

# Rettung historisch, hydrometrischer Daten

## - durch den Einsatz von Künstlicher Intelligenz

Anschrift:

**Pro Konzept GmbH**

Dalbergstr. 11  
67482 Venningen

Rechtsform:

GmbH

Handelsregisternr:

Amtsgericht Landau, HRB 31066

Gründungsjahr:

1994 / 2011

Inhaberin und

Geschäftsführerin:

Petra Greiffenhagen (100% Eigentümerin)

Ihre Ansprechpartner:

**Michael Heil**

[m.heil@it-plan.com](mailto:m.heil@it-plan.com)

**Petra Greiffenhagen**

Geschäftsführerin

Mobil: 0172 / 76 84 763

Festnetz: 06323 / 958 26 400

[petra.greiffenhagen@prokonzept.de](mailto:petra.greiffenhagen@prokonzept.de)

## Die Pro Konzept GmbH

**Unternehmensberatung und Generalunternehmer für die**

- Logistik
- Verarbeitung
- Digitalisierung
- Sicherung

von Daten und Dokumenten sowie die Beschaffung von  
sensiblen Wissen und Informationen.

**Kurz: Wir kümmern uns um alles!**



# Rettung historisch, hydrometrischer Daten

- durch den Einsatz von Künstlicher Intelligenz

## Exkurs

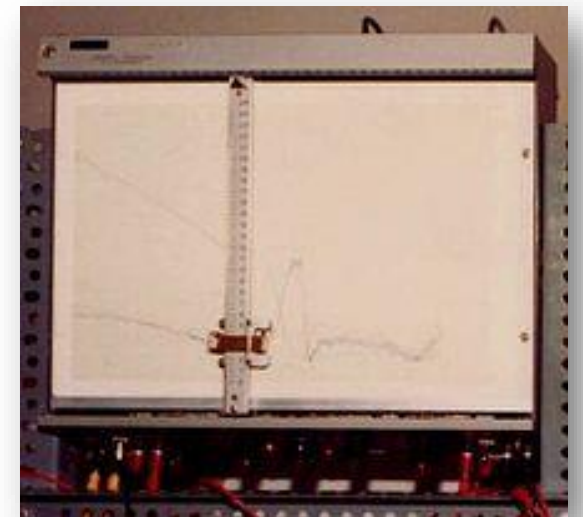
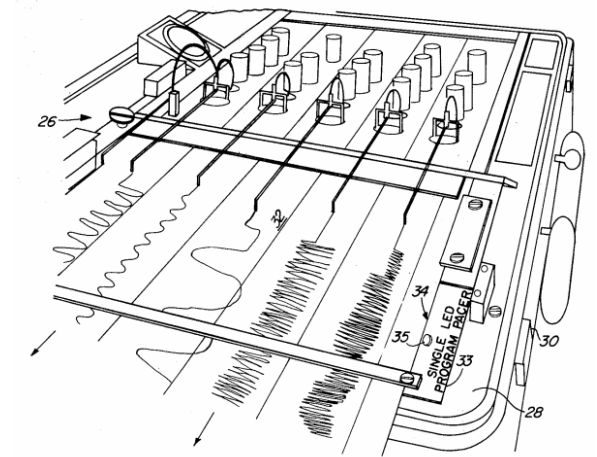
- Weltweit existieren Millionen von Papierbögen mit hydrometrischen Daten in Archiven, teilweise bis zu **100 Jahre** alt
- Auch bei optimaler Lagerung **gehen diese Daten** in absehbarer Zeit, durch Alterungsprozesse, **verloren**
- Durch Scannen wird ein digitales *Abbild* erzeugt, das archiviert und damit gesichert wird

... eine **Digitalisierung** liegt aber erst durch **Transformation** des *Bildes* in **diskrete Daten** vor!



## Die Ausgangslage

- In der Regel werden die Papiere in Archiven mit zum Teil wechselnden **klimatischen Bedingungen** (Feuchtigkeit, Temperatur, Licht) gelagert
- Diese Umwelteinflüsse verändern die Beschaffenheit des Papiers, es wird **wellig, bleicht-** oder **trocknet** aus
- **Unterschiedliche**, nicht DIN-konforme **Ablagetypen** und **-formate**
- Papier teilweise **über 100 Jahre** alt und kaum noch lesbar
- **Zusätzlich** zu der Ganglinie des XT-Schreibers **existieren** weitere **Artefakte** (Stempel, Kommentare, Hilfslinien)
- **Keine** ausreichend **klassifizierten Daten** als Lernbasis **vorhanden**.  
Das manuelle klassifizieren ist zu Zeit- und Kostenintensiv.



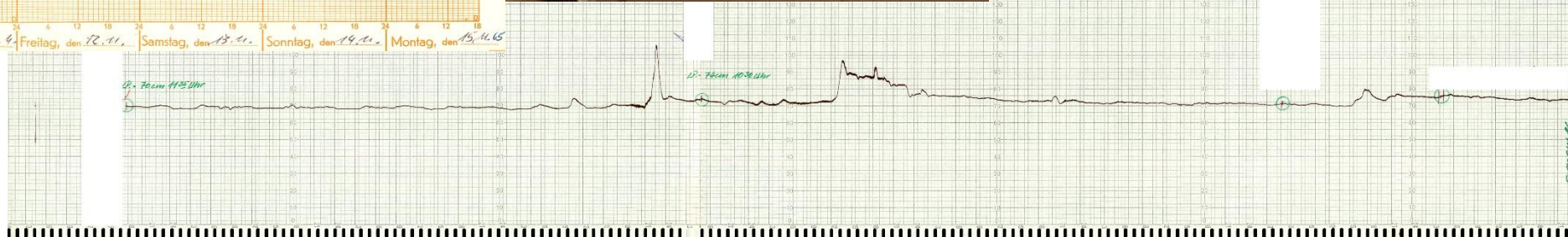
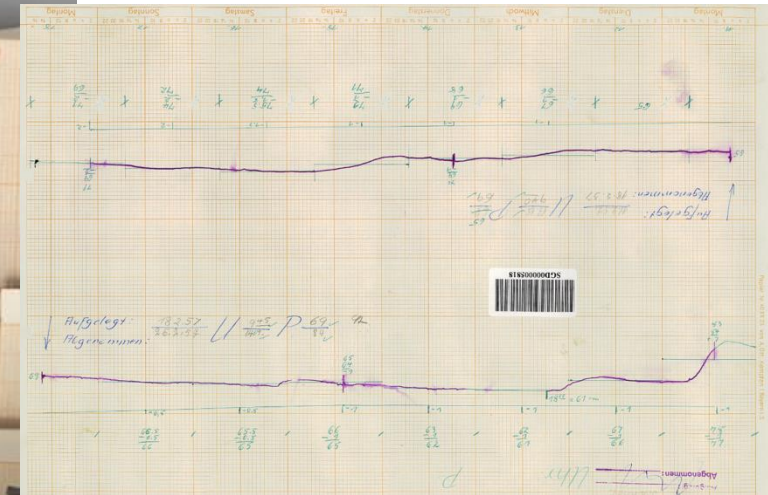
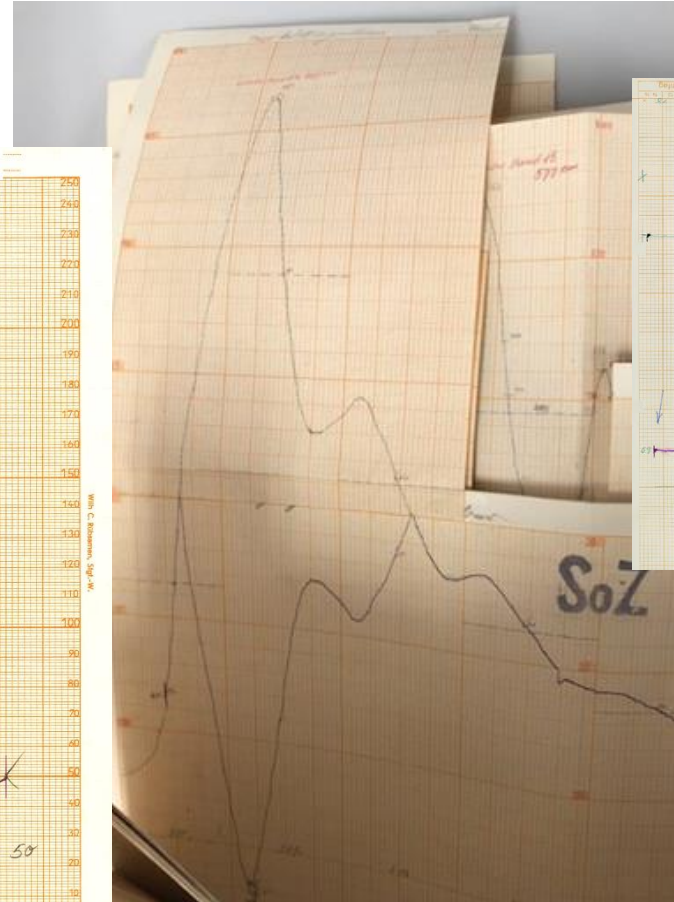
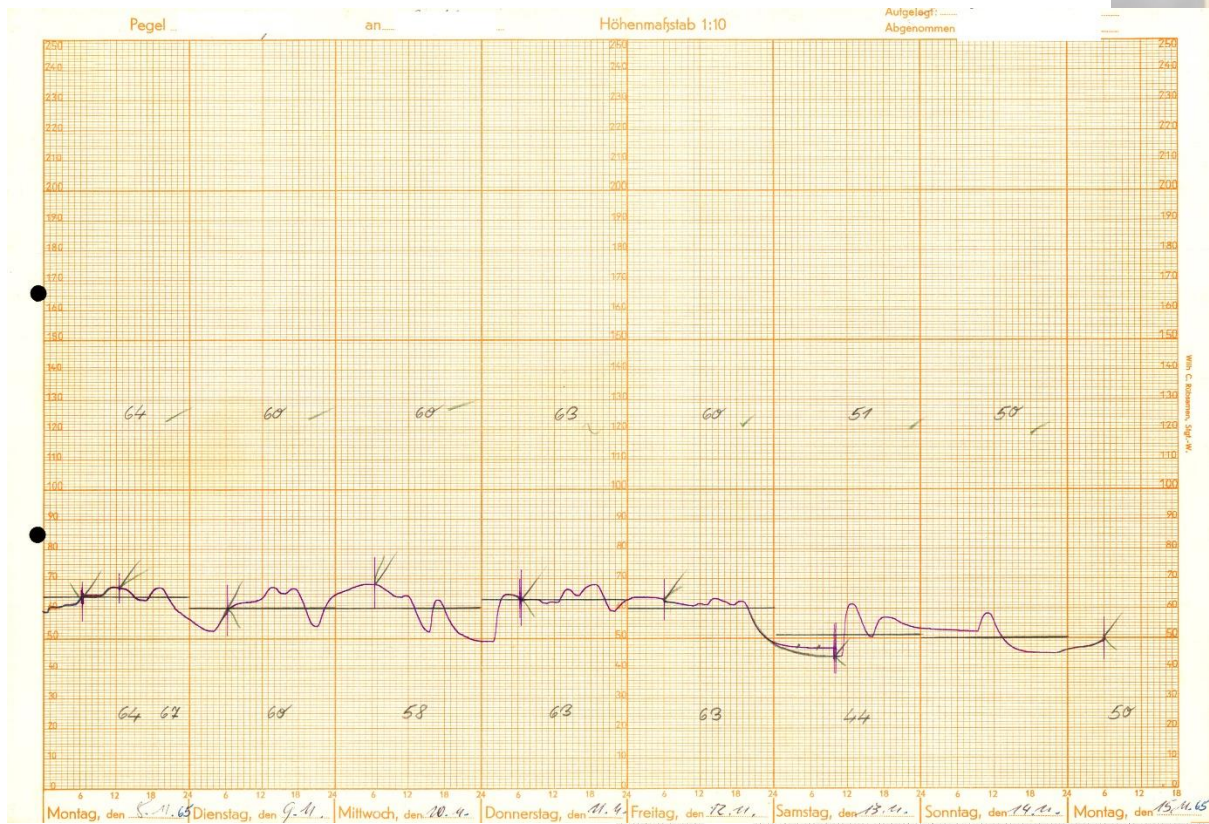
## Beispiele

PRO KONZEPT 



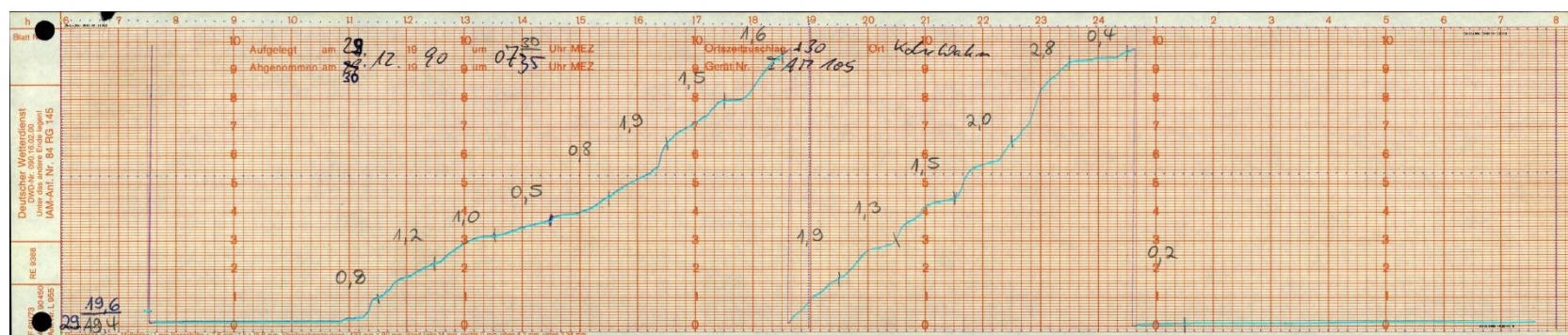
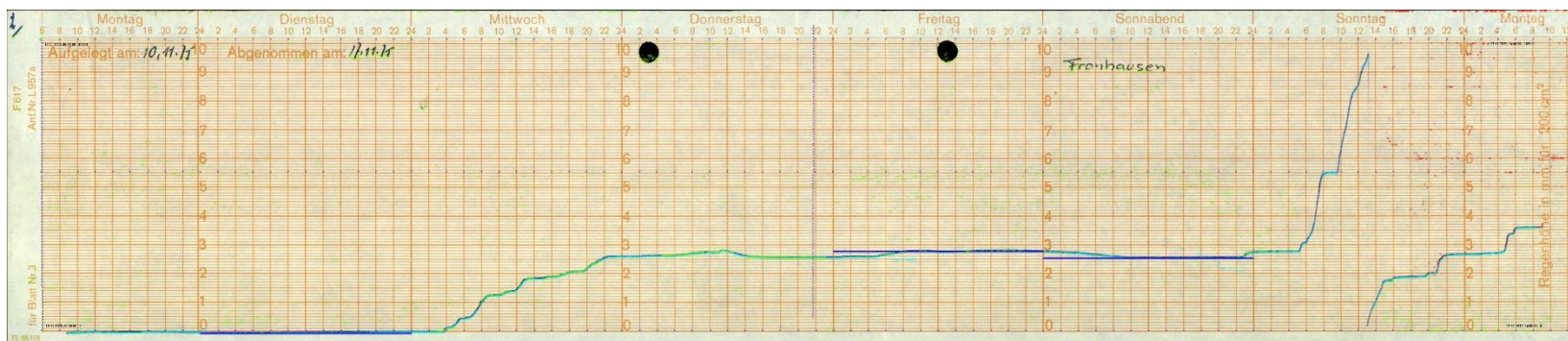
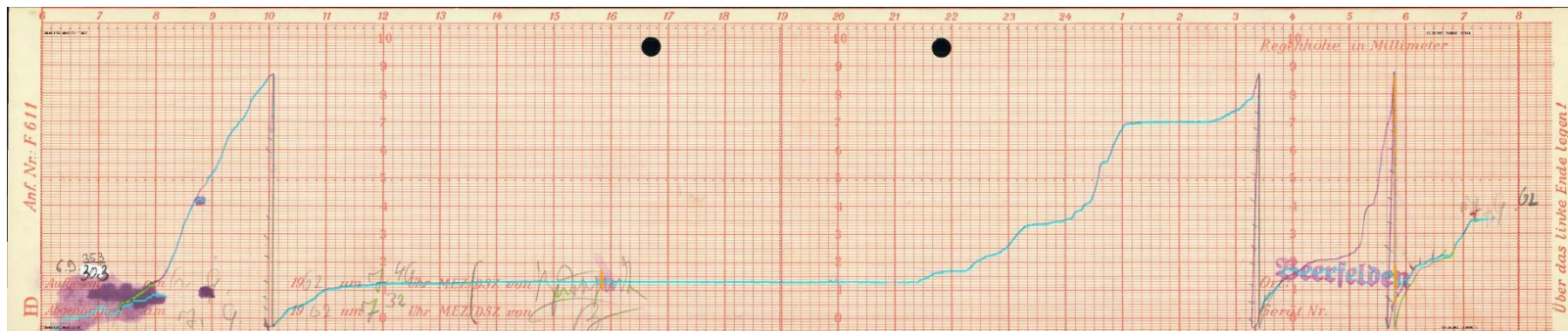
## Beispiele

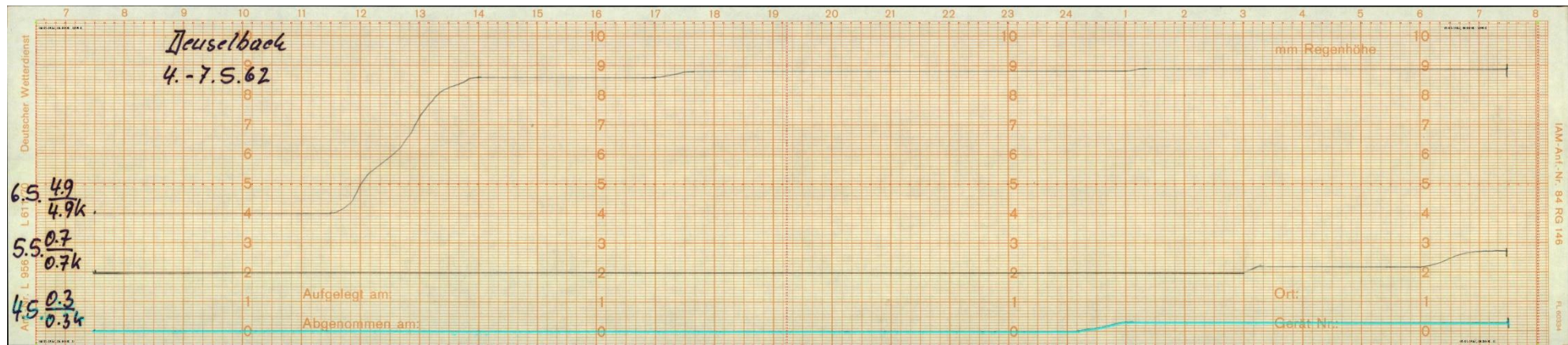
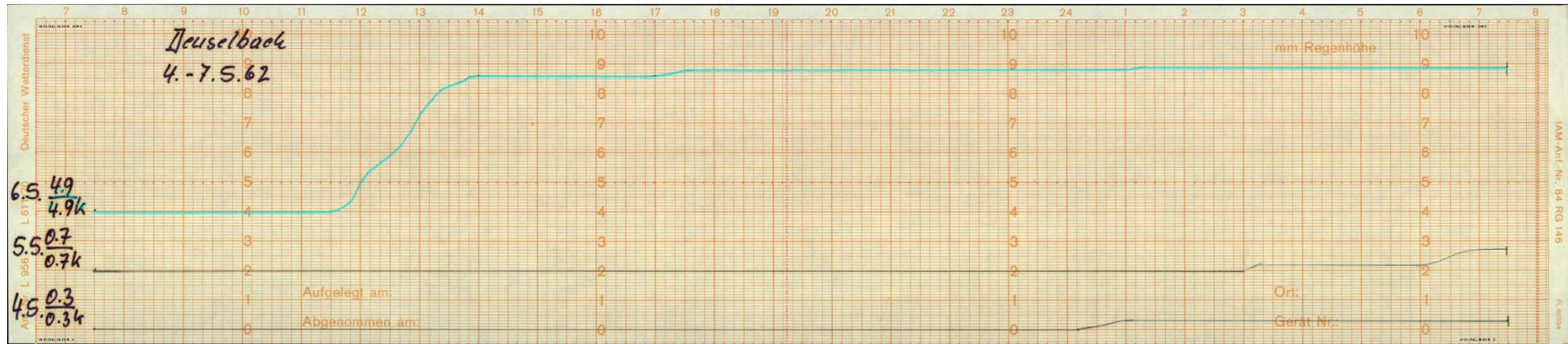
In der bisherigen Verarbeitung wurden über 300 Varianten ermittelt.

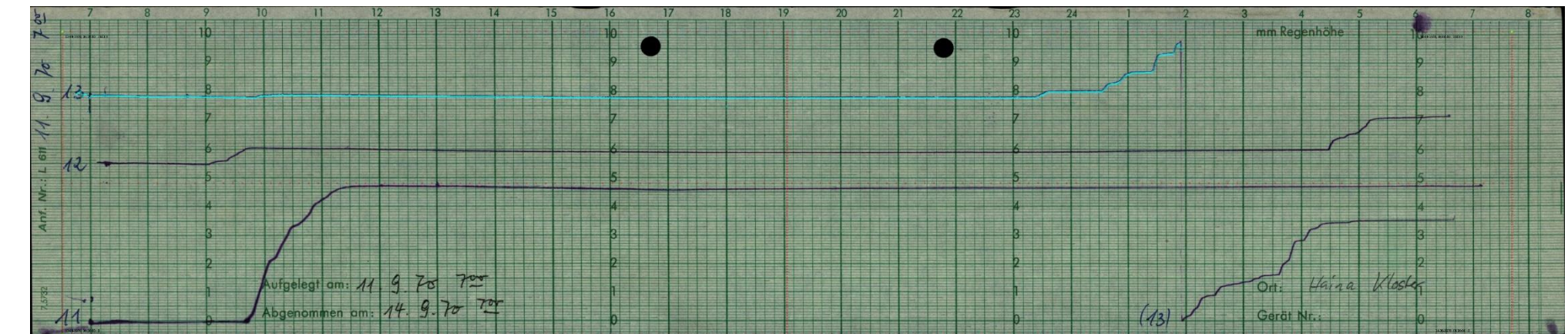
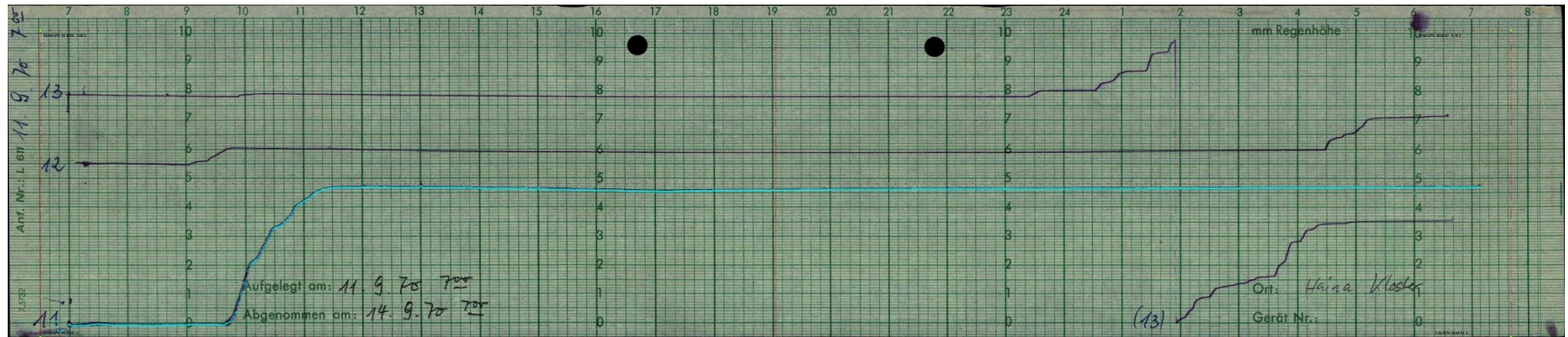
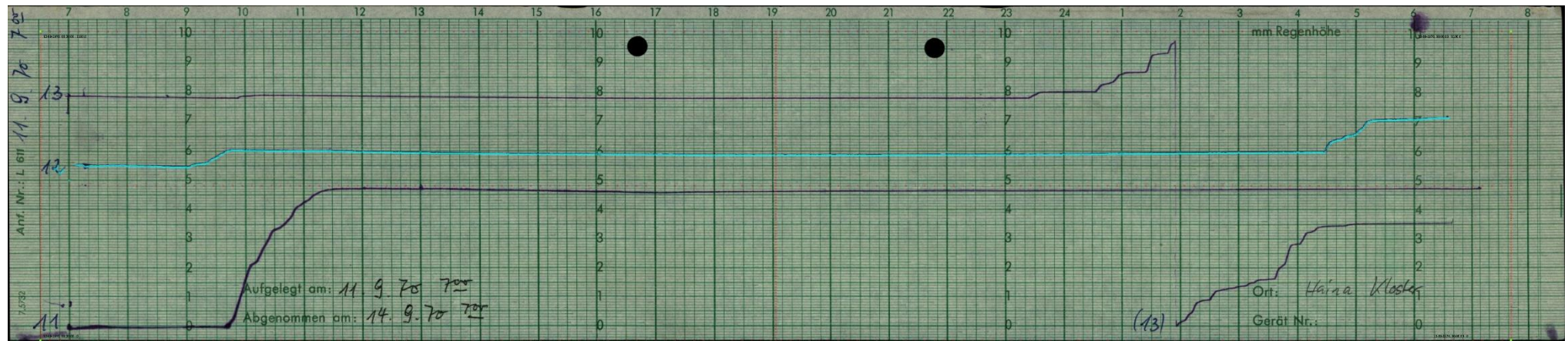


# Neues Projekt: Niederschlagsbögen

PRO KONZEPT







## Die Zielsetzung

1. Erkennen der Ganglinie auf dem Bild
2. Eliminieren von Artefakten und stärken der Ganglinie
3. Ermitteln des Bezugssystems (aufgedrucktes Koordinatensystem)
4. Korrigieren von Digitalisierungsfehlern
5. Transformation der Bild-Pixel in diskrete Wertepaare (Zeit/Höhe)

**Und das so autonom, schnell und exakt wie möglich!**

## Die Herausforderung

1. Sehr große Mengen an Daten,  
→ daher extrem schnelle Verfahren erforderlich
2. Keine qualifizierten-, etikettierten Daten vorhanden  
→ *klassisches machine learning* nicht möglich
3. Sehr heterogene Daten in Ausprägung und Qualität  
→ Einsatz *klassischer Algorithmen* in Kombination mit sehr großen Datenmengen nicht möglich
4. Nicht das Erkennen eines Pegelbogens ist das primäre Ziel, sondern **die exakte Vermessung** der Gang Linie auf dem Bild!

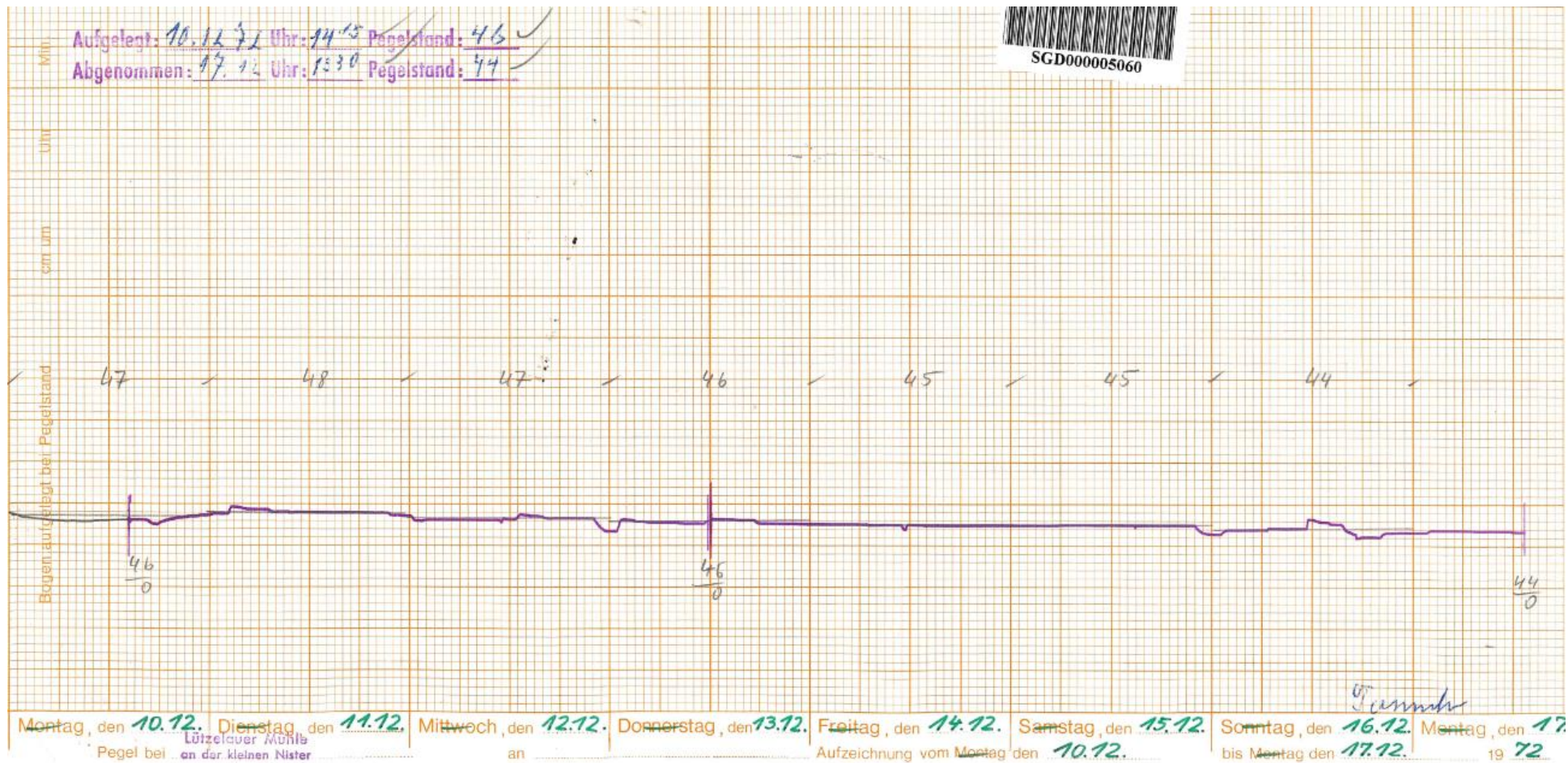


## Der Lösungsansatz

1. Verschiedene **machine learning** und weitere KI-Ansätze zu verbinden
2. Einsatz **klassischer Algorithmen**, um Zielvorgaben zu definieren wo qualifizierte Beispieldaten fehlen
3. Einsatz **machine learning**, wo klassische Algorithmen die Komplexität durch heterogene Daten nicht beherrschen
4. Aufbau von qualifizierten Datensätzen, um den Einsatz von **maschine learning** auszubauen

## Erkennen der Ganglinie auf dem Bild

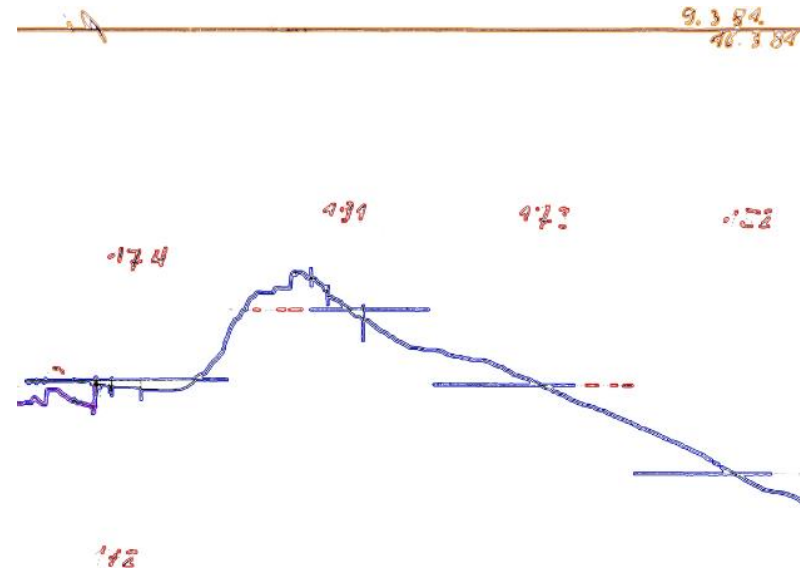
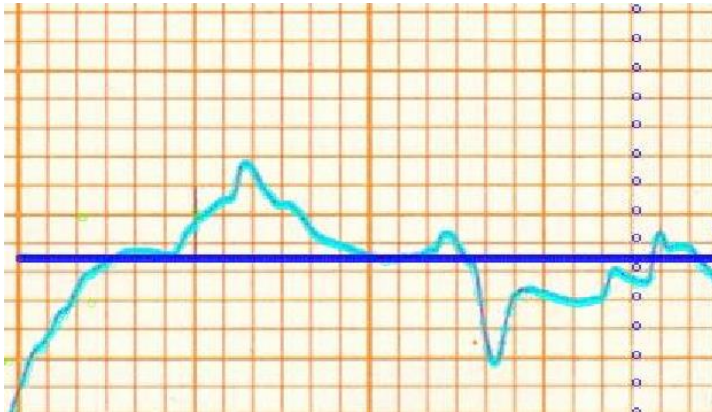
- Durch Einsatz von **Segmentierung** werden auf dem Bild Objekte separiert





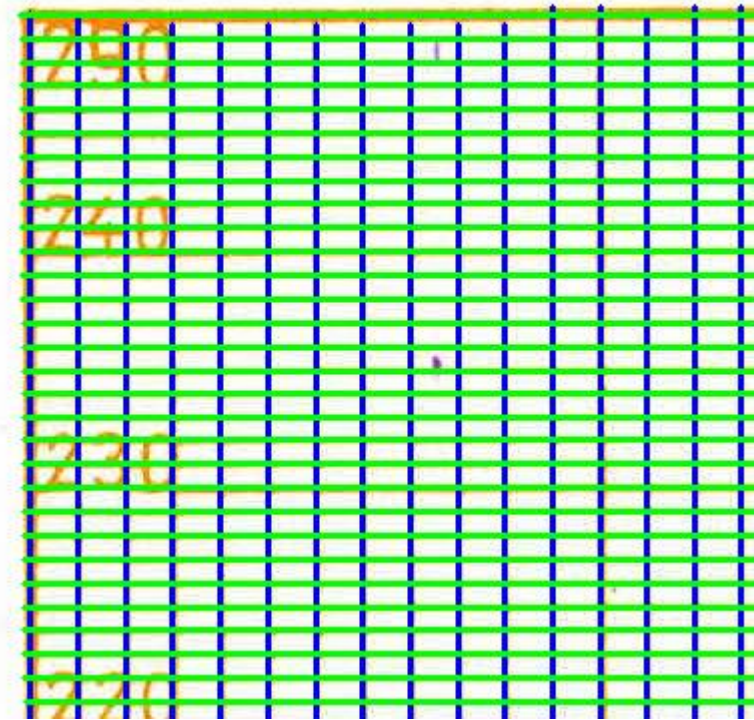
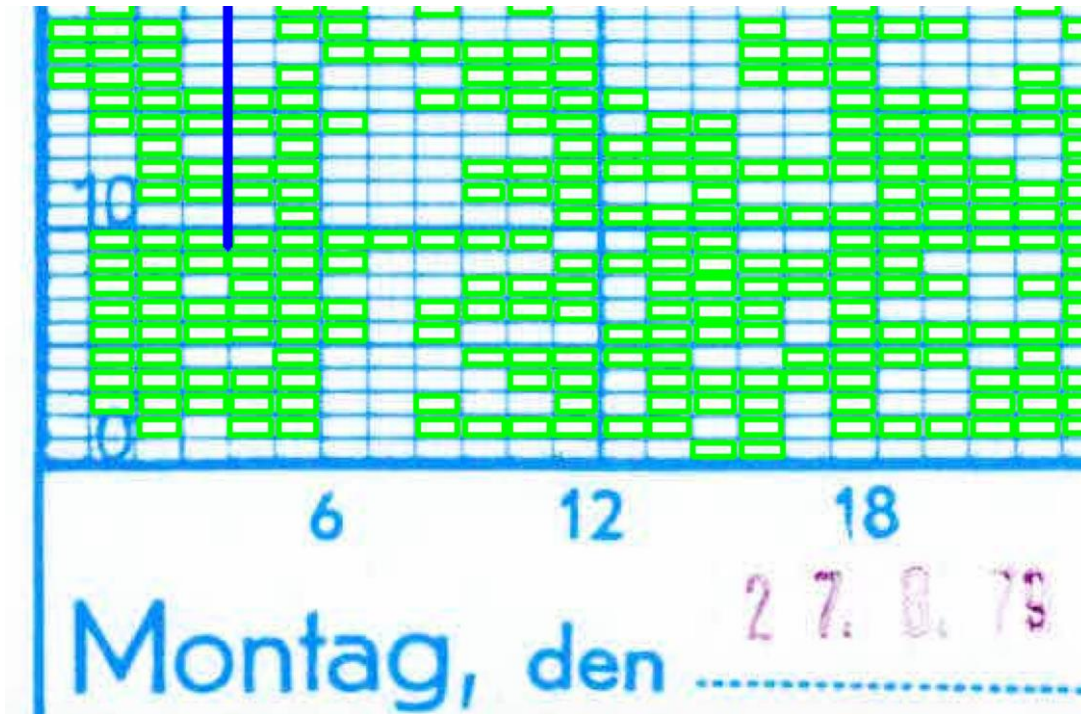
## Eliminieren von Artefakten und stärken der Ganglinie

- Objekte werden zurzeit durch klassische Algorithmen kategorisiert, die mathematische und stochastische Funktionen verwenden
- Diese Klassifizierung wird verwendet um:
  - die Ganglinie zu stärken (z.B. Überdeckung der Gang Linie durch einen Mittelwert erkennen)
  - Artefakte wie zum Beispiel Stempel zu eliminieren
  - und qualifizierte Daten aufzubauen



## Ermitteln des Bezugssystems (aufgedrucktes Koordinatensystem)

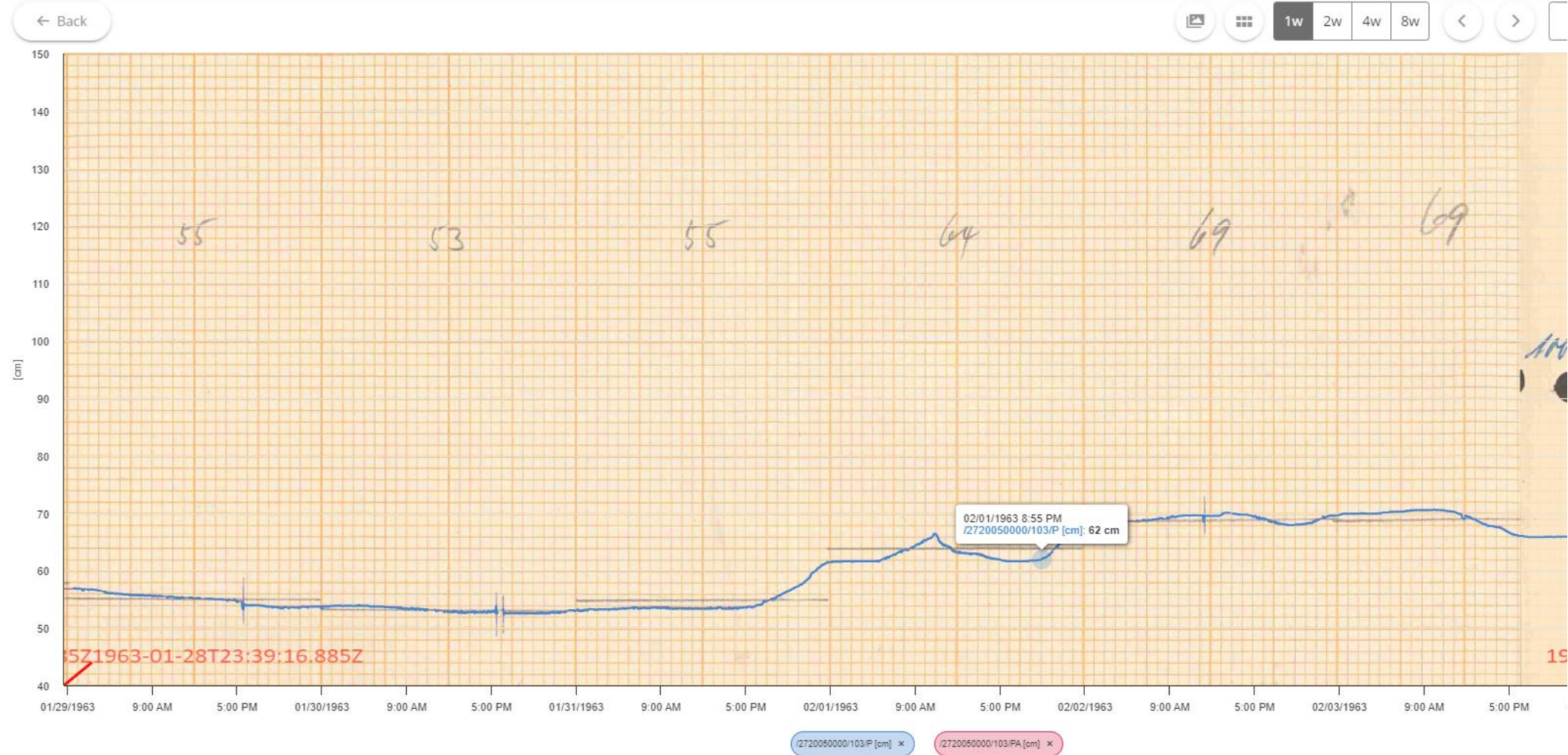
- Durch Einsatz von Techniken aus dem Bereich **computer vision** werden die Linien und die daraus resultierenden *Kästchen* des Koordinatensystems erkannt und *vermessen*.



## Korrektur von Digitalisierungsfehlern und Transformation der Bild-Pixel in diskrete Wertepaare (Zeit/Höhe)

- Beim Scannen des Papierbogens erzeugt:
  - jede Falte eine Stauchung
  - jede Welle eine Krümmung
  - jede Abweichung beim Einzug des Blattes ein Delta entlang der X-Achse
- Diese Fehler werden mit mathematischen Algorithmen eliminiert
- Jeder Pixel der Gang Linie wird als Punkt (Zeit/Höhe) ins Bezugssystem transformiert
- Im Zielsystem wird die Ganglinie mit dem digitalisierten Bild hinterlegt und visualisiert.





← Back



1w

2w

4w

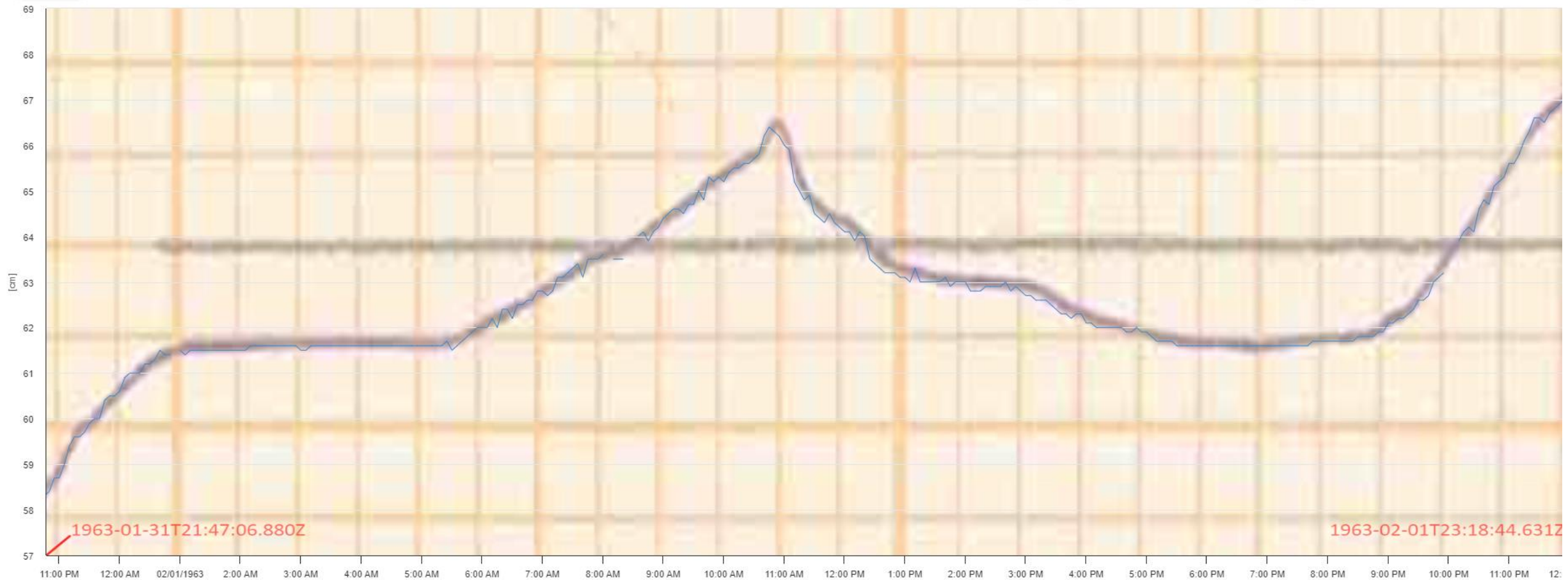
8w



01/31/1963

-

02/02/1963



/2720050000/103/P [cm] x

/2720050000/103/PA [cm] x

## Ausblick auf zukünftige Entwicklung

- Durch den Aufbau immer größerer und besser qualifizierter Datenmengen ist der Einsatz weitere ***machine learning*** Algorithmen möglich:
  - um noch exakter zwischen Artefakten wie Stempeln und der Ganglinie zu differenzieren
  - die bisher nötige manuelle Vorsortierung weiter zu minimieren
  - Abweichungen von der Norm zu erkennen und im besten Fall zu beherrschen
  - die Qualitätskontrolle effizienter und besser zu machen

## Die Zielsetzung nach wie vor.....

1. Sicherung der Pegelbögen durch Digitalisierung
2. Erschließen der Bilder durch Indizierung
3. Erkennen der Ganglinie auf dem Bild
4. Eliminieren von Artefakten und stärken der Ganglinie
5. Ermitteln des Bezugssystems (aufgedrucktes Koordinatensystem)
6. Korrigieren von Digitalisierungsfehlern
7. Transformation der Bild-Pixel in diskrete Wertepaare (Zeit/Höhe)

**Und das so autonom, schnell und exakt wie möglich!**

## Maschinell gelernte Expertise aus 1 Mio Pegelbögen

Einsatz verschiedener, individuell trainierter, KI-Modelle für:

1. Überprüfung von Gewässer und Pegel Bezeichnung  
→ ermöglicht falsch einsortierte Bögen zu finden und korrekt zuzuordnen
2. Ermittlung des Koordinatensystems „Raster“ und „Nullpunkt“  
→ manuelles prüfen/vorsortieren nach Koordinatenraster entfällt  
→ manuelles prüfen/setzen des Koordinatenursprungs entfällt
3. Blattindividuelle Anwendung des optimalen Farbprofils  
→ manuelles prüfen/vorsortieren nach Farbprofil entfällt
4. Erkennen von Start- und Enddatum  
→ manuelle Indizierung entfällt

Das macht:

- die Verarbeitung schneller, planbarer und qualitativ besser.
- Die Qualitätskontrolle effizienter und besser

# Nachtrag – Qualitäts-Management

Das Qualitäts-Management wurde in den letzten Jahren kontinuierlich erweitert.

Die **blattspezifische** Qualitätskontrolle bleibt weiterhin Grundlage des Verfahrens und wurde um eine **pegelbasierte** Betrachtung ergänzt.

Der neue Ansatz betrachtet damit zusätzlich den **Gesamtzusammenhang** aller Bögen einer Pegelstation.

## Ziel:

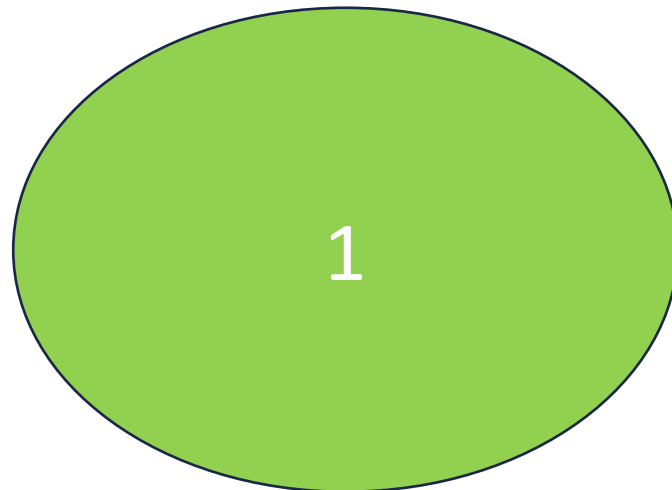
- Qualität nicht nur auf Bildebene bewerten
- Zusammenhänge über längere Zeiträume sichtbar machen
- Fehler automatisiert erkennen, einordnen und, soweit möglich, korrigieren
- manuelle Prüfungen gezielt auf relevante Fälle konzentrieren

## Überprüfung von Gewässer und Pegel Bezeichnung

- Durch KI-basierte Handschriftenerkennung und Abgleich mit den Daten vom Kunden

| Pegel-ID | Pegel-Station | Pegelbezeichnung | Gewässer | Historische-ID | Alias-Pegel-ID | Station-Kürzel | Gewässer-Kürzel | Gewässer-ID |
|----------|---------------|------------------|----------|----------------|----------------|----------------|-----------------|-------------|
|----------|---------------|------------------|----------|----------------|----------------|----------------|-----------------|-------------|

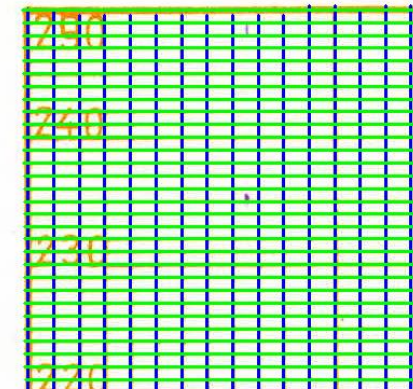
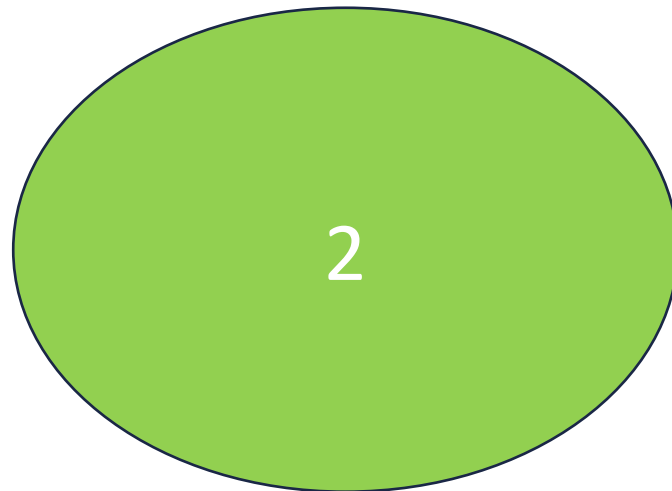
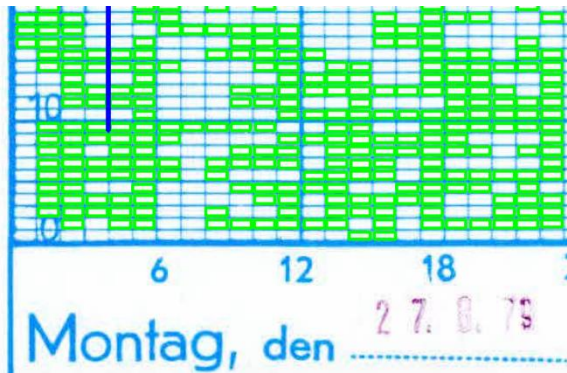
- Vorteil: Falschablagen können erkannt und korrigiert werden
- Lücken in der Zeitachse werden dadurch geschlossen



## Ermitteln des Bezugssystems (aufgedrucktes Koordinatensystem)

Durch Einsatz eines Computer-Vision-ResNet-Modells werden:

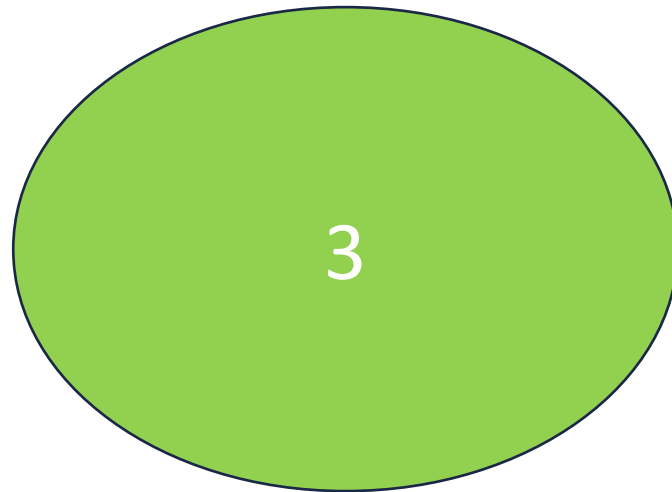
- die Linien und die daraus resultierenden *Kästchen* des Koordinatensystems erkannt und *vermessen*.
- die Bildkoordinate des Koordinatenursprungs auf wenige Pixel genau bestimmt



## Indizierung von Start- und Enddatum

Durch ein Vision Transformer-Modell:

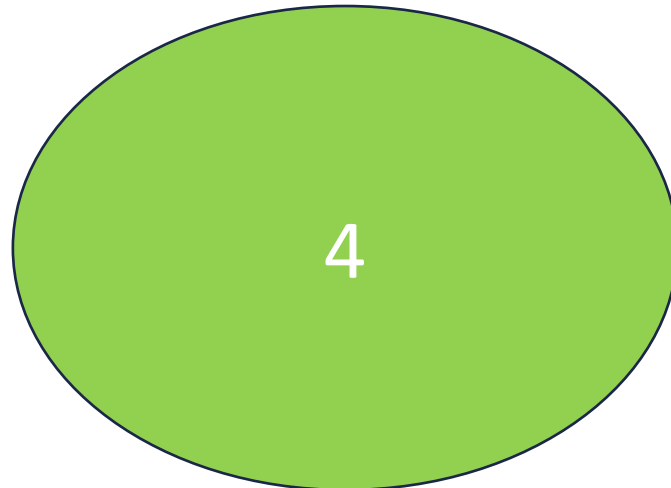
- erkennt handschriftliche Datumsangaben
- und interpretiert diese kontextbasiert als Start- und Enddatum.
- Die manuelle Erfassung von Einzelblatt-Bögen entfällt.



## Blattindividuelle Anwendung des optimalen Farbprofils

Durch den Einsatz eines KI-Modells zur adaptiven Farbprofil-Ermittlung:

- entfällt die Vorsortierung der Bilder nach Farbprofilen durch den Scandienstleister
- wird für jedes Bild automatisch das optimale Farbprofil bestimmt und angewendet
- verbessert sich die Qualität des Gesamtsystems, insbesondere bei:
  - unterschiedlich verblassten Papierdrucken durch Lagerung und Alterung
  - unterschiedlich intensiven Stiftfarben durch variierenden Verbrauchsstand



## Ihre Vorteile im Überblick

- Auch sehr große Archive lassen sich digitalisieren, durch KI gestützte Indizierung der Scans.
- Spezielle Vorarbeit im Kundenarchiv entfällt\*. Die Basissortierung nach Pegel und Datum reicht aus.
- Durch KI gestützte Werkzeuge werden Standardfälle sicher verarbeitet und Sonderfälle erkannt.
- Durch den blattbezogenen Ansatz sichert die KI auch bei großer Varianz eine optimale Verarbeitung.
- Manuelle Qualitätssicherung wird effizienter und gezielter eingesetzt und durch KI unterstützt.



\* Bezieht sich auf Pegelbögen im Einzelblattformat

## Abgrenzung der neuen QM-Ebene

### Blattbezogene Prüfung:

Es werden technische Merkmale erhoben, die Hinweise auf die Qualität der Verarbeitung geben:

- Fragmentierung der Ganglinie
- Anzahl erkannter Messpunkte
- und vieles mehr

### Pegelbezogene Prüfung:

Die Ergebnisse der **blattbezogenen** Prüfung werden in einen größeren Kontext gestellt:

- Prüfung der zeitlichen Reihenfolge → Erkennung von Lücken/ Überlappungen in der Zeitachse
- Erkennung von Mustern und Auffälligkeiten innerhalb der Gesamtserie
- Einordnung einzelner Bögen im fachlichen Gesamtkontext

## Pegelbasiertes Qualitäts-Management

- **Erkennen von Lücken**

Die Folge der Bögen sollte zeitlich geschlossen sein.

Lücken weisen auf mögliche Fehler bei der Indizierung von Start- oder Enddatum hin.

- **Erkennen von Überlappungen**

Zeitliche Überschneidungen zeigen an, dass Bögen doppelt, falsch oder widersprüchlich eingeordnet wurden.

- **Vergleich in zeitlicher Folge**

Durch die chronologische Betrachtung lassen sich Muster erkennen, die auf Auffälligkeiten einzelner Bögen hindeuten.

- **Automatische Korrektur von Fehlern**

Vor allem bei der Indizierung können viele Fehler automatisch erkannt und korrigiert werden.

- **Gezielte Markierung von Prüffällen**

Auffällige Bögen werden gezielt für eine manuelle Kontrolle ausgewählt.

**Damit wird die Qualitätssicherung umfassender, robuster und effizienter.**

## Heatmaps für den schnellen Gesamtüberblick

Zur pegelbasierten Qualitätsbewertung werden verschiedene Heatmaps erzeugt.

Sie zeigen unterschiedliche Sichten auf die Daten und machen Auffälligkeiten über den gesamten Pegel schnell sichtbar.

Eine ausführliche Beschreibung der unterschiedlichen Sichten erfolgt im Anhang.

### Nutzen der Heatmaps

- schneller Überblick über den kompletten Pegel
- Auffälligkeiten sofort sichtbar
- Prüffälle können gezielt identifiziert werden
- technische und fachliche Probleme werden in ihrem zeitlichen Zusammenhang erkennbar

# Anhang – Heatmaps

- ✓ **Anhang – Heatmap 1: Abdeckung des Koordinatensystems**
- ✓ **Anhang – Heatmap 2: Ausnutzung des Koordinatensystems durch die ROI**
- ✓ **Anhang – Heatmap 3: Geschlossenheit der Kontur**
- ✓ **Anhang – Heatmap 4: Qualität der Quantisierung**
- ✓ **Anhang – Heatmap 5: Kontinuität der ROI**
- ✓ **Anhang – Heatmap 6: Zustand des Koordinatensystems**

## Heatmap 1 – Abdeckung des Koordinatensystems

Diese Heatmap zeigt, ob das erkannte Koordinatensystem groß genug ist, um den relevanten Bereich des Bildes abzudecken.

Bei randlosen Bildern liegt der Wert im Idealfall nahe bei **1,0**.

Bei Bildern mit Rand liegt der Wert typischerweise noch bei etwa **0,9**.

Sinkt der Wert deutlich unter **0,9**, deutet dies auf mögliche Fehler hin:

- falsche Indizierung des Zeitraums
- zu viele oder zu wenige Tage auf dem Bogen
- falsche Angaben zum Raster
- fehlerhafte Konfiguration des Koordinatensystems

### Prüffall:

Bei sehr kleinen Werten wird der Bogen als möglicher Fall für ein fehlerhaftes Koordinatensystem markiert.

## Heatmap 2 – Ausnutzung des Koordinatensystems durch die ROI

Diese Heatmap zeigt, wie viel des Koordinatensystems durch die erkannte Kontur beziehungsweise Region of Interest genutzt wird.

Bei Bögen mit durchgehender Ganglinie sollte der Wert nahe bei **1,0** liegen.

Kleinere Werte können darauf hinweisen, dass:

- nur ein Teil des Bogens eine Ganglinie enthält
- die Ganglinie sehr schwach erkannt wurde
- die Farbpalette nicht optimal eingerichtet ist
- keine verwertbare Kontur gefunden wurde

### Sonderwerte:

- **0,0**: keine Kontur gefunden
- **-1,0**: Bogen fehlt
- **-2,0**: zeitliche Überlappung mit einem anderen Bogen

### Prüffall:

Kleine Werte zeigen Bögen, bei denen die Konturerkennung geprüft werden sollte.

## Heatmap 3 – Geschlossenheit der Kontur

Diese Heatmap zeigt, wie geschlossen die erkannte Ganglinie innerhalb der Region of Interest ist.

Der optimale Wert liegt bei **1,0**.

Das bedeutet, dass die erkannte Kontur die ROI möglichst vollständig abdeckt.

Kleinere Werte zeigen an, dass zwischen den Konturstücken Lücken vorhanden sind.

Je kleiner der Wert ist, desto stärker ist die Ganglinie unterbrochen.

Mögliche Ursachen sind:

- schwach erkennbare Ganglinie
- Lücken in der Stiftlinie
- Störungen oder Artefakte im Bild
- nicht optimal eingerichteter Farbraum

### Prüffall:

Bögen mit großen Lücken in der Kontur werden für die manuelle Prüfung markiert.

## Heatmap 4 – Qualität der Quantisierung

Diese Heatmap zeigt das Verhältnis der tatsächlich quantisierten Messpunkte zu den theoretisch möglichen Messpunkten.

Bei optimal quantisierten Bögen liegt der Wert bei **1,0 \***.

Kleinere Werte zeigen an, dass innerhalb der erkannten Kontur nicht alle Punkte der Ganglinie als Messpunkte übernommen werden konnten.

Mögliche Ursachen sind:

- schwache Ganglinie
- ungünstiger Farbraum
- unterbrochene Linienführung
- Bildstörungen oder Artefakte

### **Prüffall:**

Kleine Werte weisen auf eine mögliche schlechte Quantisierung hin.

(\*) In Sonderfällen kann der Wert über **1,0** liegen. Dies zeigt, dass eine schwache Ganglinie besser erkannt wurde, als es die einzelnen Konturstücke zunächst erwarten lassen.

## Heatmap 5 – Kontinuität der ROI

Diese Heatmap zeigt, wie stark die erkannte Region of Interest fragmentiert ist.

Der optimale Wert ist **1**.

Das bedeutet, dass nur eine zusammenhängende Kontur erkannt wurde.

Steigt der Wert, zeigt dies an, dass die Kontur in mehrere Einzelteile zerfallen ist.

Hohe Werte können auf folgende Probleme hinweisen:

- stark fragmentierte Ganglinie
- Unterbrechungen in der Linienführung
- schwache oder wechselnde Stiftfarbe
- Störungen im Bild
- unvollständige Konturerkennung

### **Nutzen:**

Die Heatmap macht sichtbar, bei welchen Bögen die Ganglinie nicht als stabile zusammenhängende Struktur erkannt wurde.

## Heatmap 6 – Zustand des Koordinatensystems

Diese Heatmap zeigt, auf welcher Grundlage das Koordinatensystem eines Bogens erkannt wurde.

Dabei werden unterschiedliche Zustände klassifiziert.

Die Bewertung reicht von vollständig erkannten Koordinatensystemen bis hin zu Koordinatensystemen, die nur auf Basis der Konfiguration erzeugt wurden.

Mögliche Bewertungsstufen sind:

- Koordinatensystem vollständig erkannt
- Koordinatensystem teilweise erkannt
- Koordinatensystem durch erkannte Linien ergänzt
- Koordinatensystem weitgehend aus Konfiguration abgeleitet
- Koordinatensystem vollständig künstlich aus Konfiguration erzeugt

### **Wichtig:**

Dieser Wert wird vor allem in Kombination mit anderen Heatmaps bewertet.

So lässt sich erkennen, ob Auffälligkeiten bei der Quantisierung oder Konturerkennung möglicherweise durch ein unsicher erkanntes Koordinatensystem verursacht wurden.

## Heatmap 6 – Zustand des Koordinatensystems

Diese Heatmap zeigt, auf welcher Grundlage das Koordinatensystem eines Bogens erkannt wurde.

Dabei werden unterschiedliche Zustände klassifiziert.

Die Bewertung reicht von vollständig erkannten Koordinatensystemen bis hin zu Koordinatensystemen, die nur auf Basis der Konfiguration erzeugt wurden.

Mögliche Bewertungsstufen sind:

- Koordinatensystem vollständig erkannt
- Koordinatensystem teilweise erkannt
- Koordinatensystem durch erkannte Linien ergänzt
- Koordinatensystem weitgehend aus Konfiguration abgeleitet
- Koordinatensystem vollständig künstlich aus Konfiguration erzeugt

### **Wichtig:**

Dieser Wert wird vor allem in Kombination mit anderen Heatmaps bewertet.

So lässt sich erkennen, ob Auffälligkeiten bei der Quantisierung oder Konturerkennung möglicherweise durch ein unsicher erkanntes Koordinatensystem verursacht wurden.

## Ihre Vorteile im Überblick

- **Ein Ansprechpartner** für den kompletten Prozess => wir kümmern uns um alles!
- **Automatisierte Datenextraktion** durch speziell für die Kurvenauslesung entwickelte **Software**
- **Tiefgehende Analysemöglichkeiten** und **individuelle** Datenlieferung
- **Jegliche Wünsche umsetzbar** durch skalierbare Prozesse
- **Keine** manuelle -und daher fehleranfällige- Kurvenanalyse
- Massenverarbeitung der Dokumente möglich und daher **sensationelle Durchlaufzeiten**



## Weitere Anwendungsgebiete

- **Pegelstandmessungen** von Kanälen, Flüssen und Seen
- **Fließgeschwindigkeiten** in der Natur oder in industriellen Anlagen
- **Temperatur** und **Luftfeuchtigkeit** von Räumen mit wertvollem Inhalt (Museen, Archivräume, etc.)
- **Seismografische Auswertungen** (Erdbeben)
- **EKG-Aufzeichnungen** und **CTG Herzton-Wehenschreiber**
- **Druckänderungen** von Luftdrucken und Gasen
- Ältere **Flugschreiber** in der kommerziellen Luftfahrt



**Machen Sie jetzt den Schritt  
in die digitale Zukunft**

—

**wir erstellen gerne Ihr  
individuelles Angebot!**