

# Machine Learning und Messdaten

## Parameterfreie Detektion von Trockenwettertagen und Fremdwasseranalyse

Dominik Leutnant, Malte Henrichs und Mathias Uhl

Institut für Wasser-Ressourcen-Umwelt (IWARU),  
Fachhochschule Münster, Corrensstr. 25, 48149 Münster

**Kurzfassung:** Zur modellgestützten Optimierung des Entwässerungssystems der KA Dortmund-Deusen wurde ein Schmutzfrachtmodell erstellt und anhand von Messdaten kalibriert. Hierzu wurden langjährige Messungen der Drosselabflüsse von Mischwasserbehandlungsanlagen verwendet. Aufgrund der räumlichen Ausdehnung des Einzugsgebiets und der damit verbundenen hohen Anzahl an Mischwasserbehandlungsanlagen waren Methoden zum Umgang mit umfangreichen Datenbeständen erforderlich. Es wurde daher ein Algorithmus entwickelt, der ausschließlich anhand einer Drosselabfluss-Zeitreihe i) Trockenwettertage effizient und präzise detektiert und ii) eine plausible Differenzierung zwischen Schmutz- und Fremdwasseranteilen ermöglicht. Die erzielten Ergebnisse wurden verwendet, um die Modellansätze zur Abbildung von Tages- und Jahresgängen ohne zeit- und rechenintensive Parameterschätzverfahren zu kalibrieren. Die entwickelte Methodik ist unabhängig vom Einzugsgebiet und der Modellanwendung einsetzbar. Sie eignet sich neben dem Preprocessing zur Modellkalibrierung ebenfalls zur Anomalieerkennung in Messdaten.

**Keywords:** Trockenwettertagesgang, machine learning, Modellkalibrierung, Fremdwasseranalyse

## 1 Einleitung

Sowohl der demografische als auch der klimatische Wandel machen die Anpassung bzw. Optimierung bestehender Entwässerungssysteme zu

einer der wichtigsten Aufgabe der Siedlungsentwässerung. Simulationsmodelle zur Abschätzung der Entlastungscharakteristik von Bauwerken zur Misch- und Niederschlagswasserbehandlung dienen hierbei als elementare Werkzeuge, um etwaige Änderungen des Systems analysieren zu können. Voraussetzung dafür ist jedoch ein kalibriertes bzw. realitätsnahes Simulationsmodell.

Je nach Modelltyp und Modellstruktur muss eine Kalibrierung unterschiedliche Prozesse (z. B. Abflussbildung, Abflusskonzentration, Abflusstransport) berücksichtigen. Modellspezifische Ansätze zur Abbildung von frequentativen Abflüssen (z. B. Schmutzwassertagesgang, Fremdwasserjahresgang) sind systemnah zu parametrisieren. Insbesondere bei stark fremdwassergeprägten Einzugsgebieten kann eine differenzierte Betrachtung von Schmutz- und Fremdwasser erforderlich sein, um eine hohe Modellgüte mit realitätsnahen Modellergebnissen zu erhalten. Zwar können einzugsgebietsspezifische Trockenwetterabflussspenden z. B. anhand der angeschlossenen Einwohner und einem einwohnerspezifischen Wasserverbrauch geschätzt werden, jedoch sollten Messdaten zur Ermittlung von Abwassermengen und ihrer tages- und jahreszeitlichen Dynamik erhoben werden (ATV-DVWK-M 165, 2004).

Die Aufbereitung der Messdaten zur Kalibrierung beeinflusst die Modellgüte und ist somit von hoher Relevanz. Insbesondere bei großräumigen Einzugsgebieten mit zahlreichen Mischwasserbehandlungsanlagen (MBWA) kann dieser Vorgang ressourcenintensiv sein und zu subjektiven Ergebnissen führen. Durch eine IT-gestützte prozess- und problemorientierte Messdatenanalyse kann die Kalibrierung der Modellansätze zur Abbildung von Tages- und Jahresgängen stark vereinfacht werden. Dieses „messdatenorientierte“ Vorgehen ermöglicht bereits eine quantitative Differenzierung von Schmutz- und Fremdwasseranteilen im Trockenwetterabfluss.

Der vorliegende Beitrag zeigt daher am Beispiel der Messdatenauswertung des Einzugsgebiets der Kläranlage Dortmund-Deusen, dass (i) eine dem unüberwachten Lernen („unsupervised learning“) zuzuordnende Methodik eine parameterfreie und effiziente Detektion von Trockenwettertagen ermöglicht, (ii) der Trockenwettertagesgang (Median) am Wochenende im Vergleich zum Trockenwettertagesgang (Median) an Werktagen

differiert und (iii) eine einzugsgebietsspezifische plausible Herleitung von Schmutzwassertagesgängen und Fremdwasserjahresgängen erfolgt.

## 2 Material und Methodik

### 2.1 Detektion von Trockenwettertagen

Ein Trockenwettertag ist naturgemäß abhängig vom Niederschlag. In Nordrhein-Westfalen gilt per Verwaltungsvorschrift die Definition, dass an zwei konsekutiven Tagen jeweils die Niederschlagsmenge von 0,3 mm/d nicht überschritten sein darf (MURL NRW, 2001). Basierend auf einer vorliegenden Niederschlagszeitreihe können Trockenwettertage direkt detektiert werden. Bei geringer Datenqualität der hierzu erforderlichen Niederschlagszeitreihe oder in Gebieten mit geringer Niederschlagsstationsdichte kann dieses Vorgehen jedoch zu einer fehlerhaften Zuordnung führen. Dies gilt insbesondere in ungleichmäßig überregneten Gebieten.

Um diesem Umstand zu begegnen und eine möglichst bauwerks- bzw. einzugsgebietsnahe Separation von Trocken- und Regenwettertagen zu ermöglichen, wird ein Machine Learning Algorithmus (ML) zur Mustererkennung von Trockenwetterabflüssen entwickelt und angewandt. Eine integrierte Schätzung des Fremdwasseranteils nach der Nachtminium-Methode (DWA-M 182, 2012) lässt eine Differenzierung zwischen Schmutz- und Fremdwasser zu. Die Programmierung erfolgte in der frei verfügbaren Skriptsprache R (R Core Team, 2014), der Quellcode ist unter [https://github.com/dleutnant/dwf\\_detect](https://github.com/dleutnant/dwf_detect) abrufbar. Das Vorgehen des Algorithmus wird durch die nachfolgenden sechs Schritte beschrieben:

- 1) Zunächst werden alle Trockenwettertage  $T_{TW,i}$  der Messzeitreihe  $Q_{TW}$  mittels „shape-based distance“ (Paparrizos und Gravano, 2015) und hierarchischer Clusteranalyse detektiert.
- 2) Anschließend wird anhand der detektierten  $T_{TW,i}$  der Trockenwettertagessgang im Median  $Q_{TW,median}$  (l/s) berechnet.
- 3) Eine erste Schätzung des monatlichen Fremdwasserabflusses wird basierend auf der Nachtminium-Methode durchgeführt.

- 4) Der tägliche Schmutzwasserabfluss ergibt sich dann als  $Q_s \text{ (l/s)} = \text{MIN}(0, Q_{TW, median} - Q_{FW, Monat})$ .
- 5) Die Anteile von  $Q_s$  und  $Q_{FW}$  werden dann anhand des gemessenen Trockenwetterabflusses mit der Gleichung  $Q_{TW} = a * Q_s + b * Q_{FW}$  numerisch optimiert.
- 6) Es ergeben sich optimierte Schmutzwasserabflüsse  $Q_{s, opt} = a * Q_s$  sowie Fremdwasserabflüsse  $Q_{FW, opt} = b * Q_{FW}$ .

Um die Klassifizierungsgüte des Algorithmus zu bewerten, werden die Trockenwettertage ebenfalls gemäß der Kriterien nach MURL NRW (2001) bestimmt. Beide Ergebnisse werden verglichen und einer manuellen Klassifizierung gegenübergestellt. Grundlage hierfür stellt die Konfusionsmatrix dar, die das Ergebnis der automatischen Klassifizierung und der manuellen Bewertung anhand der vier Zustände „richtig positiv“, „falsch positiv“, „richtig negativ“ und „falsch negativ“ zusammenfasst (s. Tabelle 1). Wird beispielsweise ein Trockenwettertag sowohl manuell als auch durch den Algorithmus als Trockenwettertag klassifiziert, wird dieser „richtig positiv“ gekennzeichnet. Die Konfusionsmatrix stellt zusammengefasst die Klassenstärken von Übereinstimmung und Abweichung dar.

Tabelle 1: Beispiel einer Konfusionsmatrix zur Klassifizierung von Trocken- und Regenwettertagen

		Automatische Klassifizierung	
		Trockenwettertag	Regenwettertag
Manuelle Klassifizierung	Trockenwettertag	richtig positiv	falsch negativ
	Regenwettertag	falsch positiv	richtig negativ

Weiterhin wird die Klassifizierungsgüte anhand der statistischen Gütekriterien i) Sensitivität („Richtig-Positiv-Rate“, Gleichung 1) und Genauigkeit („Präzision“, Gleichung 2) numerisch bewertet. Die Sensitivität beschreibt den korrekt klassifizierten Anteil an Trockenwettertagen im Verhältnis zur Summe der korrekt klassifizierten Trockenwettertage und falsch klassifizierten Regenwettertage. Die Genauigkeit hingegen drückt den Anteil an richtig klassifizierten Trockenwettertagen aller als Trockenwettertag klassifizierten Tage aus.

$$\text{Sensitivität} = \text{richtig positiv} / (\text{richtig positiv} + \text{falsch negativ}) \quad (1)$$

$$\text{Genauigkeit} = \text{richtig positiv} / (\text{richtig positiv} + \text{falsch positiv}) \quad (2)$$

## 2.2 Fallstudie Dortmund-Deusen

Für das Einzugsgebiet der Kläranlage Dortmund-Deusen (Abbildung 1) wurde ein hydrologisches Schmutzfrachtmodell mit der Software MOMENT 9 (BGS Wasser, 2016) zur Systemoptimierung erstellt. Die Einzugsgebietsgröße beträgt 160 km<sup>2</sup> (versiegelte Fläche 41,5 km<sup>2</sup>) und umfasst den östlichen Teil des Emschereinzugsgebiets. Zur Kalibrierung wurden Messdaten (Drosselabflüsse und Wasserstände) über 2 hydrologische Jahre an 12 Mischwasserbehandlungsanlagen verwendet, an denen Fremdwasserjahresgänge und Schmutzwassertagesgänge bestimmt wurden.



Abbildung 1: Einzugsgebiet der Kläranlage Dortmund-Deusen

## 2.3 Schmutz- und Fremdwassermodellierung in MOMENT

Die Parametrisierung der häuslichen und gewerblichen Schmutzwasserabflussanteile erfolgt in MOMENT über die kanalisierte Teilfläche. Es können sowohl Abflüsse als auch Abflussspenden je Teilfläche definiert werden. Mittels Faktoren einer Tagesganglinie lassen sich tageszeitliche Ausprägungen abbilden.

Fremdwasser wird neben einem flächenspezifischen Fremdwasserabfluss über einen niederschlagsbedingten Nachlauf abgebildet. Hierzu wird zunächst ein Teil des auf den durchlässigen Flächen infiltrierten Niederschlags konzeptionell dem Kanal als Fremdwasser zugeführt. Über den Flächenparameter  $a_{QF}$  wird der maßgebliche Flächenanteil des Einzugsgebiets definiert. Zur Beschreibung der Abflusskonzentration des Fremdwasserabflusses wird eine lineare Speicherkaskade verwendet, dessen Retentionskonstante  $k_{QF}$  zu kalibrieren ist. Die Speicherkonstante beträgt  $n = 2$  und ist programmintern fixiert. Jahreszeitliche Schwankungen des flächenspezifischen Fremdwasserabflusses können über Faktoren der Monatsmittelwerte berücksichtigt werden.

## 3 Ergebnisse

### 3.1 Detektion von Trockenwettertagen

Die Güte des entwickelten Klassifizierers wird exemplarisch an Messdaten der MWBA Röhrenstraße (vgl. Abbildung 1) ermittelt. Anhand der dem MWBA-Einzugsgebiet zuzuordnenden Niederschlagszeitreihe werden Trocken- und Regenwettertage gemäß der Kriterien nach MURL NRW (2001) detektiert. Die sowohl manuelle als auch ML-basierte Klassifizierung wird an der Zeitreihe des Drosselabflusses durchgeführt. Tabelle 2 zeigt die Konfusionsmatrix der Klassifizierung, Tabelle 3 die ermittelten statistischen Kenngrößen.

Tabelle 2: Methodenspezifische Konfusionsmatrix der Klassifizierung von Trocken- und Regenwettertagen

		Automatische Klassifizierung MURL NRW (2001)		Automatische Klassifizierung Machine Learning (ML)	
		TW	RW	TW	RW
Manuelle Klassifizierung	TW	248	147	161	234
	RW	8	326	1	333

Tabelle 3: Statistische Kenngrößen der Klassifizierungsgüte der zwei angewandten Methoden

	Sensitivität	Genauigkeit
MURL NRW (2001)	<b>0.63</b>	0.968
Machine Learning (ML)	0.41	<b>0.994</b>

Die manuelle Auswertung der Drosselabflusszeitreihe ergab 395 Trockenwetter- und 334 Regenwettertage.

Die Klassifizierung nach MURL NRW (2001) führte zu 248 richtigen und 147 falschen Zuordnungen von Trockenwettertagen. In 8 Fällen wurden Regenwettertage einem Trockenwettertag zugeordnet. 326 Regenwettertage wurden korrekt erkannt.

Die Anwendung der ML-Methode zeigt eine höhere Genauigkeit (0,994) in der Bestimmung von Trockenwettertagen (161 richtig, 1 falsch). Allerdings werden mehr Trockenwettertage als Regenwettertage interpretiert (234). Dies führt im Vergleich zu MURL NRW (2001) zu einer geringeren Sensitivität.

Durch die geringere Anzahl an richtig klassifizierten Trockenwettertagen („richtig positiv“) verändert sich der zur Modellierung verwendete Trockenwettertagesgang (Median) (s. Abbildung 2). Es wird ersichtlich, dass die Tagesgänge der manuell und der nach MURL NRW (2001) ermittelten Trockenwettertage vergleichbar sind. Die geringen Residuen bestätigen dies. Der Tagesgang der ML-Methode zeigt zwischen 06:00 Uhr und 18:00 Uhr eine maximale Abweichung von ca. 10 % (entspricht 20 l/s).

Der Trockenwetterabfluss steigt hier eher und erzielt einen höheren Spitzenabfluss. Durch ein früheres Abklingen der Abflusswelle ergeben sich lediglich marginale Abweichungen im Abflussvolumen (s. Tabelle 4). Die Ermittlung eines unterschiedlichen Trockenwettertagesgang (Median) ist auf die geringere Anzahl an richtig klassifizierten Trockenwettertagen zurückzuführen. Der Algorithmus kann die natürliche Dynamik im Trockenwetterabfluss zwischen 06:00 Uhr und 18:00 Uhr nur bedingt korrekt interpretieren und klassifiziert kleine Abweichungen gegenüber dem Clusterzentrum bereits als Regenwettereinfluss. Die Trennschärfe zur Clustierung der Trockenwettertage könnte in diesem Fall reduziert werden, um eine größere Schwankung zuzulassen. Dies kann jedoch zu einer verminderten Genauigkeit führen.

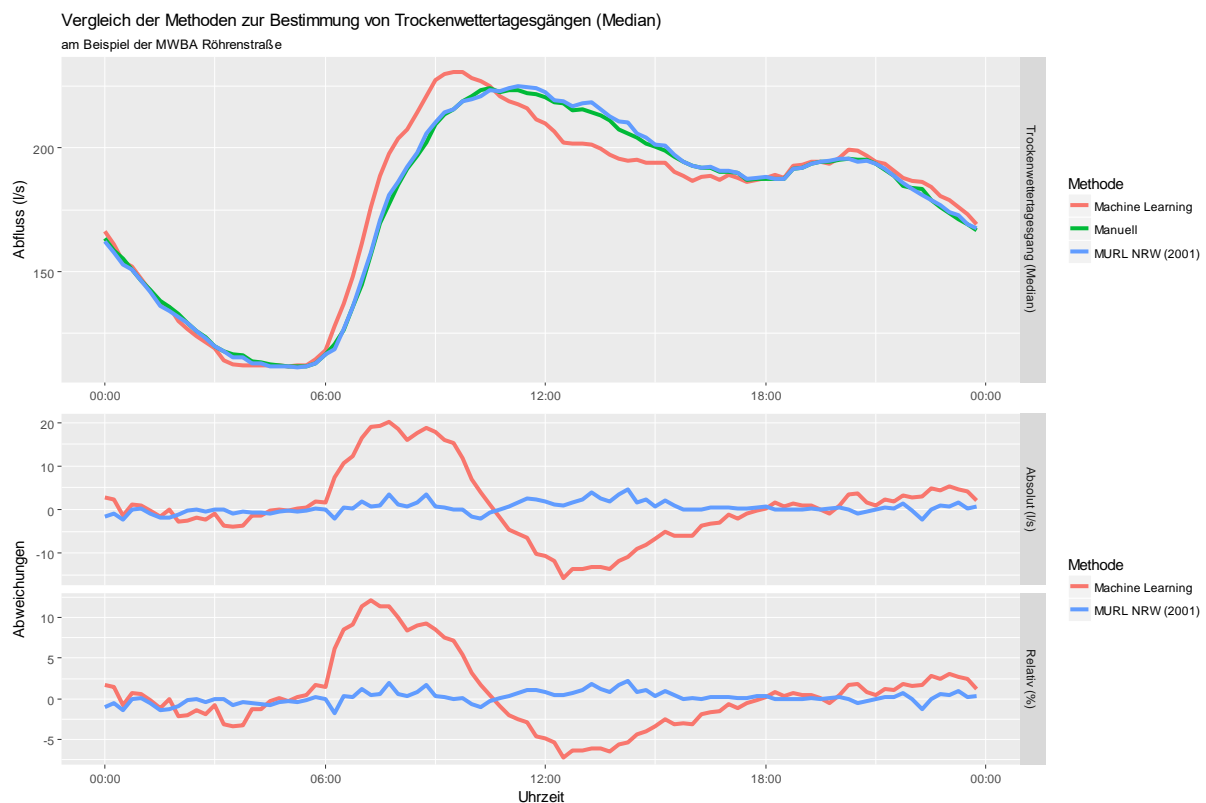


Abbildung 2: Vergleich der sowohl manuell als auch automatisch ermittelten Trockenwettertagesgänge am Beispiel der MWBA Röhrenstraße



Tabelle 4: Abflussvolumen des Median Trockenwetterabflusses (pro Tag)

Manuell	MURL NRW (2001)	Machine Learning
15265 m <sup>3</sup>	15301 m <sup>3</sup> (+0,2 %)	15328 m <sup>3</sup> (+0,4 %)

Die Abbildungen 3 bis 5 zeigen die mit Hilfe der entwickelten ML-Methodik ermittelten Trockenwettertagesgänge (differenziert nach Werk- und Wochenendtagen), Fremdwasserjahresgänge und Schmutzwassertagesgänge an 12 MWBA im Einzugsgebiet der KA Dortmund-Deusen. Die nach Werktag und Wochenende differenzierte Betrachtung des Trockenwetterabflusses in Abbildung 3 zeigt einen am Wochenende um ca. 1h verzögerten Trockenwetterabfluss. Die Darstellung des Fremdwasserjahresgangs in Abbildung 5 zeigt eine Auffälligkeit im Juli an der MWBA Menglinghauser Straße. Hier können extreme Niederschläge im Jahr 2014 als Ursache gelten, die zu einem langanhaltenden Fremdwasseranlauf im Einzugsgebiet führten.

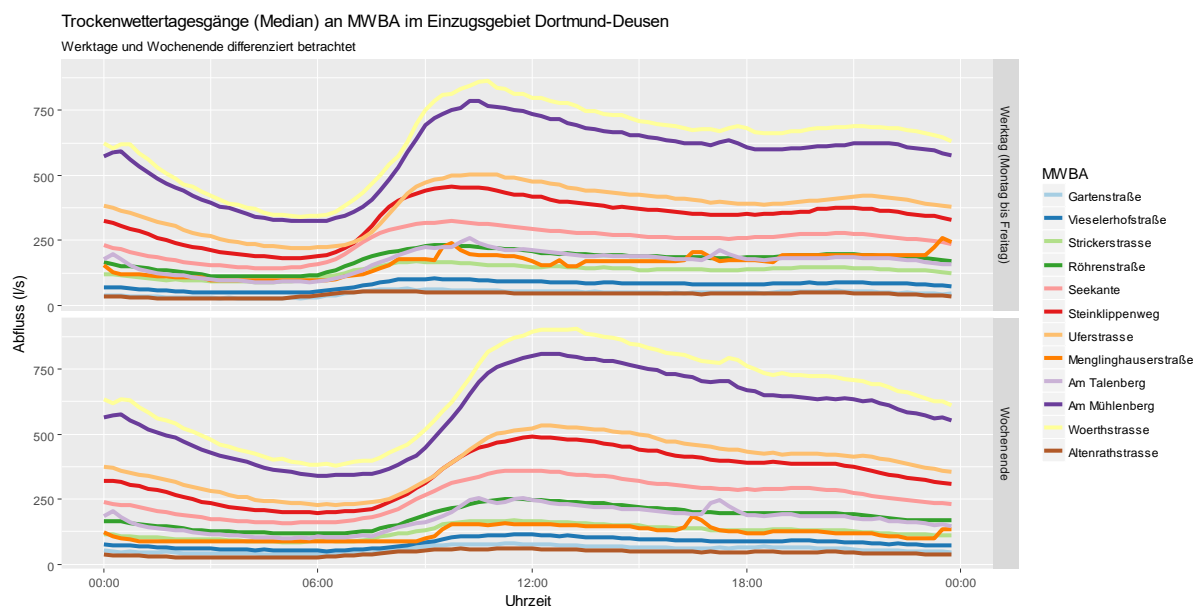


Abbildung 3: Trockenwettertagesgänge an ausgewählten MWBA im Einzugsgebiet Dortmund-Deusen, differenziert nach Werktagen und Wochenende

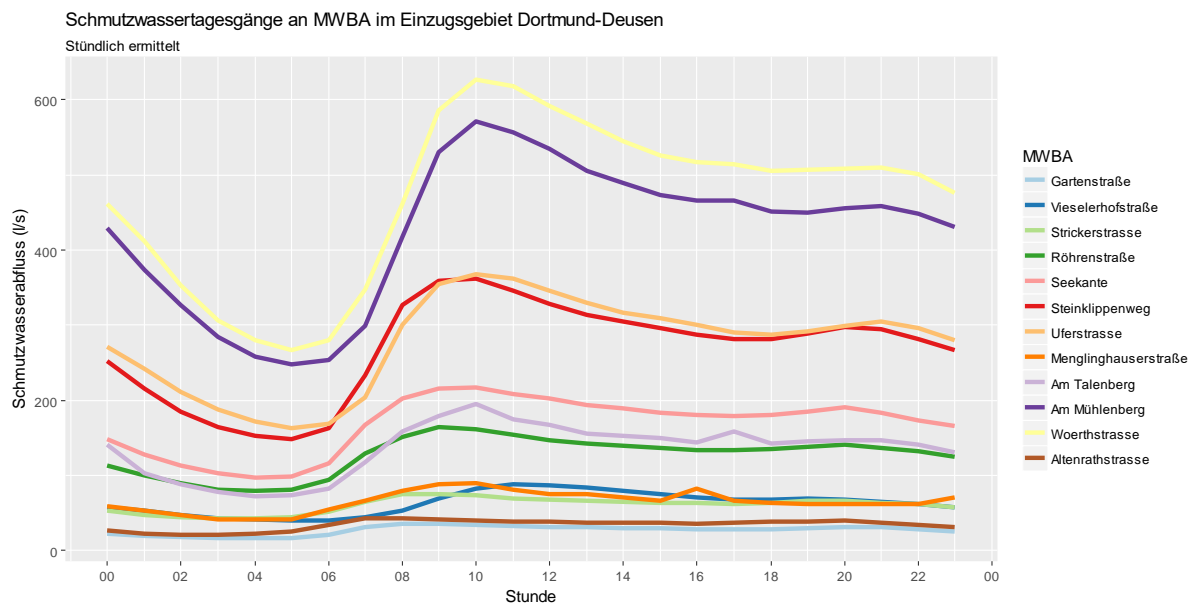


Abbildung 4: Schmutzwassertagesgänge an ausgewählten MWBA im Einzugsgebiet Dortmund-Deusen

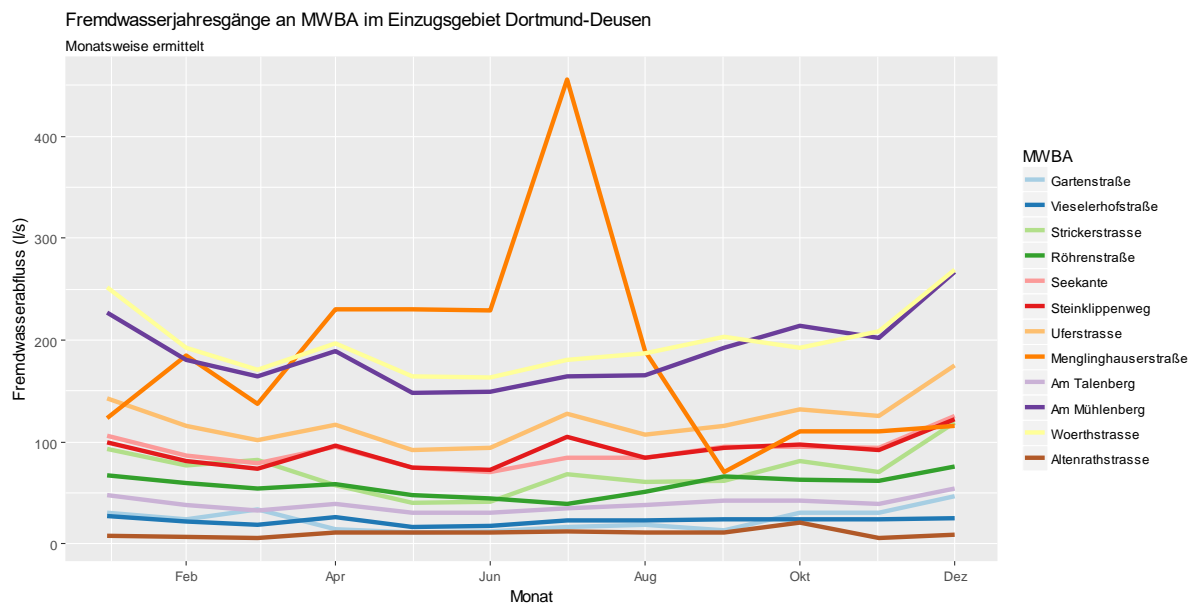


Abbildung 5: Fremdwasserjahresgänge an ausgewählten MWBA im Einzugsgebiet Dortmund-Deusen.

### 3.2 Einzugsgebietspezifische Kalibrierung des Trockenwetterabflusses

Die mittels ML-Methodik ermittelten Tagesgänge und Fremdwasserjahresgänge werden weiterhin verwendet, um die entsprechenden Abflussspenden und Faktoren im jeweiligen Einzugsgebiet zu kalibrieren.

Das Ergebnis der Kalibrierung wird dem unkalibrierten Zustand vergleichend gegenübergestellt. Zur numerischen Bewertung der Kalibrierung werden die Nash-Sutcliffe-Effizienz (Nash und Sutcliffe, 1970) sowie die relative Volumenabweichung verwendet. Tabelle 5 sowie Abbildung 6 stellen exemplarisch die Anpassungsgüte des kalibrierten Modells anhand vier ausgewählter Ereignisse (Jahreszeit berücksichtigt) an der MWBA Röhrenstraße (~ 30 km<sup>2</sup> Einzugsgebiet) dar.

Durch die Kalibrierung kann die Anpassungsgüte des Modells signifikant verbessert werden. Im Mittel beträgt die Nash-Sutcliffe Effizienz der Trockenwetteranpassung 0,65 (vormals: 0,29). Die relative Volumenabweichung liegt zwischen -0,7 % und 7 %. Die geringe Volumenabweichung ist auch anhand der Zeitreihen in Abbildung 6 erkennbar. Allerdings wird hier ebenfalls deutlich, dass das Modell die gemessenen Minima und Maxima des Tagesgangs nicht exakt wiedergeben kann.

Tabelle 5: Abweichungsmaße des kalibrierten und unkalibrierten Trockenwetterabflusses ausgewählter Ereignisse

	Nash-Sutcliffe-Effizienz (-)		Volumenabweichung (%)	
	kalibriert	unkalibriert	kalibriert	unkalibriert
2014-02-04	0.62	0.29	-0,7	4
2014-05-30	0.72	0.43	-2,3	-5
2014-09-04	0.54	0.03	7	-12
2014-12-02	0.89	0.33	-2,5	-2,2

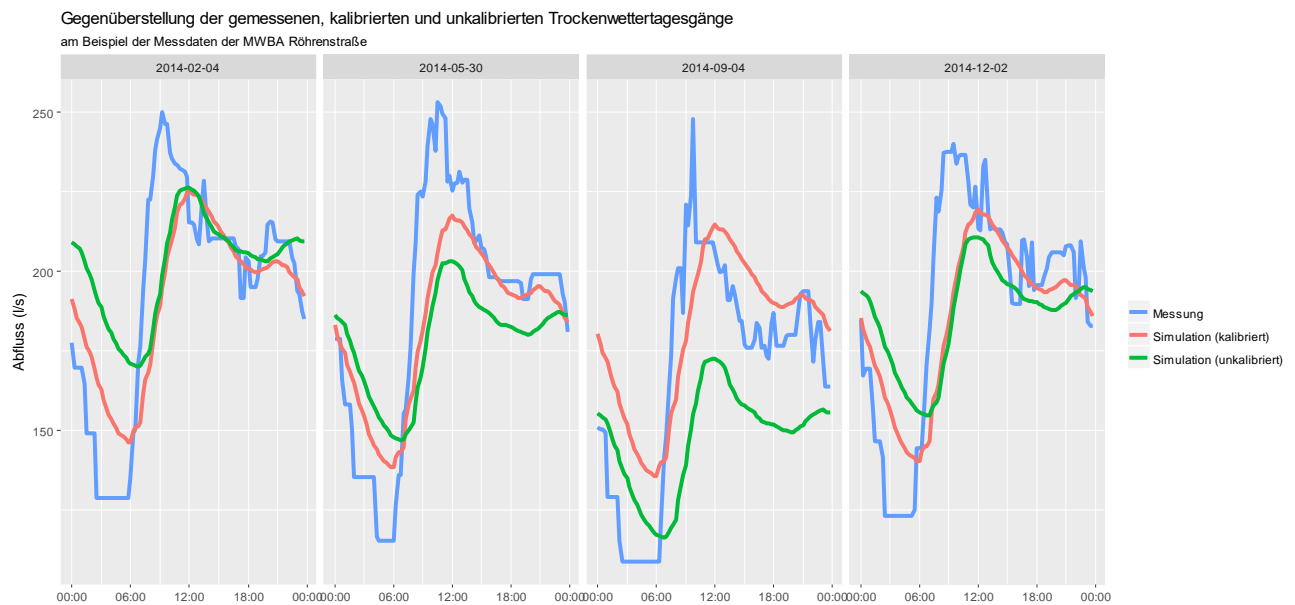


Abbildung 6: Trockenwettertagesgänge ausgewählter Ereignisse (Messung vs. Simulation)

## 4 Zusammenfassung und Ausblick

Im vorliegenden Beitrag werden Messdaten an Mischwasserbehandlungsbauwerken im Einzugsgebiet der Kläranlage Dortmund-Deusen mithilfe statistischer Verfahren automatisiert analysiert und ausgewertet. Es wurde eine für die Praxis robuste Machine Learning (ML) Methodik entwickelt, die eine automatisierte Analyse von Trockenwettertagesgängen vornimmt und in die Abflusskomponenten Fremdwasser und Schmutzwasser aufteilt. Anhand der Fallstudie wurden die mittels ML-Methodik erzielten Tages- und Jahresgänge in einen stochastischen Modellansatz überführt, der zur robusten Kalibrierung des entsprechenden Fremdwasser-Modellbausteins des hydrologischen Schmutzfrachtmodells MOMENT verwendet wurde.

Die ML Methodik bietet gegenüber dem klassischen Vorgehen zur Bestimmung von Trockenwettertagen die Vorteile, i) auch in Einzugsgebieten mit geringer Niederschlagsstationsdichte oder ii) bei unzureichender Qualität der Niederschlagsdaten einsetzbar zu sein, sowie iii) die Ermittlung von Schmutzwassertagesgängen zu unterstützen. Die Repräsentativität der

Methodik hängt maßgeblich von Messdaten ab. Eine wachsende Datenbasis kann somit zur Verbesserung der Methode beitragen.

## 5 Danksagung

Die Arbeiten erfolgten im Rahmen des Projektes „Entwicklung einer Optimierungsstrategie für Mischwassernetze am Beispiel Dortmund Deusen“ im Auftrag der Emschergenossenschaft/Lippeverband. Die Autoren möchten besonders Herrn Dr.-Ing. Jürgen Mang (EG/LV) für die konstruktive Mitarbeit danken.

## 6 Literatur

- ATV-DVWK-M 165 (2004). Anforderungen an Niederschlag-Abfluss-Berechnungen in der Stadtentwässerung. In: ATV-DVWK-Merkblatt 165.
- BGS Wasser (2016). MOMENT 9, Modellierung von Mischwasserentlastungen, Handbuch Version 9. Darmstadt: Brandt Gerdes Sitzmann Wasserwirtschaft GmbH.
- DWA-M 182 (2012). Fremdwasser in Entwässerungssystemen außerhalb von Gebäuden. April 2012, Hennef (Sieg): Deutsche Vereinigung für Wasserwirtschaft, Abwasser und Abfall.
- MURL NRW (2001). Verwaltungsvorschrift zur Ermittlung der Jahresschmutzwassermenge bei Einleitung von mit Niederschlagswasser vermischem Schmutzwasser. In: RdErl. d. Ministeriums für Umwelt, Raumordnung und Landwirtschaft - IV B 6 - 031 003 0101/IV B 5 - 676/5-28728 v. 4.2.1991.
- Nash J.E. und Sutcliffe J.V. (1970). River flow forecasting through conceptual models part I - A discussion of principles. In: Journal of Hydrology, 10(3), S. 282–290.
- Paparrizos J. und Gravano L. (2015). k-shape: Efficient and accurate clustering of time series. In: Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data. , S. 1855–1870: ACM.
- R Core Team (2014). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

**Korrespondenz an:**

Dominik Leutnant, M.Sc.  
Corrensstr. 25, 48149, Münster, Deutschland  
Tel.: +49 251 / 8365274  
Email: [leutnant@fh-muenster.de](mailto:leutnant@fh-muenster.de)  
GIT: <https://github.com/dleutnant>