., Li, X., Sun, S., Feng, W., Liu, Z., Xu, J., Zhang, Z., Cheng, D., Zhu, C., Cheng, T., Zhao, Q., Li, B., Lu, X., Zhu, R., Wu, Y., Dai, J., Wang, J., Shi, J., Ouyang, W., Loy, C. C., and Lin, D.: MMDetection: Open MMLab Detection Toolbox and Benchmark, *arXiv [preprint]*, <https://doi.org/10.48550/arXiv.1906.07155>, 2019.
- Cheng, Z.-Q., Dai, Q., Li, S., Sun, J., Mitamura, T., and Hauptmann, A. G.: ChartReader: A Unified Framework for Chart Derendering and Comprehension without Heuristic Rules, *arXiv [preprint]*, <https://doi.org/10.48550/arXiv.2304.02173>, 2023.
- Corona-Fernandez, R. D. and Santoyo, M. A.: TIITBA – a Graphical Interface for Historical Seismograms, Vectorization, Analysis and, Correction, Zenodo [code], <https://doi.org/10.5281/zenodo.6272822>, 2023.
- Dangendorf, S., Wahl, T., Mudersbach, C., and Jensen, J.: The Seasonal Mean Sea Level Cycle in the Southeastern North Sea, *J. Coastal Res.*, 165, 1915–1920, <https://doi.org/10.2112/SI65-324.1>, 2013.
- Dangendorf, S., Calafat, F. M., Arns, A., Wahl, T., Haigh, I. D., and Jensen, J.: Mean Sea Level Variability in the North Sea: Processes and Implications, *J. Geophys. Res.-Oceans*, 119, 2014JC009901, <https://doi.org/10.1002/2014JC009901>, 2014.
- Das, S., Ma, K., Shu, Z., Samaras, D., and Shilkrot, R.: DeWarpNet: Single-Image Document Unwarping With Stacked 3D and 2D Regression Networks, in: *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, Seoul, Korea (South), 131–140, ISBN 978-1-7281-4803-8, <https://doi.org/10.1109/ICCV.2019.00022>, 2019.

- Davila, K., Kota, B. U., Setlur, S., Govindaraju, V., Tensmeyer, C., Shekhar, S., and Chaudhry, R.: ICDAR 2019 Competition on Harvesting Raw Tables from Infographics (CHART-Infographics), in: 2019 International Conference on Document Analysis and Recognition (ICDAR), IEEE, Sydney, Australia, 1594–1599, ISBN 978-1-7281-3014-9, <https://doi.org/10.1109/ICDAR.2019.00203>, 2019.
- De Smeth, K., Comer, J., and Murphy, C.: Hydrometric Data Rescue and Extension of River Flow Records: Method Development and Application to Catchments Modified by Arterial Drainage, *Geosci. Data J.*, 11, 176–196, <https://doi.org/10.1002/gdj3.206>, 2024.
- Donig, S. and Rehbein, M.: Für Eine “Gemeinsame Digitale Zukunft”. Eine Kritische Verortung Der Digital History, *Geschichte in Wissenschaft und Unterricht*, 72, 527–545, 2022.
- Drevon, D., Fursa, S. R., and Malcolm, A. L.: Inter-coder Reliability and Validity of WebPlotDigitizer in Extracting Graphed Data, *Behav. Modif.*, 41, 323–339, <https://doi.org/10.1177/0145445516673998>, 2017.
- Faber, E.: Studien Über Die Verbesserung Der Schiffbarkeit Der Donau von Kelheim Bis Nach Ulm, A. Troschel, Berlin-Grünwald, 1903.
- Featherstone, W. E., Penna, N. T., Filmer, M. S., and Williams, S. D. P.: Nonlinear Subsidence at Fremantle, a Long-recording Tide Gauge in the Southern Hemisphere, *J. Geophys. Res.-Oceans*, 120, 7004–7014, <https://doi.org/10.1002/2015JC011295>, 2015.
- Federal Agencies Digital Guidelines Initiative: Technical Guidelines for Digitizing Cultural Heritage Materials, 3rd edn., https://www.digitizationguidelines.gov/guidelines/FADGITechnicalGuidelinesforDigitizingCulturalHeritageMaterials_ThirdEdition_05092023.pdf (last access: 5 March 2025), 2023.
- Fields, C. and Kennington, C.: Vision Language Transformers: A Survey, *arXiv [preprint]*, <https://doi.org/10.48550/arXiv.2307.03254>, 2023.
- Gallant, A. J. E. and Gergis, J.: An Experimental Streamflow Reconstruction for the River Murray, Australia, 1783–1988, *Water Resour. Res.*, 47, 2010WR009832, <https://doi.org/10.1029/2010WR009832>, 2011.
- Hannah, J. and Bell, R. G.: Regional Sea Level Trends in New Zealand, *J. Geophys. Res.-Oceans*, 117, 2011JC007591, <https://doi.org/10.1029/2011JC007591>, 2012.
- Hawkins, E.: Data Rescue with AI: Google Gemini Has Changed the Game, *Substack*, <https://climatelabbook.substack.com/p/data-rescue-with-ai> (last access: 5 March 2025), 2025.
- Hawkins, E., Burt, S., Brohan, P., Lockwood, M., Richardson, H., Roy, M., and Thomas, S.: Hourly Weather Observations from the Scottish Highlands (1883–1904) Rescued by Volunteer Citizen Scientists, *Geosci. Data J.*, 6, 160–173, <https://doi.org/10.1002/gdj3.79>, 2019.
- Hawkins, E., Brohan, P., Burgess, S. N., Burt, S., Compo, G. P., Gray, S. L., Haigh, I. D., Hersbach, H., Kuyper, K., Martínez-Alvarado, O., McColl, C., Schurer, A. P., Slivinski, L., and Williams, J.: Rescuing historical weather observations improves quantification of severe windstorm risks, *Nat. Hazards Earth Syst. Sci.*, 23, 1465–1482, <https://doi.org/10.5194/nhess-23-1465-2023>, 2023a.
- Hawkins, E., Burt, S., McCarthy, M., Murphy, C., Ross, C., Baldock, M., Brazier, J., Hersee, G., Huntley, J., Meats, R., O’Grady, J., Scrimgeour, I., and Silk, T.: Millions of Historical Monthly Rainfall Observations Taken in the UK and Ireland Rescued by Citizen Scientists, *Geosci. Data J.*, 10, 246–261, <https://doi.org/10.1002/gdj3.157>, 2023b.
- Helaire, L. T., Talke, S. A., Jay, D. A., and Mahedy, D.: Historical Changes in Lower Columbia River and Estuary Floods: A Numerical Study, *J. Geophys. Res.-Oceans*, 124, 7926–7946, <https://doi.org/10.1029/2019JC015055>, 2019.
- Hohensinner, S., Lager, B., Sonnlechner, C., Haidvogel, G., Gierlinger, S., Schmid, M., Krausmann, F., and Winiwarter, V.: Changes in Water and Land: The Reconstructed Viennese Riverscape from 1500 to the Present, *Water History*, 5, 145–172, <https://doi.org/10.1007/s12685-013-0074-2>, 2013a.
- Hohensinner, S., Sonnlechner, C., Schmid, M., and Winiwarter, V.: Two Steps Back, One Step Forward: Reconstructing the Dynamic Danube Riverscape under Human Influence in Vienna, *Water History*, 5, 121–143, <https://doi.org/10.1007/s12685-013-0076-0>, 2013b.
- Holgate, S., Matthews, A., Woodworth, P., Richards, L., Tamisiea, M., Bradshaw, E., Foden, P., Gordon, K., Jevrejeva, S., and Pugh, J.: New Data Systems and Products at the Permanent Service for Mean Sea Level, *J. Coastal Res.*, 29, 493, <https://doi.org/10.2112/JCOASTRES-D-12-00175.1>, 2012.
- Huffman, G. J., Adler, R. F., Behrangi, A., Bolvin, D. T., Nelkin, E. J., Gu, G., and Ehsani, M. R.: The New Version 3.2 Global Precipitation Climatology Project (GPCP) Monthly and Daily Precipitation Products, *J. Climate*, 36, 7635–7655, <https://doi.org/10.1175/JCLI-D-23-0123.1>, 2023.
- Hunter, J., Coleman, R., and Pugh, D.: The Sea Level at Port Arthur, Tasmania, from 1841 to the Present, *Geophys. Res. Lett.*, 30, 2002GL016813, <https://doi.org/10.1029/2002GL016813>, 2003.
- Inayatillah, A., Haigh, I. D., Brand, J. H., Francis, K., Mortley, A., Durrant, M., Fantuzzi, L., Palmer, E., Miller, C., and Hogarth, P.: Digitising Historical Sea Level Records in the Thames Estuary, UK, *Scientific Data*, 9, 167, <https://doi.org/10.1038/s41597-022-01223-7>, 2022.
- Jaklič, A., Šajn, L., Derganc, G., and Peer, P.: Automatic Digitization of Pluviograph Strip Charts, *Meteorol. Appl.*, 23, 57–64, <https://doi.org/10.1002/met.1522>, 2016.
- Jung, D., Kim, W., Song, H., Hwang, J.-I., Lee, B., Kim, B., and Seo, J.: ChartSense: Interactive Data Extraction from Chart Images, in: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, ACM, Denver Colorado USA, 6706–6717, ISBN 978-1-4503-4655-9, <https://doi.org/10.1145/3025453.3025957>, 2017.
- Kass, M., Witkin, A., and Terzopoulos, D.: Snakes: Active Contour Models, *Int. J. Comput. Vision*, 1, 321–331, <https://doi.org/10.1007/BF00133570>, 1988.
- Kelln, J., Dangendorf, S., Jensen, J., Patzke, J., and Fröhle, P.: Monthly Sea Level from Tide Gauge Stations at the German Baltic Coastline (AMSeL_Baltic Sea), PANGAEA [data set], <https://doi.org/10.1594/PANGAEA.904737>, 2019.
- Kendon, M., Doherty, A., Hollis, D., Carlisle, E., Packman, S., McCarthy, M., Jevrejeva, S., Matthews, A., Williams, J., Garforth, J., and Sparks, T.: State of the UK Climate 2023, *Int. J. Climatol.*, 44, 1–117, <https://doi.org/10.1002/joc.8553>, 2024.
- Khan, M. J. U., Van Den Beld, I., Wöppelmann, G., Testut, L., Latapy, A., and Pouvreau, N.: Extension of a high temporal resolution sea level time series at Socoa (Saint-Jean-de-Luz,

- France) back to 1875, *Earth Syst. Sci. Data*, 15, 5739–5753, <https://doi.org/10.5194/essd-15-5739-2023>, 2023.
- Kim, G., Hong, T., Yim, M., Nam, J., Park, J., Yim, J., Hwang, W., Yun, S., Han, D., and Park, S.: OCR-Free Document Understanding Transformer, in: *Computer Vision – ECCV 2022*, edited by: Avidan, S., Brostow, G., Cissé, M., Farinella, G. M., and Hassner, T., vol. 13688, Springer Nature Switzerland, Cham, 498–517, ISBN 978-3-031-19814-4 978-3-031-19815-1, https://doi.org/10.1007/978-3-031-19815-1_29, 2022.
- Kiss, A. and Laszlovszky, J.: 14th-16th-Century Danube Floods and Long-Term Water-Level Changes in Archaeological and Sedimentary Evidence in The Western and Central Carpathian Basin: An Overview with Documentary Comparison, *Journal of Environmental Geography*, 6, 1–11, <https://doi.org/10.2478/jengeo-2013-0001>, 2013.
- Koutsoyiannis, D., Zarkadoulas, N., Angelakis, A. N., and Tchobanoglous, G.: Urban Water Management in Ancient Greece: Legacies and Lessons, *J. Water Res. Pl.*, 134, 45–54, [https://doi.org/10.1061/\(ASCE\)0733-9496\(2008\)134:1\(45\)](https://doi.org/10.1061/(ASCE)0733-9496(2008)134:1(45)), 2008.
- Lakkis, S. G., Canziani, P. O., Rodriguez, J. O., and Yuchechen, A. E.: Early meteorological records from Corrientes and Bahía Blanca, Argentina: Initial ACRE-Argentina data rescue and related activities, *Geosci. Data J.*, 10, 328–346, <https://doi.org/10.1002/gdj3.176>, 2023.
- Lal, J., Mitkari, A., Bhosale, M., and Doermann, D.: LineFormer: Line Chart Data Extraction Using Instance Segmentation, in: *Document Analysis and Recognition – ICDAR 2023*, edited by: Fink, G. A., Jain, R., Kise, K., and Zanibbi, R., vol. 14191, Springer Nature Switzerland, Cham, 387–400, ISBN 978-3-031-41733-7, https://doi.org/10.1007/978-3-031-41734-4_24, 2023 (code available at: <https://github.com/TheJaeLal/LineFormer>, last access: 18 August 2025).
- Latapy, A., Ferret, Y., Testut, L., Talke, S., Aarup, T., Pons, F., Jan, G., Bradshaw, E., and Pouvreau, N.: Data Rescue Process in the Context of Sea Level Reconstructions: An Overview of the Methodology, Lessons Learned, Up-to-date Best Practices and Recommendations, *Geosci. Data J.*, 10, 396–425, <https://doi.org/10.1002/gdj3.179>, 2023.
- Lee, K., Joshi, M., Turc, I. R., Hu, H., Liu, F., Eisenschlos, J., Khandelwal, U., Shaw, P., Chang, M.-W., and Toutanova, K.: Pix2Struct: Screenshot Parsing as Pretraining for Visual Language Understanding, *Proceedings of Machine Learning Research*, 202, 18893–18912, 2023.
- Liu, F., Eisenschlos, J., Piccinno, F., Krichene, S., Pang, C., Lee, K., Joshi, M., Chen, W., Collier, N., and Altun, Y.: DePlot: One-shot Visual Language Reasoning by Plot-to-Table Translation, in: *Findings of the Association for Computational Linguistics: ACL 2023*, Association for Computational Linguistics, Toronto, Canada, 10381–10399, <https://doi.org/10.18653/v1/2023.findings-acl.660>, 2023.
- Lorrey, A. M., Pearce, P. R., Allan, R., Wilkinson, C., Woolley, J.-M., Judd, E., Mackay, S., Rawhat, S., Slivinski, L., Wilkinson, S., Hawkins, E., Quesnel, P., and Compo, G. P.: Meteorological Data Rescue: Citizen Science Lessons Learned from Southern Weather Discovery, *Patterns*, 3, 100495, <https://doi.org/10.1016/j.patter.2022.100495>, 2022.
- Lucas, M., Lang, M., Renard, B., and Le Coz, J.: A comprehensive uncertainty framework for historical flood frequency analysis: a 500-year-long case study, *Hydrol. Earth Syst. Sci.*, 28, 5031–5047, <https://doi.org/10.5194/hess-28-5031-2024>, 2024.
- Luo, J., Li, Z., Wang, J., and Lin, C.-Y.: ChartOCR: Data Extraction from Charts Images via a Deep Hybrid Framework, in: *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, IEEE, Waikoloa, HI, USA, 1916–1924, ISBN 978-1-6654-0477-8, <https://doi.org/10.1109/WACV48630.2021.00196>, 2021.
- Macdonald, N. and Black, A. R.: Reassessment of Flood Frequency Using Historical Information for the River Ouse at York, UK (1200–2000), *Hydrolog. Sci. J.*, 55, 1152–1162, <https://doi.org/10.1080/02626667.2010.508873>, 2010.
- Masry, A., Long, D., Tan, J. Q., Joty, S., and Hoque, E.: ChartQA: A Benchmark for Question Answering about Charts with Visual and Logical Reasoning, in: *Findings of the Association for Computational Linguistics: ACL 2022*, Association for Computational Linguistics, Dublin, Ireland, 2263–2279, <https://doi.org/10.18653/v1/2022.findings-acl.177>, 2022.
- Masry, A., Kavehzadeh, P., Do, X. L., Hoque, E., and Joty, S.: UniChart: A Universal Vision-language Pretrained Model for Chart Comprehension and Reasoning, *arXiv [preprint]*, <https://doi.org/10.48550/arXiv.2305.14761>, 2023.
- Matsumoto, H., Tanioka, Y., Nishimura, Y., Tsuji, Y., Namegaya, Y., Nakasu, T., and Iwasaki, S.-I.: Review of Tide Gauge Records in the Indian Ocean, *J. Earthq. Tsunami*, 3, 1–15, <https://doi.org/10.1142/S1793431109000378>, 2009.
- McInnes, K. L., O’Grady, J. G., Hague, B. S., Gregory, R., Hoeko, R., Trenham, C., and Stephenson, A.: Digitizing the Williamstown, Australia Tide-Gauge Record Back to 1872: Insights Into Changing Extremes, *J. Geophys. Res.-Oceans*, 129, e2024JC020908, <https://doi.org/10.1029/2024JC020908>, 2024.
- McLoughlin, P. J., McCarthy, G. D., Nolan, G., Lawlor, R., and Hickey, K.: The Accurate Digitization of Historical Sea Level Records, *Geosci. Data J.*, *gdj3.256*, <https://doi.org/10.1002/gdj3.256>, 2024.
- Mcrae, M.: Turning the Tides on Historical Sea Level Records Is Giving up Valuable Secrets, Australia’s National Science Agency, <https://www.csiro.au/en/news/All/Articles/2018/May/turning-the-tides-on-historical-sea-level-records-is-giving-up-valuable-secrets> (last access: 5 March 2025), 2018.
- Methani, N., Ganguly, P., Khapra, M. M., and Kumar, P.: PlotQA: Reasoning over Scientific Plots, in: *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, IEEE, Snowmass Village, CO, USA, 1516–1525, ISBN 978-1-7281-6553-0, <https://doi.org/10.1109/WACV45572.2020.9093523>, 2020.
- Mitchell, M., Muftakhidinov, B., Winchen, T., Schaik, B. V., Wilms, A., Kylesower, Kensington, Jędrzejewski-Szmek, Z., Badger, T. G., and Badshah400: Markumtmitchell/Engauge-Digitizer: Version 12.1 Directory Dialogs Start in Saved Paths, Zenodo [code], <https://doi.org/10.5281/zenodo.3558440>, 2019.
- Murdy, J., Orford, J., and Bell, J.: Maintaining Legacy Data: Saving Belfast Harbour (UK) Tide-Gauge Data (1901–2010), *GeoResJ*, 6, 65–73, <https://doi.org/10.1016/j.grj.2015.02.002>, 2015.
- Navarro, L. M., Armstrong, C. G., Changeux, T., Frisch, D., Gil-Romera, G., Kaim, D., McClenachan, L., Munteanu, C., Szabó, P., Baranov, V., Blanco-Garrido, F., Camarero, J. J., García, M. B., Grace, M., Izdebski, A., Morueta-Holme, N., Pando, F., Schouten, R., Spitzig, A., Svenning, J.-C., Tribot, A.-S., Viana,

- D. S., and Clavero, M.: Integrating Historical Sources for Long-Term Ecological Knowledge and Biodiversity Conservation, *Nature Reviews Biodiversity*, <https://doi.org/10.1038/s44358-025-00084-3>, 2025.
- Needham, J., Ling, W., and Lu, G.: *Civil Engineering and Nautics*, vol. 4, Science and Civilisation in China, Physics and Physical Technology, Cambridge Univ. Press, Cambridge, ISBN 978-0-521-07060-7, 1971.
- Neppel, L., Renard, B., Lang, M., Ayrat, P.-A., Coeur, D., Gaume, E., Jacob, N., Payrastré, O., Pobanz, K., and Vinet, F.: Flood Frequency Analysis Using Historical Data: Accounting for Random and Systematic Errors, *Hydrolog. Sci. J.*, 55, 192–208, <https://doi.org/10.1080/02626660903546092>, 2010.
- Paulescu, D., Rogoza, M., Popa, M., and Radulian, M.: Digitized Database of Old Seismograms Recorder in Romania, *Acta Geophys.*, 64, 963–977, <https://doi.org/10.1515/acgeo-2016-0039>, 2016.
- Poco, J. and Heer, J.: Reverse-Engineering Visualizations: Recovering Visual Encodings from Chart Images, *Comput. Graph. Forum*, 36, 353–363, <https://doi.org/10.1111/cgf.13193>, 2017.
- Rao, M. P., Cook, E. R., Cook, B. I., D'Arrigo, R. D., Palmer, J. G., Lall, U., Woodhouse, C. A., Buckley, B. M., Uriarte, M., Bishop, D. A., Jian, J., and Webster, P. J.: Seven Centuries of Reconstructed Brahmaputra River Discharge Demonstrate Underestimated High Discharge and Flood Hazard Frequency, *Nat. Commun.*, 11, 6017, <https://doi.org/10.1038/s41467-020-19795-6>, 2020.
- Rehbein, M.: Künstliche Intelligenz Und Datenmobilisierung Zwischen Geschichtswissenschaft Und Archiv. Über Forschung In Der Digitalen Transformation Und Die Notwendigkeit Des Umdenkens, *Archivalische Zeitschrift*, 101, 11–27, <https://doi.org/10.17613/5m74-sk48>, 2024.
- Rehbein, M.: Bavarian Danube Water Level Reconstruction (1826–1894), Zenodo [data set], <https://doi.org/10.5281/zenodo.17296750>, 2025.
- Rehbein, M.: Kulturgut Und Digitalisierung, in: *Clio Guide – Ein Handbuch Zu Digitalen Ressourcen Für Die Geschichtswissenschaften*, 3rd edn., edited by: Daniel, S., Enderle, W., Hohls, R., Meyer, T., Prellwitz, J., Prinz, C., Schuhmann, A., and Schwandt, S., Clio Online, Berlin, <https://doi.org/10.60693/wxbk-e572>, 2026.
- Rehbein, M. and Ernst, M.: Erschließung Handschriftlicher Dokumente Zwischen Fachwissen, Citizen Science Und KI, *Bibliothek Forschung und Praxis*, 47, 503–513, <https://doi.org/10.1515/bfp-2023-0043>, 2023.
- Savva, M., Kong, N., Chhajta, A., Fei-Fei, L., Agrawala, M., and Heer, J.: ReVision: Automated Classification, Analysis and Redesign of Chart Images, in: *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, ACM, Santa Barbara California USA, 393–402, ISBN 978-1-4503-0716-1, <https://doi.org/10.1145/2047196.2047247>, 2011.
- Scheurmann, K.: Die Flussgeschichtliche Entwicklung Bayerischer Flüsse Im 19. Und 20. Jahrhundert, *Laufener Spezialbeiträge und Laufener Seminarbeiträge*, 5, 29–35, 1981.
- Schrader, L.: Grundzüge Der Ritter'schen Methode, Wasserstandsbeobachtungen Aufzutragen Und Abzubilden Und Benützung Derselben Zur Darstellung Des Sommerhochwassers Der Elbe Im Jahre 1882, *Zeitschrift des Oesterreichischen Ingenieur- und Architekten-Vereins*, 36, 55–59, 1884.
- Short, M. A., Norman, R. S., Pillans, B., De Deckker, P., Usback, R., Opydyke, B. N., Ransley, T. R., Gray, S., and McPhail, D. C.: Two Centuries of Water-Level Records at Lake George, NSW, Aust. J. Earth Sci., 68, 453–472, <https://doi.org/10.1080/08120099.2020.1821247>, 2021.
- Spanoudi, S., Golfinopoulos, A., and Kalavrouziotis, I.: Water Management in Ancient Alexandria, Egypt. Comparison with Constantinople Hydraulic System, *Water Supp.*, 21, 3427–3436, <https://doi.org/10.2166/ws.2021.128>, 2021.
- Sušin, N. and Peer, P.: Open-source Tool for Interactive Digitisation of Pluviograph Strip Charts, *Weather*, 73, 222–226, <https://doi.org/10.1002/wea.3001>, 2018.
- Talke, S. A. and Jay, D. A.: Nineteenth Century North American and Pacific Tidal Data: Lost or Just Forgotten?, *J. Coastal Res.*, 29, 118, <https://doi.org/10.2112/JCOASTRES-D-12-00181.1>, 2013.
- Talke, S. A., Mahedy, A., Jay, D. A., Lau, P., Hilley, C., and Hudson, A.: Sea Level, Tidal, and River Flow Trends in the Lower Columbia River Estuary, 1853–Present, *J. Geophys. Res.-Oceans*, 125, e2019JC015656, <https://doi.org/10.1029/2019JC015656>, 2020.
- Teleti, P., Hawkins, E., and Wood, K. R.: Digitizing weather observations from World War II US naval ship logbooks, *Geosci. Data J.*, 11, 314–329, <https://doi.org/10.1002/gdj3.222>, 2024.
- Ullmann, A., Pons, F., and Moron, V.: Tool Kit Helps Digitize Tide Gauge Records, *Eos, Transactions American Geophysical Union*, 86, 342–342, <https://doi.org/10.1029/2005EO380004>, 2005.
- Unnikrishnan, A. and Shankar, D.: Are Sea-Level-Rise Trends along the Coasts of the North Indian Ocean Consistent with Global Estimates?, *Global Planet. Change*, 57, 301–307, <https://doi.org/10.1016/j.gloplacha.2006.11.029>, 2007.
- Wang, K., Ellsworth, W. L., Beroza, G. C., Williams, G., Zhang, M., Schroeder, D., and Rubinstein, J.: Seismology with Dark Data: Image-Based Processing of Analog Records Using Machine Learning for the Rangely Earthquake Control Experiment, *Seismol. Res. Lett.*, 90, 553–562, <https://doi.org/10.1785/0220180298>, 2019.
- Wang, M., Yang, F., Liao, X., Wang, B., Gao, K., Zhang, L., Guo, W., Jiang, J., Yan, B., Xu, Y., and Wan, Q.: A Novel Approach to Automatically Digitize Analog Seismograms, *Seismol. Res. Lett.*, 96, 192–206, <https://doi.org/10.1785/0220240220>, 2025.
- Wasson, R., Sundriyal, Y., Chaudhary, S., Jaiswal, M. K., Morthekai, P., Sati, S., and Juyal, N.: A 1000-Year History of Large Floods in the Upper Ganga Catchment, Central Himalaya, India, *Quaternary Sci. Rev.*, 77, 156–166, <https://doi.org/10.1016/j.quascirev.2013.07.022>, 2013.
- WBGU: EU-Wasserpolicy Im Aufbruch: Die EU-Wasserresilienzstrategie Als Chance Nutzen, Tech. Rep. 14, WBGU, Berlin, https://www.wbgu.de/fileadmin/user_upload/wbgu/publikationen/politikpapiere/pp14_2025/pdf/wbgu_pp14.pdf (last access: 5 March 2025), 2025a.
- WBGU: Empfehlungen für die internationale Wassergovernance: Policy Brief, Policy Brief, Wissenschaftlicher Beirat d. Bundesregierung Globale Umweltveränderungen, Berlin, ISBN 978-3-946830-48-1, 2025b.
- Weigmann: Aus Dem Wasserbaulichen Arbeitsgebiet Der Bayerischen Staatsbauverwaltung, *Die Bautechnik*, 13, 483–487, 1935.

- Westcott, N., Cooper, J., Andsager, K., Stoecker, L., and Shein, K.: Status and Climate Applications of the 19th Century Forts and Volunteer Observer Database, *Journal of Applied and Service Climatology*, 2021, 1–14, <https://doi.org/10.46275/JOASC.2021.09.001>, 2021.
- Wetter, O.: The potential of historical hydrology in Switzerland, *Hydrol. Earth Syst. Sci.*, 21, 5781–5803, <https://doi.org/10.5194/hess-21-5781-2017>, 2017.
- Wilkinson, C., Brönnimann, S., Brugnara, Y., Brohan, P., Valente, M. A., Gilabert, A., Jourdain, S., Roucaute, E., Crouthamel, R., Manola, B.-I., and COMPO, GILBERT: Best Practice Guidelines for Climate Data Rescue, Copernicus Climate Change Service, <https://doi.org/10.24381/X9RN-MP92>, 2019.
- Willems, H. and Dahms, J.-M. (Eds.): The Nile: Natural and Cultural Landscape in Egypt: Proceedings of the International Symposium Held at the Johannes Gutenberg-Universität Mainz, 22 & 23 February 2013, transcript, ISBN 978-3-8394-3615-8, <https://doi.org/10.14361/9783839436158>, 2017.
- Winiwarter, V., Schmid, M., and Dressel, G.: Looking at Half a Millennium of Co-Existence: The Danube in Vienna as a Socio-Natural Site, *Water History*, 5, 101–119, <https://doi.org/10.1007/s12685-013-0079-x>, 2013.
- WMO, W. M. O.: Guidelines on Best Practices for Climate Data Rescue, no. 1182 in WMO, World Meteorological Organization, Geneva, Switzerland, ISBN 978-92-63-11182-1, 2024.
- Woodworth, P. L. and Blackman, D. L.: Changes in Extreme High Waters at Liverpool since 1768, *Int. J. Climatol.*, 22, 697–714, <https://doi.org/10.1002/joc.761>, 2002.
- Woodworth, P. L., Hunter, J. R., Marcos, M., Caldwell, P., Menéndez, M., and Haigh, I.: Towards a Global Higher-frequency Sea Level Dataset, *Geosci. Data J.*, 3, 50–59, <https://doi.org/10.1002/gdj3.42>, 2016.
- World Meteorological Organization: State of Global Water Resources 2024, Tech. rep., World Meteorological Organization, <https://doi.org/10.59327/WMO/WATER/2024>, 2025.
- Xue, N., Wu, T., Bai, S., Wang, F.-D., Xia, G.-S., Zhang, L., and Torr, P. H. S.: Holistically-Attracted Wireframe Parsing: From Supervised to Self-Supervised Learning, *IEEE T. Pattern Anal.*, 45, 14727–14744, <https://doi.org/10.1109/TPAMI.2023.3312749>, 2023.
- Yepes, A. J., Zhong, X., and Burdick, D.: ICDAR 2021 Competition on Scientific Literature Parsing, arXiv [preprint], <https://doi.org/10.48550/arXiv.2106.14616>, 2021.
- Zhao, H., Jia, J., and Koltun, V.: Exploring Self-Attention for Image Recognition, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Seattle, WA, USA, 10073–10082, ISBN 978-1-7281-7168-5, <https://doi.org/10.1109/CVPR42600.2020.01009>, 2020.