

Zum Einsatz automatisierter Pipelines zur Datenqualifizierung im Talsperrenmanagement

Benjamin Mewes¹ und Ihno Ihnen²

¹Okeanos Smart Data Solutions, Bochum, Deutschland

²OttHydromet, Kempten, Deutschland

Der Beitrag untersucht die automatisierte Datenqualifizierung von Bauwerksdaten in der Talsperrenwirtschaft, insbesondere die Verbindung zwischen dem SICA-Algorithmus von Okeanos und der Messdatenhaltung Aquarius von Aquatic Information. Die zuverlässige Erfassung und Auswertung dieser Daten sind entscheidend für die Sicherheit und Effizienz von Talsperren. Der Fokus liegt auf den neuesten Entwicklungen im Bereich der automatisierten Datenqualifizierung, einschließlich künstlicher Intelligenz und maschinellem Lernen zur Fehlererkennung und -korrektur. Der SICA-Algorithmus zeichnet sich durch den Einsatz unbewachten Lernens aus, was eine flexible und adaptive Lösung für die Bauwerksanalyse ermöglicht. Ein praktisches Beispiel illustriert die Anwendung des SICA-Algorithmus in Verbindung mit Aquarius und liefert Einblicke zur Weiterentwicklung der automatisierten Datenqualifizierung in der Talsperrenwirtschaft.

This article examines the automated data qualification of structural data in the dam industry, in particular the connection between the SICA algorithm from Okeanos and the Aquarius measurement data management system from Aquatic Information. The reliable collection and evaluation of this data is crucial for the safety and efficiency of dams. The focus is on the latest developments in automated data qualification, including artificial intelligence and machine learning for error detection and correction. The SICA algorithm is characterized using unsupervised learning, which enables a flexible and adaptive solution for structural analysis. A practical example illustrates the application of the SICA algorithm in conjunction with Aquarius and provides insights into the further development of automated data qualification in the dam management.

1. Einleitung

Im Talsperrenmanagement fallen große Datenmengen unterschiedlichen Ursprungs an. Die Datenniveaus dieser Daten unterscheiden sich dabei in ihrer räumlichen sowie zeitlichen Skala. Das alleinige Aufnehmen dieser ist nach teils Jahrzehnten der Sammlung nicht ausreichend. Um diesen wasserwirtschaftlichen Schatz zu heben, gilt es diese Daten über automatisierbare Pipelines zu verarbeiten und Auswertungsroutinen bereitzustellen und damit auszuwerten.

Somit werden reine Datensensoren, wie Datenbanksysteme zu modernen Datenhaltungen aufgewertet und der manuelle wasserwirtschaftliche Aufwand zur Systembetrachtung reduziert. Anhand eines Praxisbeispiels zeigen wir wie eine einfache Datenpipeline aufgebaut und ein dezentral arbeitender Gruppierungs-Algorithmus zur Detektion von unterschiedlichen Abflusssituationen an einem Bauwerk aufgebaut sein kann.

2. Pipeline Struktur zur Datenauswertung

In der Talsperrenwirtschaft ist ein effizientes Messdatenmanagementsystem von entscheidender Bedeutung. Diese Systeme ermöglichen die Erfassung, Speicherung und Verwaltung einer Vielzahl von Daten, die für die Überwachung und den Betrieb von Talsperren und ihren Einrichtungen erforderlich sind. Dabei reicht die Bandbreite der erfassten Daten von hydrologischen

Messungen über Strukturüberwachung bis hin zu Umweltparametern.

Die kontinuierliche Überwachung und Aufzeichnung dieser Daten ermöglichen es den Betreibern, den Zustand der Talsperre sowie potenzielle Risiken frühzeitig zu erkennen und entsprechende Maßnahmen zu ergreifen. Ein effektives Messdatenmanagementsystem gewährleistet dabei nicht nur die Zuverlässigkeit und Integrität der Daten, sondern auch deren Zugänglichkeit für verschiedene Anwendungen und Analysen.

Zusätzlich zur Erfassung und Verwaltung der Daten spielt die automatische Prozessierung und Verarbeitung eine wichtige Rolle. Datenpipelines ermöglichen die automatisierte Übertragung, Transformation und Analyse von Datenströmen beispielsweise über REST-Schnittstellen, was die Effizienz bei der Datenverarbeitung erheblich steigert. Durch die Implementierung von Datenpipelines können zeitkritische Analysen durchgeführt werden, um beispielsweise Anomalien oder Trends in den Messdaten zu identifizieren.

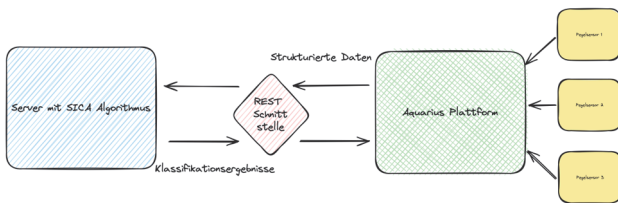


Abbildung 1: Aufbau der Datenpipeline im Praxisbeispiel zwischen Datenhaltungsplattform Aquarius und SICA-Algorithmus mittels REST-Schnittstelle.

Darüber hinaus ermöglichen Datenpipelines die nahtlose Integration verschiedener Datenquellen und -formate, was eine ganzheitliche Analyse und Entscheidungsfindung unterstützt. Sie sind somit ein wesentliches Instrument für die automatisierte Überwachung und den Betrieb von Talsperren, da sie eine schnelle und präzise Verarbeitung großer Datenmengen ermöglichen und gleichzeitig die Fehleranfälligkeit durch manuelle Eingriffe reduzieren. Insgesamt tragen effektive Messdatenmanagementsysteme und Datenpipelines wesentlich zur Sicherheit, Effizienz und Nachhaltigkeit in der Talsperrenwirtschaft bei.

2.1 Beschreibung der Datenplattform Aquarius von Aquatic Informatics

AQUARIUS ist eine Plattform für das Management von großen Datenmengen. Sie verwandelt gesammelte Rohdaten in verwertbare Informationen - für aussagekräftige Analysen, optimierte Abläufe und die Weitergabe wichtiger Erkenntnisse an die richtigen Personen zur richtigen Zeit.

Mit AQUARIUS können Wasserdaten in Echtzeit erfasst, verarbeitet und weitergeben werden, um bessere und schnellere Entscheidungen zu treffen.

2.2 Beschreibung der Fallstudie Bauwerk Ahausen

Die in diesem Beitrag verwendeten Daten stammen von drei Pegelmessstellen des Damms in Ahausen, der vom Ruhrverband betrieben wird. Die drei Abflusspegel P1, P2 und P3 messen in unregelmäßigen zeitlichen Abständen den Wasserstand in einem definierten Gerinne. Über eine W-Q-Beziehung werden diese Daten in einen Volumenstrom umgerechnet.

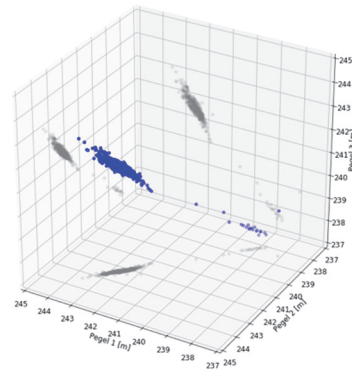


Abbildung 2: Plot der drei Pegel P1-P3 (nicht normalisiert dargestellt) unterhalb des Bauwerks Ahausen. Man sieht eindeutig eine Häufung im mittleren Bereich mit einer Häufung im Niedrigwasserbereich, sowie wenigen lokalen Ausreißern in höhere Wasserstände.

Die vorlegte Datenreihe umfasst Daten zwischen 1996 und 2020. Die drei Zeitreihen liegen bereinigt, ohne Fehlerwerte, für Zeitstempel auf 0:00h mit einem Umfang von je 910 Einträgen vor. Diese wurde in die Datenplattform Aquarius importiert und hinterlegt. Um das Beispiel einfach zu halten, wurden keine weiteren externen Daten hinzugezogen.

Für eine bessere Lesbarkeit der Daten werden diese mittels einer linearen Interpolation verbunden und in einem Dashboard mit interaktiven Elementen dargestellt.

3. Beschreibung des SICA-Algorithmus

Die Hauptanforderung an den Algorithmus zur Situationsanalyse des Abflussverhaltens am Stauwerk Ahausen ist die automatische Erkennung der Abflusskonstellation an drei Pegeln, ohne dass die Anzahl von Gruppen vorab festgelegt werden muss. Eine detaillierte Beschreibung der Abflüsse und der Hydrogeologie vor der Analyse war daher nicht erforderlich. Für diese Anwendung wurde der SICA-Algorithmus über eine REST-Schnittstelle in Aquarius integriert und verwendet.

Der SICA-Algorithmus (Similarity-based Cluster Analysis) wurde speziell für diese Anwendung entwickelt und vernachlässigt daher die Häufigkeit von Messsituationen bei der Bestimmung der Gruppen [1]. Zudem wurde der SICA-Algorithmus bereits in einigen Talsperrenmonitoring-Projekten eingesetzt [2]. Dadurch können auch kürzere Datenreihen mit weniger als 1.000 Datenpunkten statistisch analysiert werden, wie die Datensätze vom Dam in Ahausen aufweisen. Wie fast alle Algorithmen beginnt die Suche des SICA mit einer Distanzanalyse, die die Unterschiede zwischen den einzelnen Datenpunkten misst. Im Gegensatz zur Dichteanalyse betrachtet der SICA die Distanz zwischen den Datenpunkten als Maß für Unähnlichkeit anstelle von Ähnlichkeit. Anhand dieser Distanzverteilungen identifiziert der SICA ähnliche

Datenpunkte und bildet sukzessive Verknüpfungen zwischen ihnen, um Gruppen/Cluster zu bilden.

Diese Verknüpfungen werden dann als ein einziger Datenpunkt betrachtet, unabhängig von der Anzahl der dahinterstehenden Datenpunkte. Dies ermöglicht eine umfassende Gruppeneinteilung der Messwerte, einschließlich dünn besetzter Ausnahmefälle, die besonders wichtig für die Bauwerksüberwachung sind. Auch dieser Algorithmus erfordert eine Parametrisierung, jedoch mit nur zwei Parametern im Vergleich zu anderen Verfahren. Die Parameter ζ_1 und ζ_2 definieren Grenzen für die Ähnlichkeitsanalyse und steuern den Prozess der Reduktion der Anzahl der Clusterzentren.

3.1 Anbindung von SICA mittels REST-Schnittstelle

Um SICA mittels REST an Aquarius anzubinden, wurde der Algorithmus in der Programmiersprache Python implementiert und als gesonderte Funktion bereitgestellt. Der Aufbau orientiert sich dabei an gängigen Bibliotheken aus dem Bereich des maschinellen Lernens.

Zunächst wird der Algorithmus als Objekt initialisiert und mit Parametern versehen. Dabei muss ein POST-Befehl mit den variablen Einstellungen versehen an die IP des Servers mit dem SICA-Algorithmus gesendet werden. Diese werden schlussendlich als Konfigurationsdatei im JSON-Format auf Serverseite abgelegt und verwaltet. So können ähnliche Operationen ohne weitere Konfiguration durchgeführt werden.

Nach der Konfiguration können Datenpakete mit den Zeitreihen über die REST-Schnittstelle geschickt werden. Diese können z.B. als Zeitreihenobjekte mit Zeitstempel geschickt werden. Die Datenhaltungsplattform, in diesem Falle Aquarius, muss sicherstellen, dass das Zeitformat in einem POSIX-Format vorliegt und Fehlwerte mit NA gekennzeichnet sind.

Durch die Übermittlung der ersten Datenreihe wird der SICA-Algorithmus angelernt und als trainiertes Objekt auf dem Server abgelegt. Somit reduziert sich der Zeitaufwand bei neu eintreffenden Daten in der Eingruppierung mit bestehenden Clusterzentren auf Sekundenbruchteile. Das initiale Training benötigt auf einem M1-ARM Prozessor weniger als 50s für knapp 3.500 Datensätze.

Die Rückgabe der Ergebnisse erfolgt über einen Zeitstempel an die Aquarius Plattform. Hierbei wird der mehrdimensionale Datenraum auf eine 2D-Zeitreihe mit dem Clusterzentrum als Integer-Wert zurückgegeben, das als Ergebnis in einem beliebigen Dimensionsraum grafisch dargestellt werden kann (siehe Abb. 2).

4. Ergebnisse

Der SICA-Algorithmus ist in der Lage aus der reinen Wasserstandsbeobachtung in den Pegeln P1 – P3 eine Differenzierung durchzuführen. Insgesamt wurden 9 statistisch eindeutige Abflusssituationen über eine Laufzeit von 24 Jahren identifiziert. Dabei treten die Cluster 5 – 7

über die Betrachtungszeit gehäuft auf. Hierbei verhalten sich alle Pegel ähnlich. Eine klassische Mittelwassersituation wurde erkannt.

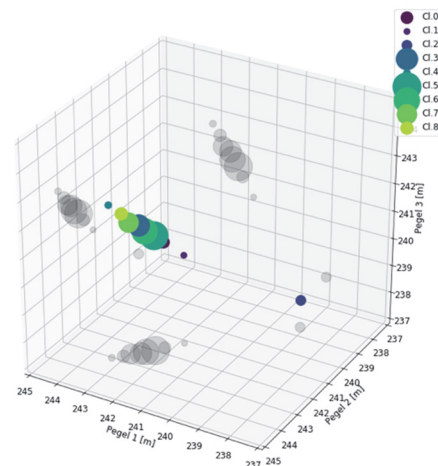


Abbildung 3: Gruppierung der Pegeldata nach einer SICA-Analyse ($\zeta_1 = 0,5$, $\zeta_2 = 10$).

Auffällig ist der Cluster 0 der eine hohe Anzahl kleiner Wasserstände an allen Pegeln identifiziert. Dabei sticht heraus, dass der Übergang zwischen den Clustern 0 und der Häufung zwischen Cluster 5-7 nicht gemessen wird. Hier ist eine Veränderung des Messintervalls an den Pegelgebern hin zu einer Ereignis-basierten Messung empfehlenswert.

Innerhalb der Cluster 5-7 sind Übergänge im Zusammenspiel der Pegel 1 – 3 zu erkennen. Der Cluster 6 hat einen erhöhten Wasserstand am Pegel 3, während die Pegel 1 & 2 noch im unteren mittleren Niveau verharren. Cluster 8 zeichnet sich durch einen stark erhöhten Pegel 1 aus, während Pegel 2 & 3 im Mittelwasser verbleiben. Die Größe dieser Gruppe ist als mittelgroß (ca. 20 Mitglieder) zu bezeichnen.

5. Fazit

Der Einsatz von Datenpipelines im Talsperren-Management ist eine sinnvolle Ergänzung zum Einsatz von Datenhaltungen in Form von Plattformen, um dem wachsenden Berg an Daten Einhalt zu gebieten. Durch die Anbindung des SICA-Algorithmus über eine REST-Schnittstelle kann der Funktionsumfang der Datenhaltung per se massiv und dynamisch erweitert werden, ohne dass grundlegende Erweiterungen an der Verarbeitung von den grundlegenden Messdaten aus dem Bauwerksmonitoring notwendig werden. Der hier gezeigte Ansatz hilft moderne Forschungsansätze schnell und effizient in die wasserwirtschaftliche Praxis zu überführen.

Danksagung / Angaben zu Fördermittelgebern

Die Autoren bedanken sich bei Frau Juliane Neumann (Okeanos Smart Data Solutions) für die Bereitstellung von SICA über eine REST-Schnittstelle.

Kontaktinformationen (optional)

Viktoriastr. 29 44787 Bochum
Dr. Benjamin Mewes
Benjamin.mewes@okeanos.ai

Literaturverzeichnis

- [1] Opiel, H. und Fischer, S., 2020: A New Unsupervised Learning Method to Assess Clusters of Temporal Distribution of Rainfall and their Coherence with Flood Types. *Water Resources Research*, 56, 6, DOI: 10.1029/2019WR026511
- [2] Türksoy, Ö., Mewes, B., Demisch, G., & Last, K. (2023). Datengetriebene Bauwerksanalyse am Baldeneysee. *WASSERWIRTSCHAFT*, 113, 39 - 42.