



# Maschinelles Lernen in der Siedlungswasserwirtschaft

Johannes Sappl · Matthias Harders · Wolfgang Rauch

Online publiziert: 21. Mai 2019  
 © Der/die Autor(en) 2019

**Zusammenfassung** Seit dem Beginn des Siegeszugs von Maschinellern Lernen vor einigen Jahren ist die schiere Anzahl an Möglichkeiten für Anwendungen und die damit verbundene Flut an Publikationen kaum mehr zu durchblicken. Dennoch ist bislang die Menge an realisierten Anwendungen der Technologie in der Praxis überschaubar. In diesem Artikel erläutern wir die theoretischen Grundlagen und verschaffen uns einen Überblick über die Methodik sowie deren Anwendung in der Siedlungswasserwirtschaft.

**Schlüsselwörter** Maschinelles Lernen · Siedlungswasserwirtschaft · Wasserversorgung · Entwässerung · Abwasserreinigung

## Machine Learning in Urban Water Management

**Abstract** Since the beginning of the triumphal march of machine learning a few years ago, it is almost impossible to keep track of the sheer number of possibilities for applications and the flood of publications associated with them. Nevertheless, the number of applications of the technology implemented in practice is still moderate. In this article, we explain the theoretical basics and gain an overview of the methodology and its application in urban water management.

DI J. Sappl (✉) ·  
 Univ.-Prof. DI Dr. W. Rauch  
 Arbeitsbereich Umwelttechnik,  
 Universität Innsbruck,  
 Technikerstraße 13, 6020 Innsbruck,  
 Österreich  
[johannes.sappl@uibk.ac.at](mailto:johannes.sappl@uibk.ac.at)

Univ.-Prof. DI Dr. W. Rauch  
[wolfgang.rauch@uibk.ac.at](mailto:wolfgang.rauch@uibk.ac.at)

Univ.-Prof. Dipl.-Inf. Dr. M. Harders  
 Forschungsgruppe Interaktive  
 Grafik und Simulation, Universität  
 Innsbruck, Technikerstraße  
 21A, 6020 Innsbruck, Österreich  
[matthias.harders@uibk.ac.at](mailto:matthias.harders@uibk.ac.at)

**Keywords** Machine Learning · Urban Water Management · Water Supply · Urban Drainage · Wastewater Treatment

## 1 Einleitung

Im Sinne der Modellierung kann grundsätzlich zwischen den beiden Kategorien der physikalisch basierten Modelle, welche wir in dieser Arbeit gänzlich beiseitelassen, sowie der auf Daten basierenden statistischen Modellierung unterschieden werden. Eine stringente Klassifizierung der Methoden der datenbasierten Modellierung ist de facto nicht möglich, da dies je nach Problemstellung und Fachgebiet divergent definiert wird. Die Methoden umfassen den Bereich der klassischen Statistik (z. B. lineare oder logistische Regression) bis hin zu Artificial Neural Networks (ANN)<sup>1</sup> und Deep Learning (DL)-Methoden, wobei für letztere Verfahren die Bezeichnung Machine Learning (ML) geläufig ist.

Der in der Literatur oft verwendete Begriff Artificial Intelligence (AI) geht über die Modellierung auf Basis von Daten weit hinaus und umfasst auch heuristische Lösungsverfahren zur Entscheidungsfindung wie z. B. Genetische Algorithmen. ML-Verfahren sind daher eine Untergruppe der AI.

ML-Modelle können wiederum anhand der zur Verfügung stehenden Daten kategorisiert werden, an welche sie angepasst werden sollen. Man spricht dabei – je nachdem, ob zu gegebenen Eingangsdaten ein gewünschter Ausgangswert vorliegt oder nicht – von überwachtem bzw. unüberwachtem Lernen. Als gesonderter Fall muss das sogenannte Bestärkende oder Reinforcement Learning (RL) betrachtet werden, bei dem das Modell nur durch Vorgabe einer (negativen) Belohnung nach dem Versuch-und-Irrtum-Prinzip von selbst eine Gewinnstrategie entwickelt, siehe Abb. 1. Im Sinne der Klas-

<sup>1</sup> Wir verwenden ab jetzt die englischen Termini.

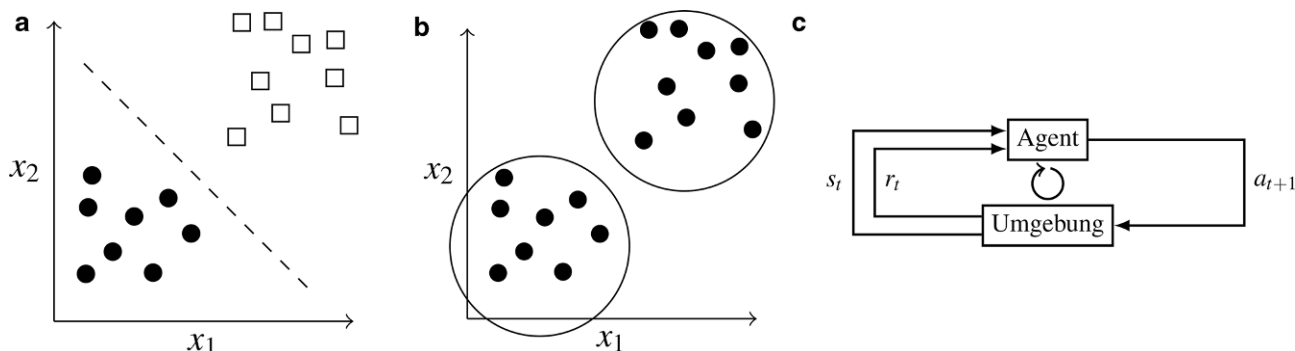
sifizierung ist zu beachten, dass auch statistische Verfahren wie z. B. Clusteranalyse oder Principal Component Analysis (PCA) in der Literatur dem unüberwachtem Lernen zugeordnet werden. Andererseits sind diese Verfahren nicht (zumindest nicht direkt) dem Bereich ML zugehörig.

Lässt man die Google Suchanfragen im Netz zu den englischen Begriffen AI, ML und DL als Gradmesser für die Popularität dieser Thematik gelten, so scheint das Interesse daran nach wie vor ungebrochen zu sein, vgl. Abb. 2. Da heutzutage alles (Un-)Mögliche aufgezeichnet, archiviert und für eventuelle spätere Anwendungszwecke gespeichert wird, wurde die perfekte Ausgangslage für den Einsatz von selbstlernenden Modellen geschaffen. Speziell auch in der Siedlungswasserwirtschaft (SWW) kann mit dem Einsatz von Smart Metern eine große Menge an Daten zur Verfügung gestellt werden. In Souza et al. (2000) wird die Entwicklung von Smarten Städten in Zusammenhang mit ML- und DL-Methoden anhand von 39 relevanten Publikationen untersucht. Drei davon analysieren Techniken zur Prognose von Zeitreihen des urbanen Wasserverbrauchs und sind damit eindeutig der SWW zuzuordnen. Die generellen Möglichkeiten sind noch um einiges vielseitiger und reichen über dezentrale Steuerung von Pumpen in Entwässerungssystemen bis hin zur assistierten Zustandserhebung von Kanalrohren.

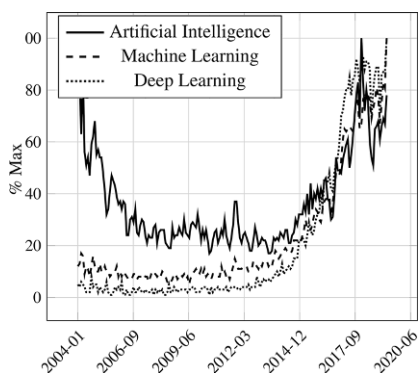
Im Folgenden beschäftigen wir uns primär mit der Methodik des ML und geben einen kurzen Einblick in die historische Entwicklung sowie die dafür notwendigen theoretischen Grundlagen. Weiters führen wir drei Gründe an, weshalb die rasante Entwicklung dieser Methoden in den letzten Jahren, auch im Bereich der SWW, überhaupt erst möglich war.

## 2 Machine Learning – Vom Neuron zum Netzwerk

Der Begriff ML bezieht sich auf computergestützte Methoden mit dem Ziel,



**Abb. 1** Unterscheidung nach Methodik: **a** (Überwachtes Lernen) Bestimme eine trennende Gerade (*gestrichelt*) zwischen den markierten Trainingsdaten (*Kreis, Quadrat*) mit Merkmalen  $(x_1, x_2)$  zur binären Klassifikation. **b** (Unüberwachtes Lernen) Finde Muster in nicht unterscheidbaren Daten mit Merkmalen  $(x_1, x_2)$ . **c** (RL) Finde basierend auf einem Zustand  $s_t$  und einer Belohnung  $r_t$  zum Zeitpunkt  $t$  eine Aktion  $a_{t+1}$  für den nächsten Zeitschritt  $t + 1$  welche  $r_{t+1}$  maximiert und beginne von Neuem



**Abb. 2** Ungebrochener Trend: Die weltweiten Google Suchanfragen seit 2004 in % des Maximums offenbaren einen merklichen Anstieg mit dem Durchbruch von ML und DL im Jahre 2012. Da der AI schon lange bekannte Algorithmen wie PCA zugeordnet werden, war reges Interesse daran bereits schon vor 2012 vorhanden. (<https://trends.google.com/>)

wiederkehrende sinnvolle Muster aus oft sehr umfangreichen Datensätzen zu extrahieren. Im Gegensatz zu konventionellen Ansätzen, bei denen ein menschlicher Programmierer explizit Regeln für das Erkennen solcher Gesetzmäßigkeiten erstellen muss, „erlernt“ ein ML-Modell basierend auf den Daten inhärent eine in gewissem Sinne optimale Vorgehensweise. Es sammelt dabei Erfahrung aus Beobachtungen, den Trainingsdaten, und verwandelt diese mittels Adjustierung seiner internen Parameter in domänenspezifisches Wissen. Seit Anfang der 2000er-Jahre wird ML mit großem Erfolg, z.B. zur Erkennung von Objekten und Segmentierung von Regionen in Bildern, eingesetzt.

## 2.1 Die Natur als Vorbild

Das menschliche Gehirn als Ursprung unserer kognitiven Leistungsfähigkeit ist eine biologische Rechenmaschine bestehend aus 100 Mrd. Neuronen und dem Zehn- bis Fünzigfachen davon an Gliazellen (Herculano-Houzel 2012). Obwohl seine Masse nur 2 % zum Gesamtkörpergewicht beiträgt, verbraucht es doch rund 20 % der vom menschlichen Metabolismus zur Verfügung gestellten Energie.

Das Gehirn ermöglicht es uns, Muster sowie Zusammenhänge zu erkennen und komplexe Aufgabenstellungen zu bearbeiten. Ein Strom verschiedenster äußerer sensorischer Reize wird dabei umgewandelt in konkretes Handeln und Interagieren mit unserer Umwelt. Auch wenn das Gehirn den Menschen zu sehr vielschichtigen Gedankengängen bemächtigt, sind jene Prozesse, die dabei auf mikroskopischem Level ablaufen, erstaunlich rudimentär. Die einzelnen Neuronen leiten entweder Strom weiter oder verbleiben in einem Ruhezustand und codieren damit ein binäres Signal, welches an und für sich noch wenig Möglichkeiten bietet, Informationen zu verwerten. Erst die Komposition vieler dieser elementaren Komponenten macht das Gehirn zu einem ausdrucksstarken Instrument der Datenverarbeitung, das im Zentrum des menschlichen Nervensystems steht.

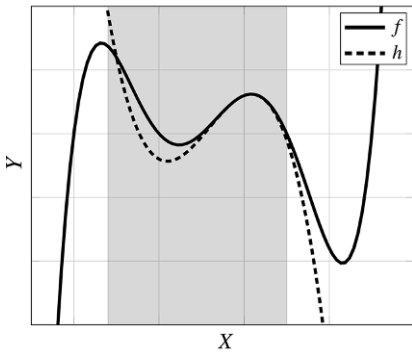
## 2.2 Über Bits und Bytes

Die theoretischen Grundlagen sind für die wichtigsten ML-Verfahren, d.h. ANN, Convolutional Neural Networks (CNN) und Support-Vector Machines (SVM), allgemein gültig. Man beach-

te, dass eine SVM als Spezialfall eines ANNs im Rahmen von überwachtem Lernen betrachtet werden kann.

Am Beginn jedes ML-Problems mit überwachtem Lernen steht ein  $N$ -elementiger Datensatz  $(X, Y)$  der sich zusammensetzt aus Eingangsvariablen  $X = (x_j)_{j=1}^N$  in einem meistens hochdimensionalen Raum und dem gewünschten Resultat  $Y = (y_j)_{j=1}^N$ . Ist  $Y$  nicht bekannt, so spricht man vom, wie schon anfangs erwähnt, unüberwachten Lernen, also das Verarbeiten von nicht markierten Daten, wie z.B. Clusteranalyse mit dem  $k$ -Means-Algorithmus. Optimalerweise ist der Datensatz  $(X, Y)$  von hoher Qualität, d.h. es werden standardisierte Datenformate eingehalten, er enthält keine fehlenden Werte oder Duplikate und die Anzahl an erklärenden Variablen für eine vorgegebene Problemstellung ist ausreichend. Bis zu einem gewissen Grad kann die Methodik ungenügende Datenqualität z.B. aufgrund von Messfehlern kompensieren, jedoch sinkt mit der Güte der Daten auch die Performance des ML-Modells.

Wir treffen die theoretische Annahme, dass eine uns unbekanntes Abbildung  $f: X \rightarrow Y$  existiert, welche jedem Wert  $x_j$  das korrekte Ergebnis  $y_j$  zuordnet, also  $f(x_j) = y_j$ . Das Ziel unseres ML-Algorithmus ist es, diese Abbildung  $f$  auf den uns zur Verfügung stehenden Daten bestmöglich mit einem ANN, CNN oder einer SVM, im Folgenden mit  $h$  bezeichnet, zu approximieren (Shalev-Shwartz, Ben-David 2014). Sobald ein solches Modell jedoch Vorhersagen für noch nie zuvor gesehene Eingangswerte treffen muss, lässt die Güte der Annäherung  $h$  sehr schnell zu wünschen übrig, siehe Abb. 3.



**Abb. 3** Schematische Darstellung von Approximation mit ML-Modellen: Die unbekannte Abbildung  $f$ , welche jedem Wert  $x_j \in X$  das korrekte Ergebnis  $y_j \in Y$  aus dem Datensatz  $(X, Y)$  zuordnet, wird mit einem ML Modell, hier als  $h$  bezeichnet, angenähert. Der schraffierte Bereich kennzeichnet die Verfügbarkeit von Trainingsdaten. Außerhalb dieses Gebiets offenbart sich ein Nachteil jedes datengetriebenen Modells, nämlich der Mangel an Fähigkeit zur Extrapolation, der sich in einem rapide zunehmenden Fehler widerspiegelt. Zusätzlich hängt das Ergebnis der Modellierung sehr stark von der Qualität der ursprünglichen Daten ab

Inspiziert von den in Abschn. 2.1. beschriebenen biologischen Prozessen setzt sich  $h$  aus miteinander verschachtelten linearen Abbildungen zusammen, deren einzelne Komponenten, in Abb. 4 schematisch als Kreise dargestellt, ebenfalls Neuronen genannt werden. Da die Hintereinanderausführung linearer Funktionen wiederum linear

ist, werden Nichtlinearitäten als Aktivierungsfunktionen, in den nachfolgenden Ausführungen als  $\sigma$  bezeichnet, dazwischen geschaltet. In den letzten Jahren hat sich hier die sogenannte Rectified Linear Unit (ReLU) als De-facto-Standard etabliert (LeCun et al. 2015; siehe Abb. 5). Insgesamt lässt sich  $h$  mit  $n-1$  versteckten Schichten schreiben als:

$$h(x_j) = W_n \sigma(W_{n-1}(\dots \sigma(W_1 x_j + b_1) + \dots)) + b_n = \hat{y}_j. \quad (1)$$

Die zu bestimmenden Parameter  $W_1, \dots, W_n$  sowie  $b_1, \dots, b_n$ , auch Gewichte bzw. Bias-Terme genannt, werden zufällig initialisiert und dann anhand der Daten  $(X, Y)$  optimiert. Dazu bedarf es einer Fehler- oder Kostenfunktion  $\ell$ , welche den Abstand zwischen dem Ergebnis des Modells  $\hat{y}_j$  zum tatsächlichen Wert  $y_j$  misst. Das Trainieren von  $h$  bedeutet die empirische Verlustminimierung auf den Daten, genauer:

$$\sum_{j=1}^N \ell(y_j, \hat{y}_j) \rightarrow \min \quad (2)$$

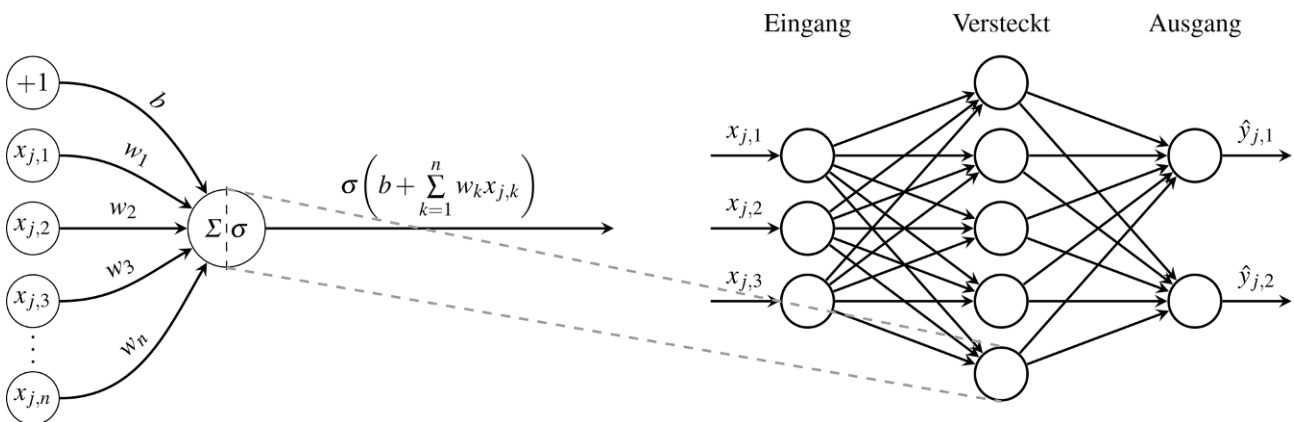
mittels einer Feinabstimmung der Gewichte des Modells. Üblicherweise verwendet man bloß 80% aller Daten zur Bestimmung der  $W_j, b_j$  laut Gl. 2. Auf den verbleibenden 20%, dem Testdatensatz, wird das fertig trainierte Modell auf Überanpassung, also eine un-

erwünschte hohe Genauigkeit auf den Trainingsdaten, getestet.

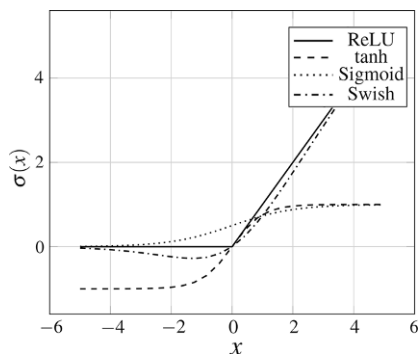
### 2.3 Weitere Ansätze

Wie schon eingangs erwähnt, ist die Klassifizierung der Methoden in diesem Bereich schwierig. Auch werden fallweise AI-Algorithmen fälschlicherweise dem Bereich ML zugeordnet. Die folgende Darstellung soll nur einen kurzen Abriss über diverse Verfahren geben, die in diesem Zusammenhang häufig genannt werden.

In Entscheidungsbäumen wird an jeder Verzweigung ein Attribut der Eingangsdaten mit einem zu optimierenden Grenzwert verglichen, bis ein Endknoten erreicht wird. Dieses Verfahren wurde mittlerweile weitgehend durch ML und DL abgelöst. Heuristische Methoden wie z.B. Genetische Algorithmen, Simulated Annealing und Ameisenalgorithmen verwenden – wie auch ANNs – biologisch (bzw. im Fall von Simulated Annealing physikalisch) basierte Konzepte, sind jedoch nicht dem Bereich ML zuzuordnen. Dies gilt auch für Fallbasiertes Schließen, wo eine neue Problemstellung anhand einer Ansammlung ähnlicher bereits gelöster Probleme aus der Vergangenheit bewältigt wird. Statistische Verfahren wie z.B. PCA aber auch  $k$ -Means-Clustering sind weiterhin in Gebrauch, oft jedoch nur als Vorbearbeitungsschritt für detailliertere Verfahren.



**Abb. 4** Schematische Darstellung eines Neuronalen Netzwerks: Die Eingangsparameter  $x_j$  aus dem Trainingsdatensatz  $X$  werden mit Gewichten  $w_j$  multipliziert, ein Bias-Term  $b$  addiert und auf das Ergebnis dieser affinen Abbildung anschließend punktwise eine Nichtlinearität  $\sigma$  angewendet (*links*). In ein ANN kann prinzipiell eine beliebige Anzahl von versteckten Schichten und Neuronen eingebaut werden (*rechts*). Mit dem einhergehenden Anstieg an Parametern nimmt aber nicht nur die Komplexität des ML-Modells, sondern auch die Schwierigkeit des Optimierungsprozesses zu. Falls das gewünschte Resultat  $y_j \in Y$  bekannt ist, wird das Ergebnis  $\hat{y}_j$  des Modells mit diesem bezüglich einer Kostenfunktion  $\ell$  verglichen



**Abb. 5** Nichtlineare Abbildungen als Aktivierungsfunktion: Da die Hintereinanderausführung zweier linearer Abbildungen wieder linear, ist werden in einem ANN in jedem Neuron punktweise Nichtlinearitäten auf die Daten angewendet. Die Wahl der Funktionen ist vielfältig und erfolgt in Abhängigkeit vom Problem. Die falsche Aktivierungsfunktion kann ebenfalls die Konvergenzgeschwindigkeit des Trainingsprozesses negativ beeinflussen

### 3 Eine Erfolgsgeschichte

Im Wesentlichen waren die folgenden drei Faktoren ausschlaggebend für das Wiederaufgreifen von zumindest in der Literatur bereits bekannten ML-Algorithmen:

#### 3.1 Theorie und Methoden

Erste sehr primitive Formen von ML-Modellen existieren schon seit langem. Als Beispiel sei hier das von Rosenblatt 1957 entwickelte sogenannte Perceptron erwähnt (Rosenblatt 1957). Es trennt Datenpunkte linear mittels eines Schwellenwerts und ermöglicht somit eine Unterteilung in zwei verschiedene Klassen.

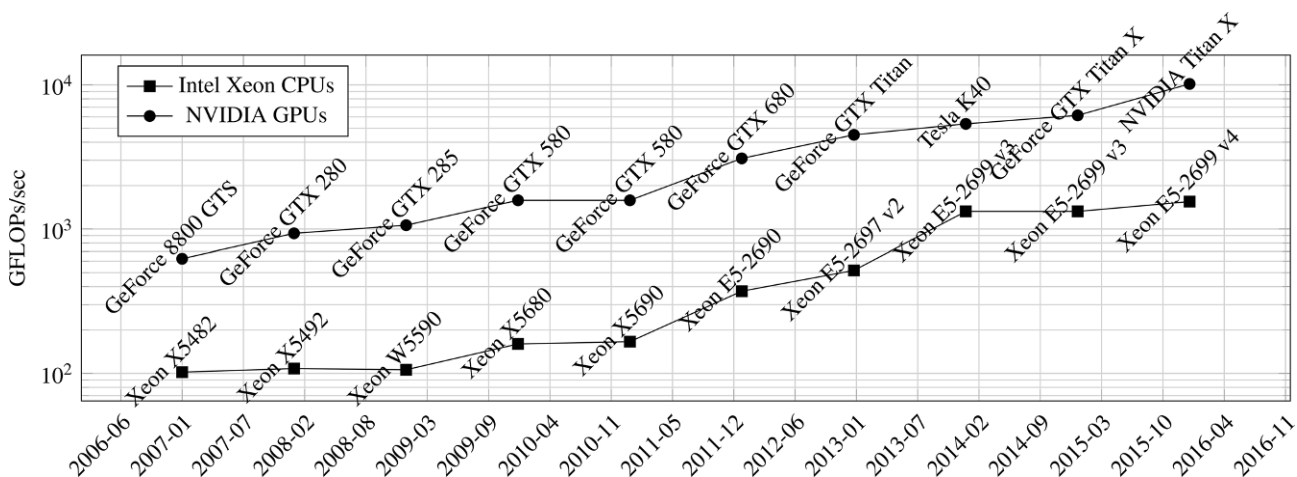
Jenes theoretische Resultat, auf dem im Prinzip alle heutigen ML-Modelle beruhen, wurde 1989 von Cybenko bewiesen (Cybenko 1989). Das „Universal Approximation Theorem“ garantiert, dass eine stetige Funktion beliebig genau mit einer wie in Gl. 1 definierten Abbildung approximiert werden kann. Auch für die Praxis relevante Resultate, wie z.B. der in Lecun et al. (1989) beschriebene Algorithmus, mit dessen Hilfe die Parameter eines ML-Modells über eine Kostenfunktion Gl. 2 optimiert werden, folgten schon sehr rasch. Trotz der fundierten theoretischen Ergebnisse stellte sich der gewünschte Erfolg in der Praxis allerdings nicht ein und das Interesse an ML flaute Ende der 80er-Jahre wieder ab. Das erneute Aufkommen von ML und DL ab dem Jahre 2012 (vgl. Abb. 2) war nur aufgrund der passenden technischen Voraussetzungen möglich.

#### 3.2 Hardware

ML auf einer herkömmlichen Central Processing Unit (CPU) ist aufgrund veralteter Software und dem immensen Rechenaufwand unpraktikabel (Steinkraus et al. 2005), daher ist die Verwendung einer Graphics Processing Unit (GPU) fast zwingend notwendig. Deren spezielle auf parallele Verarbeitung der Daten ausgerichtete Architektur erbringt für rechenintensive Anwendungen, wie z.B. das Optimieren von teilweise Millionen Modellparametern, eine erhebliche Leistungssteigerung im Gegensatz zu CPUs, vgl. dazu auch Abb. 6. Grund dafür ist, dass sich ML-Algorithmen größtenteils auf die Verwendung von simplen Operationen aus der Linearen Algebra wie innere/äußere Produkte, Addition oder Multiplikation von Matrizen und Vektoren sowie Transponieren stützen, die den hohen Grad an Parallelisierung optimal nutzen können.

#### 3.3 Sensoren und Messdaten

Während der durchschnittliche Preis eines Internet of Things (IoT)-Sensors 2004 noch 1,3\$ betrug, verringerte sich dieser Wert bis ins Jahr 2018 auf nur mehr 0,44\$, eine Fortsetzung dieses Trends in die Zukunft scheint sehr wahrscheinlich<sup>2</sup>. Mit dem Preisverfall



**Abb. 6** Entwicklung von CPU/GPU-Hardware: Trotz ähnlicher Steigerungsraten bei der Rechenleistung von CPU/GPU ist z. B. die NVIDIA Titan X um circa eine Größenordnung leistungsstärker als der Xeon E5-2699 v4. Da ML-Algorithmen typischerweise von simplen Matrix-Vektor-Operationen Gebrauch machen, ist die Architektur von Grafikkarten gegenüber jener von herkömmlichen Prozessoren klar im Vorteil. Das Auswerten der ML-Modelle kann mithilfe einer dedizierten GPU um mehr als das Dreifache beschleunigt werden. (Steinkraus et al. 2005; <https://github.com/karlrupp/cpu-gpu-mic-comparison>)

<sup>2</sup> <https://www.ennomotive.com/industrial-iot-sensor-prices/>.



Tab. 1 ML in der SWW, ausgewählte Beispiele

Anwendungsbereich	Datensatz	Methode	Resultat	Referenz
Wasserbedarf	Vier Monate stündliche Messungen von u. a. Wasserverbrauch, Temperatur, Windgeschwindigkeit, Niederschlag; Stadt in Spanien, ca. 5000 EW	Gegenüberstellung ANN, PPR, MARS, SVM, Entscheidungsbäume; evaluiert auf mit Monte Carlo erstelltem Datensatz	Fehlermetrik Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Nash-Sutcliffe; SVM bestes Modell, gefolgt von MARS, PPR	Herrera et al. (2010)
Wasserbedarf	Täglicher Niederschlag, Temperatur, Wasserverbrauch in Montreal, Kanada, 1,8 Mio. EW, 2760 km Leitungen; Daten von 2001–2009	(Nicht-)lineare Regression, gleitender Mittelwert, ANN, Wavelet Analysis und Wavelet ANNs	Transformation der Daten mittels Wavelets vor Anwendung, ANN beste Performance auf Testdatensatz mit $R^2 = 0,919$	Adamowski et al. (2012)
Wasserbedarf	Max. Temperatur, Gesamtniederschlag, Gesamtwasserverbrauch in Calgary, Kanada, 1,1 Mio. EW, 4600 km Leitungen; 2004–2006	Standard ANN, Bootstrap ANN, Wavelet Bootstrap ANN	Wavelet-Transformation verbessert Genauigkeit signifikant; $R^2 = 0,85$ für einmonatige bzw. $R^2 = 0,70$ für zweimonatige Prognose	Tiwari und Adamowski (2015)
Rohrleitungsverschleiß	Wassernetz österreichische Stadt, ca. 95.000 EW; 39.637 Rohre mit 20 Merkmalen; 3743 dokumentierte Ausfälle, Aufzeichnungen seit 1983	Gründliche Vorbehandlung der Daten, geboostete Entscheidungsbäume; Ausfallwahrscheinlichkeit in Fünfjahresschritten	Genauigkeit von 0,96; AUC von 0,93	Winkler et al. (2018)
Echtzeitkontrolle	Fallstudie in Delfland, Niederlande; 40.000 ha, davon 35 % versiegelte Fläche, 120 Pumpstationen; Niederschlagsdaten von 1994–1998	ANN, Fuzzy Adaptive System	Getestet für drei Jahre künstliche Niederschlagsereignisse in Delfland, Korrelationskoeffizient 0,959 mit Aquarius-Software	Lobbrecht und Solomatine (2002)
Kanalrohrinspektion	Taichung, Taiwan; 291 Bilder von Kanalinspektion; vier Klassen von Mängeln	Wavelet Transformation, Grauwertematrix, anschließend ANN, SVM	ANN-Genauigkeit max. 54,6 %; SVM-Genauigkeit max. 60,0 %	Yang und Su (2008)
Kanalrohrinspektion	Regina, Calgary, Kanada; CCTV-Material, 320 × 240 Pixel; Detektion von Wurzeleintritt	Vorbehandlung mit Histogramm of Oriented Gradients-Methode, SVM ohne/mit radiale/n Basisfunktionen	97,8%/86,0 % Genauigkeit auf Trainings- bzw. Testdaten	Halfawy und Hengmeechai (2014)
Kanalrohrinspektion	Georgia, US; 1260 Bilder aus CCTV-Material, Auflösung von 320 × 256 bis 1440 × 720 Pixel; vier Klassen von Mängeln mit Subtypen	CNN detektiert interessante Regionen und erstellt Vektor mit Merkmalen mit konstanter Länge, ANN klassifiziert	Zwei Experimente; max. 83 % gemittelte Genauigkeit, Infiltration am besten klassifiziert	Cheng und Wang (2018)
Emulation	Luzern, Schweiz; künstlich generierte Daten, SWMM-Modelle von zwei verschiedenen städtischen Einzugsgebieten	Singulärwertzerlegung, nicht-negative Matrix-Faktorisierung, Gaußprozess-Emulation	Gemittelter Fehler zwischen 6 und 7,9 % für Adliswil-Einzugsgebiet, Berechnung um Faktor bis zu $5 \times 10^4$ beschleunigt	Carbajal et al. (2017)
Regenabfluss	Xinfengjiang-Stausee, China; täglicher Niederschlag und Abfluss von 2000–2014; 12 Jahre Trainings-, Rest Testdaten	Eine Extreme Learning Machine (ELM), erweitert um Verfahren zur intelligenteren Bestimmung der initialen Parameter	Abhängig von Komplexität des Modells Korrelationskoeffizient $R^2$ zwischen 0,827 und 0,892 auf Testdatensatz	Niu (2018)
ARA-Steuerung	527 Tage an ARA-Daten mit 38 verschiedenen Parametern; frei zum Download verfügbar unter Dua und Graff (2017)	Fallbasiertes Schließen	Getestet auf 15 von Experten gewählten Szenarien, Wertung von 0–10 (am besten), durchschnittliches Resultat 8,8	Sanchez-Marre und Cortes (1997)
Modellprädiktive Regelung	Reale Daten der Wuquiangxi-Papiermühle in Yuanling, China; sechs Inputs, chemischer/biologischer Sauerstoffbedarf als Output	ANN optimiert mit Gradientenabstieg	Nichtlinearer Zusammenhang Schmutzentfernung, chemische Zugaben gelernt; ANN hält gesetzliche Schadstoffkonzentrationen ein	Zeng et al. (2003)
Biogasproduktion	Betriebsparameter der Russafah-Biogasanlage, Jordanien aus 2003; u. a. pH-Wert, TS, TSS	ANN mit zwei versteckten Schichten, Sigmoid-Aktivierungsfunktion	Korrelation echter Daten mit Modell $R^2 = 0,87$ ; Bestimmung besserer Betriebsparameter, dadurch Steigerung Methanproduktion 6,9 %	Qdais et al. (2010)

Tab. 1 (Fortsetzung)

Anwendungsbereich	Datensatz	Methode	Resultat	Referenz
Abwasserreinigung	Generiert mit ASM3 sowie EA-WAG Bio-P	SVM	Max. Fehler COD, Ammonium, TN, TP-Konzentration im Abwasser 20 %; anoxische Tankgröße reduziert, Rezirkulation verringert	Fang et al. (2011)
ARA-Steuerung	BSM1; frei zum Download verfügbar unter <a href="http://apps.ensic.inpl-nancy.fr/benchmarkWWTP/">http://apps.ensic.inpl-nancy.fr/benchmarkWWTP/</a>	Verstärkendes Lernen; Modell bestimmt Sollwert für gelösten Sauerstoff im Belebtschlammverfahren abhängig von Kosten Energie/Bußgelder	KI-Modell übertrifft BSM1 Kontrollstrategie	Hernandez-Del-Olmo et al. (2012)

setzt auch gleichzeitig die gesteigerte Verfügbarkeit und damit ein umfangreicheres Sammeln von Messdaten jeglicher Art ein. Wie z. B. in Reinecke et al. (2018) beschrieben, werden in einer Abwasserreinigungsanlage (ARA) inzwischen hydrodynamisch relevante Prozessparameter wie Geschwindigkeit der Strömung oder der Luftblasen und sogar Blasengrößenverteilungen im Belebungsbecken erhoben. Damit konnten in weiterer Folge signifikante Energieeinsparungen von mehr als 20 % gegenüber dem Referenzbetrieb der ARA erreicht werden. Diese großen Mengen an Messdaten müssen auch archiviert werden. Die Kosten für Festplattenspeicher belaufen sich aber mittlerweile auf weniger als 0,03 \$ pro Gigabyte<sup>3</sup>.

#### 4 Metastudien

Hadjimichael et al. (2016) verschaffen sich anhand der Abstrakt- und Zitierungsdatenbank Scopus von Elsevier einen Eindruck von der Beliebtheit des Themas AI, hauptsächlich ML, im akademischen Bereich der SWW. Für die beiden Themengebiete Wasserversorgung und Abwasserbehandlung konnte ein Wandel von herkömmlichen Clustering-Methoden, verwendet in den späten 60er-Jahren, hin zu ANNs, der heute am weitesten verbreiteten Methodik, beobachtet werden. Allerdings haben viele ML-Techniken den Sprung in die Praxis oft aufgrund der dafür nötigen interdisziplinären Erfahrung noch nicht geschafft und bleiben akademische Gedankenexperimente.

In Ghalekhondabi et al. (2017) werden in den Jahren 2005–2015 veröffentlichte Arbeiten zur Wasserbedarfsprognose untersucht. Viele davon be-

stätigen die verbesserten Vorhersagen, wenn auch in den meisten Fällen nur für kurzfristige Zeiträume, mit innovativen ML-Modellen im Vergleich zu den klassischen. Interessanterweise wurden sehr moderne Methoden wie DL oder unüberwachtes Lernen noch nicht getestet.

Auch im Bereich der Vorhersage von Überflutungsereignissen stützen sich viele Modelle schon auf ANNs, Perceptrons, SVMs oder Entscheidungsbäume, die Anzahl an Publikationen dazu wächst beständig (Mosavi et al. 2018). Ein zu beobachtender Trend ist die Zunahme von hybriden Modellen, z. B. wenn Eingangsdaten für ein ANN zuerst mittels Wavelets transformiert werden, um die Genauigkeit der Vorhersagen, die Fähigkeit zur Extrapolation oder die Berechnungsgeschwindigkeit zu erhöhen. Die Performance der in Mosavi et al. (2018) behandelten Arbeiten wird anhand von Summe der Fehlerquadrate sowie dem Determinationskoeffizient miteinander verglichen.

Eine äußerst interessante historische Betrachtungsweise von ML im Abwasserbereich ist in Corominas et al. (2018) zu finden. Als Beweggründe hinter der Verwendung von datengetriebenen Modellen in einer ARA werden einerseits der Wunsch nach Kontrolle über die biochemischen Verfahren sowie eine daraus resultierende Stabilität der Abbauprozesse, andererseits das Profitieren aus von bis dato ungenutzten Datenfriedhöfen genannt. Mithilfe der Verhältnisgröße von Publikationen zu Zitaten wird eine Gartner-Hype-Zyklus-Grafik erstellt, laut der wir uns nach dem „Tal der Enttäuschungen“ und dem anschließenden „Pfad der Erleuchtung“ gerade auf dem „Plateau der Produktivität“ befinden. Wie auch in Ghalekhondabi et al. (2017) erwähnt, finden nur sehr wenige theoretische Arbeiten, nämlich rund 9 %, wirklich Anwendung in der Praxis. Der

Rest wurde entweder nicht unter realen Bedingungen getestet, oder nur auf einem synthetischen Datensatz validiert.

Von allgemeinem Interesse sind Eggimann et al. (2017) und Song et al. (2018). Es werden Chancen sowie Risiken von datenbasierten Ansätzen und Big Data in den einzelnen Teilbereichen der SWW aufgelistet.

Die nachfolgende Auflistung von Publikationen betreffend ML in der SWW erhebt keinen Anspruch auf Vollständigkeit. Es wurde lediglich der Versuch unternommen, interessante Fallbeispiele für den praktischen Anwender herauszufiltern. Eine Übersicht ausgewählter Arbeiten ist in Tab. 1 zu finden.

#### 5 Wasserversorgung

Einen ausführlichen Gesamtüberblick über die Fortschritte der letzten drei Jahrzehnte im Bereich der Wasserbedarfsprognose verschaffen sich House-Peters und Chang (2011). Es erfolgt eine grobe Unterteilung der Methoden basierend auf zeitlicher und räumlicher Skalierung der Daten, Behandlung von Unsicherheiten der Eingangsvariablen, der Nichtlinearität des Modells, sowie ob ein dynamischer Ansatz gewählt wurde.

Der Verbrauch an Grundwasser nimmt in vielen mediterranen Bereichen bereits sehr hohe Ausmaße an, weshalb Versorger immer mehr unter Druck stehen, den wachsenden Bedarf zu decken. Eine akkurate Prognose des stündlichen Wasserverbrauchs in städtischen Gebieten ist deshalb notwendig, um die Versorgungssicherheit gewährleisten zu können (Herrera et al. 2010). Die Autoren dieser Arbeit stellen herkömmliche ANNs weiteren statistischen Methoden wie Projection Pursuit Regression (PPR), Multivariate Adaptive Regression Splines (MARS), SVM und Entscheidungsbäumen gegenüber. Die Vorhersagegenauigkeit wird anhand

<sup>3</sup> <https://www.backblaze.com/blog/hard-drive-cost-per-gigabyte/>.

von Zeitserien für den stündlich gemessenen Wasserbedarf einer spanischen Stadt mit 5000 Einwohnern über eine Periode von vier Monaten validiert. Besonders klimatische Parameter, wie z. B. Temperatur und Niederschlag, scheinen einen dynamischen Effekt auf den stündlichen Wasserverbrauch zu haben und werden als solche in die Modelle miteinbezogen. Erstaunlicherweise übertrifft – für diese Problemstellung – eine SVM die Genauigkeit des sehr viel komplexeren ANN-Modells.

In Adamowski et al. (2012) wird mit linearer sowie nichtlinearer Regression und verschiedenen ANNs der kurzfristige Wasserbedarf prognostiziert. Der Datensatz umfasst tägliche Messwerte von Niederschlag, Temperatur und Verbrauch auf acht Jahre in Montreal, Kanada. Aufgrund des maroden Zustands des Leitungsnetzes sind Prognosen von besonderer Bedeutung, weil sie es den Planern ermöglichen, ein besseres Verständnis der Dynamik und der zugrunde liegenden Faktoren der Wassernutzung zu entwickeln. Weiters können Wartungs- und Betriebszeiten der Infrastruktur optimiert, Vorteile und Kosten von Wasserspartechnologien analysiert sowie Spitzen im Verbrauch vorhergesagt werden.

Mit Prognosen für einen längerfristigen Zeitraum von einer Woche und ein bis zwei Monaten befassen sich Tiwari und Adamowski (2015). Als erschwerender Faktor kommt hier der limitierte Datensatz von weniger als drei Jahren hinzu. Abhilfe schafft die sogenannte Bootstrap-Methode, bei der zunächst die Modelle auf endlich vielen zufällig ausgewählten Teilmengen der gesamten Daten trainiert werden. Der gemittelte Wert der Prognosen aller ANNs liefert anschließend trotz geringer Anzahl an Observationen bessere Ergebnisse. Darüber hinaus kann mit den Realisierungen der einzelnen schwächeren Modelle ein Konfidenzband für Punktschätzungen erzeugt werden.

Auch Entscheidungen zur Instandhaltung des Rohrleitungssystems können mit Unterstützung von ML- bzw. AI-Methoden getroffen und verbessert werden, so geschehen für eine mittelgroße österreichische Stadt in Winkler et al. (2018). Nach einer gründlichen Vorbehandlung der Daten wurden mehrere Entscheidungsbäume trainiert und deren gemittelttes Ergebnis als Ausfallswahrscheinlichkeit einzelner Abschnitte im Rohrnetz interpretiert. Im Gegensatz zu Blackbox-Modellen wie ANNs

bieten Bäume weiterhin die Möglichkeit zur Interpretation der Ergebnisse, da diese nur von linear getrennten einzelnen Merkmalen, wie z. B. Alter, Material oder Rohrlänge abhängen. Mit derselben Methodik wird in Kumar et al. (2018) die Wahrscheinlichkeit eines Rohrbruchs für mehr als 800 km Leitungen in Syracuse, U. S. prognostiziert. Aufgrund sinkender Einwohnerzahlen kann die Stadtverwaltung das mittlerweile überdimensionierte Leitungsnetz nicht mehr instandhalten. Zum Zeitpunkt der Veröffentlichung dieser Arbeit waren bereits 42 der Top 52 vom Modell vorhergesagten problematischen Häuserblöcke von Schäden betroffen.

Eine im großen Stil angelegte Studie untersucht Trinkwasserversorgungsmuster in 627 chinesischen Städten über einen Zeitraum von 12 Jahren (De Clercq et al. 2018). In die Analyse fließen dabei gewichtete Merkmale wie Zuverlässigkeit, Wasserqualität, Effizienz und Verbrauch mit ein. Zusätzlich werden die Städte anhand eines mit PCA entwickelten Nachhaltigkeitsindex verglichen und gereiht.

## 6 Entwässerung

Eine typische Problemstellung in der Siedlungsentwässerung ist die Echtzeitsteuerung der Anlagen, welche grundsätzlich ein Optimierungs- bzw. Entscheidungsfindungsproblem darstellt. Obwohl die verwendeten Verfahren nicht stringent dem Bereich ML, sondern dem weiter gefassten Begriff AI zuzuordnen sind, wird hier exemplarisch ein Beispiel vorgestellt: Die in den Niederlanden verwendete proprietäre Software Aquarius<sup>4</sup> bietet Möglichkeiten sowohl zur lokalen Steuerung von regulierenden Strukturen, wie z. B. Pumpstationen oder Wehre, oder einer zentralisierten dynamischen Realtime Control (RTC). Für letzteres bedarf es aktueller Messdaten von verschiedenen Knotenpunkten des Abflusssystems, um anschließend die beste Kontrollstrategie als Lösung eines Optimierungsproblems formulieren zu können (Lobbrecht und Solomatine 2002). Einziger Nachteil ist die Dimensionierung des Problems für komplexe Wassersysteme, wo die Anzahl an freien Variablen schnell mehrere 10.000 betragen kann. Hinzu kommen umfangreiche Neben-

bedingungen, die sich aus den technisch machbaren Grenzen der Steuerungselemente ergeben. Das Finden einer Lösung kann hier bis zu 10 min dauern, für RTC nicht akzeptabel. Deshalb wird in Lobbrecht und Solomatine (2002) ein AI-Modell trainiert, welches nur lokale Informationen verarbeitet, aber die globale Kontrollstrategie der Aquarius-Software zu imitieren versucht. Die AI repliziert anschließend zentralisiertes Verhalten auf lokaler Ebene in Echtzeit, ohne den Gesamtzustand des Systems zu kennen.

Inspektionen von Abwasserrohren erfolgen meist unter Zuhilfenahme von Closed-Circuit Television (CCTV) Systemen, d. h. Videoüberwachung. Das Auswerten etlicher Stunden Videomaterials mittels händischer Selektion ist eine ermüdende Tätigkeit sowie abhängig von subjektivem Empfinden. Mit ML-Methoden zur Bildverarbeitung kann diese Arbeit größtenteils automatisiert werden. In Halfawy und Hengmeechai (2014) und Yang und Su (2008) werden einzelne Frames zuerst manuell bearbeitet, bevor ein ANN und eine SVM auf Erkennung von z. B. Rissen oder Brüchen trainiert wird. Die modernere Vorgehensweise ist die Verwendung eines CNN (Cheng und Wang 2018), die besonders für Bildanalysen geeignet sind. Völlig autonom werden von diesem Modell zuerst interessante Bereiche auf den Frames markiert, bevor eine Klassifizierung der Mängel erfolgt.

Einen gänzlich anderen Ansatz verfolgen Carbajal et al. (2017). Modellbasierte Aufgaben wie Sensitivitätsanalysen, Optimierung und RTC erfordern viele kostspielige Auswertungen. Schnelle Emulatoren bieten eine Lösung für diese Effizienzanforderung und tauschen unnötige Genauigkeit gegen erhöhte Geschwindigkeit. Zu diesem Zweck wird in Carbajal et al. (2017) eine komplette Simulation mit dem physikalischen Modell SWMM (Gironás et al. 2010) mit datengetriebenen AI-Modellen emuliert. Der prozentuale Fehler bewegt sich zwischen 6 und 7,9% bei gleichzeitig bis zu mehr als  $5 \times 10^4$ -facher Beschleunigung.

Der in Abschn. 2.2. beschriebene Optimierungsprozess der Parameter eines ML-Modells kann eventuell in lokalen Minima hängenbleiben und nicht gegen ein globales Optimum konvergieren. Deshalb wird in Niu (2018) eine AI verwendet, welche eine optimale analytische Lösung im Sinne der geringsten Fehlerquadrate besitzt. Damit prognostiziert

<sup>4</sup> <https://aquaticinformatics.com/>.

stizieren die Autoren den täglichen Abfluss eines chinesischen Stausees auf einem 15 Jahre umfassenden Datensatz, welcher auch Niederschlagsereignisse beinhaltet.

## 7 Abwasserbehandlung

Da heutzutage nahezu alle Prozesse einer ARA gemessen und aufgezeichnet werden (Reinecke et al. 2018), scheint die Fülle an verfügbaren Daten wie geschaffen für ML. Bereits 1997 wurde in Sanchez-Marre und Cortes (1997) die Steuerung einer ARA mit fallbasiertem Schließen veröffentlicht. Dazu muss zunächst von einem Expertenteam eine Bibliothek von Zuständen der ARA im normalen Betrieb sowie bekannter Störfälle und deren Behebung erstellt werden. Im weiteren Verlauf übernimmt eine AI basierend auf diesen Aufzeichnungen den laufenden Betrieb, indem der aktuelle Zustand der ARA mit jenen in der Bibliothek verglichen und Parameter dementsprechend adjustiert werden. Tritt eine bis dato noch unbekannte Situation auf, wird diese als eigener Fall in die Bibliothek mitaufgenommen und somit der Erfahrungsschatz der AI ständig ausgebaut.

Zeng et al. (2003) kombinieren jeweils ein ML-Modell, um einerseits den gesamten Prozess der auf Flockung basierenden Reinigung der Abwässer einer Papiermühle abzubilden, andererseits die Steuerung der Dosierung von Zugabestoffen zu kontrollieren. Ziel dabei ist die Optimierung der Menge an notwendigen Gerinnungsmitteln in Hinblick auf Kosten bei gleichzeitigem Einhalten von gesetzlich vorgegebenen Mindeststandards im Abfluss.

Der Gärprozess in einem anaeroben Faulturm kann mit einem ANN modelliert werden (Qdais et al. 2010). Als Eingangsvariablen dienen unter anderem Temperatur sowie pH-Konzentration des Faulschlamm, der Ausgangswert des Modells ist die Menge an produziertem Methangas in  $\text{m}^3$ . So kann bei leichter Adjustierung der Betriebspa-rameter die Gasausbeute beobachtet und optimiert werden, ohne jedoch in den operativen Betrieb der echten Anlage eingreifen zu müssen. In Qdais et al. (2010) konnte somit der pro-

zentuale Anteil an Methangas in einer jordanischen Biogasanlage um circa 7% gesteigert werden.

Gleich mehrere Arbeiten (Alejo et al. 2018; Fang et al. 2011; Guo et al. 2015) befassen sich mit der Prozessoptimierung bei der Entfernung von Stickstoff als umweltschädlichste Verbindung im Abfluss einer ARA. Mit dem Einsatz von ANNs und SVMs konnte somit unter anderem das Volumen des anoxischen Tanks um 11% verkleinert, das interne Rezirkulationsvolumen von 300 auf 250% verringert sowie der Anlagenbetreiber bei zu großen Mengen an Lebensmittelabfällen frühzeitig gewarnt werden.

In Hernandez-Del-Olmo et al. (2012) wird eine AI mit RL darauf trainiert, den Sollwert für den gelösten Sauerstoff im Belebtschlammverfahren optimal zu bestimmen. Die dafür verwendete Kostenfunktion setzt sich als gewichtete Summe von Betriebskosten und Bußgeldern für überschrittene Schadstoffkonzentrationen im Abfluss der ARA zusammen. Etwaige Bedenken zwecks ethischer Vertretbarkeit bzw. Umweltschutz seien an dieser Stelle einstweilen außen vor gelassen, die Resultate sprechen jedenfalls für sich. Eine der am öftesten in ARAs zum Einsatz kommende Kontrollstrategie wurde in puncto Kosteneffizienz übertroffen.

## 8 Conclusio

Digitalisierung ist ein allgegenwärtiges Schlagwort und die Anwendung derartiger Methoden wird für fast alle Bereiche und Prozesse der Industrie vorangetrieben. Digitalisierung ist jedoch ein Überbegriff und geht weit über die Anwendung von AI bzw. noch enger gefasst die Anwendung von ML für die Modellierung hinaus. In der SWW ist Digitalisierung schon lange bekannt und die Sammlung, Analyse und Verwendung von Daten werden generell und mit Erfolg angewendet.

Dem gegenüber ist die Anwendung datengetriebener Modelle in den drei Bereichen Wasserversorgung, Entwässerung und Abwasserbehandlung noch nicht so weit verbreitet. Beispiele aus der Literatur zeigen aber durchwegs positive Effekte bei der Anwendung. Dabei

sind ML-Methoden keinesfalls als Ersatz, sondern bloß als Unterstützung für menschliche Entscheidungsträger anzusehen.

Speziell das verstärkende Lernen bietet viel Potenzial, wie die aktuellsten medienwirksam vermarkteten Erfolge für das Brettspiel Go (Silver et al. 2016) und das von Blizzard Entertainment entwickelte und veröffentlichte Science-Fiction-Echtzeit-Strategie-Videospiel StarCraft II (Vinyals et al. 2017) vermuten lassen. Diese Modelle basieren auf einem Trainingsalgorithmus, welcher von selbst eine Vorgehensweise erlernt, mit der eine vorher definierte Form der Belohnung maximiert wird. Besonders bei Datensätzen, über die wenig bis nichts bekannt ist, sind diese Algorithmen in der Lage, Informationen zu extrahieren. Trotz des immensen Erfolgs von ML bzw. AI im akademischen Sektor sowie ausgewählten Bereichen der Praxis, sind bis zur endgültigen Marktreife dieser Methoden noch einige Hürden zu überwinden. Der Einsatz interdisziplinärer Arbeitsgruppen für ML als Verbindung zwischen den Ingenieurwissenschaften, der Mathematik und der Informatik ist aus unserer Sicht unumgänglich.

**Förderung** Diese Forschung ist Teil des von der Österreichischen Forschungsförderungsgesellschaft (FFG) geförderten Projekts 850738 (bezeichnet SPHAUL).

**Funding** Open access funding provided by University of Innsbruck and Medical University of Innsbruck.

**Open Access** Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden. ■



## Literatur

- Adamowski, J., Fung Chan, H., Prasher, S.O., Ozga-Zielinski, B., Sliusariva, A.: Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. *Water Resources Research* 48(1) (2012). <https://doi.org/10.1029/2010WR009945>. URL <http://doi.wiley.com/10.1029/2010WR009945>
- Alejo, L., Atkinson, J., Guzmán-Fierro, V., Roedel, M.: Effluent composition prediction of a two-stage anaerobic digestion process: machine learning and stoichiometry techniques. *Environmental Science and Pollution Research* 25(21), 21149–21163 (2018). <https://doi.org/10.1007/s11356-018-2224-7>. URL <http://links.springer.com/10.1007/s11356-018-2224-7>
- Carbajal, J.P., Leitão, J.P., Albert, C., Rieckermann, J.: Appraisal of data-driven and mechanistic emulators of nonlinear simulators: The case of hydrodynamic urban drainage models. *Environmental Modelling & Software* 92, 17–27 (2017). <https://doi.org/10.1016/j.envsoft.2017.02.006>. URL <https://www.sciencedirect.com/science/article/pii/S1364815216307964>
- Cheng, J.C., Wang, M.: Automated detection of sewer pipe defects in closed-circuit television images using deep learning techniques. *Automation in Construction* 95, 155–171 (2018). <https://doi.org/10.1016/j.autcon.2018.08.006>. URL <https://www.sciencedirect.com/science/article/pii/S0926580518303273>
- De Clercq, D., Smith, K., Chou, B., Gonzalez, A., Kothapalle, R., Li, C., Dong, X., Liu, S., Wen, Z.: Identification of urban drinking water supply patterns across 627 cities in China based on supervised and unsupervised statistical learning. *Journal of environmental management* 223, 658–667 (2018)
- Corominas, L., Garrido-Baserba, M., Villez, K., Olsson, G., Cortés, U., Poch, M.: Transforming data into knowledge for improved wastewater treatment operation: A critical review of techniques. *Environmental modelling & software* 106, 89–103 (2018)
- Cybenko, G.: Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems* 2(4), 303–314 (1989). <https://doi.org/10.1007/BF02551274>. URL <https://doi.org/10.1007/BF02551274>
- Dua, D., Graff, C.: UCI machine learning repository (2017). URL <https://archive.ics.uci.edu/ml/datasets/water+treatment+plant>
- Eggimann, S., Mutzner, L., Wani, O., Schneider, M.Y., Spuhler, D., Moy de Vitry, M., Beutler, P., Maurer, M.: The potential of knowing more: A review of data-driven urban water management. *Environmental science & technology* 51(5), 2538–2553 (2017)
- Fang, F., Ni, B., Li, W., Sheng, G., Yu, H.: A simulation-based integrated approach to optimize the biological nutrient removal process in a full-scale wastewater treatment plant. *Chemical Engineering Journal* 174(2–3), 635–643 (2011). <https://doi.org/10.1016/j.cej.2011.09.079>. URL <https://www.sciencedirect.com/science/article/pii/S1385894711011521>
- Ghalekhondabi, I., Ardjmand, E., Young, W.A., Weckman, G.R.: Water demand forecasting: review of soft computing methods. *Environmental monitoring and assessment* 189(7), 313 (2017)
- Gironás, J., Roesner, L.A., Rossman, L.A., Davis, J.: A new applications manual for the storm water management model (swmm). *Environmental Modelling & Software* 25(6), 813–814 (2010)
- Guo, H., Jeong, K., Lim, J., Jo, J., Kim, Y.M., pyo Park, J., Kim, J.H., Cho, K.H.: Prediction of effluent concentration in a wastewater treatment plant using machine learning models. *Journal of Environmental Sciences (China)* 32, 90–101 (2015). <https://doi.org/10.1016/j.jes.2015.01.007>. URL <https://www.sciencedirect.com/science/article/pii/S1001074215001278>
- Hadjimichael, A., Comas, J., Corominas, L.: Do machine learning methods used in data mining enhance the potential of decision support systems? A review for the urban water sector. *AI Communications* 29(6), 747–756 (2016)
- Halfawy, M.R., Hengmechai, J.: Automated defect detection in sewer closed circuit television images using histograms of oriented gradients and support vector machine. *Automation in Construction* 38, 1–13 (2014). <https://doi.org/10.1016/j.autcon.2013.10.012>. URL <https://www.sciencedirect.com/science/article/pii/S0926580513001805>
- Herculano-Houzel, S.: The remarkable, yet not extraordinary, human brain as a scaled-up primate brain and its associated cost. *Proceedings of the National Academy of Sciences* 109(Supplement 1), 10661–10668 (2012)
- Hernandez-Del-Olmo, F., Gaudio, E., Nevado, A.: Autonomous adaptive and active tuning up of the dissolved oxygen setpoint in a wastewater treatment plant using reinforcement learning. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews* 42(5), 768–774 (2012). <https://doi.org/10.1109/TSMCC.2011.2162401>. URL <http://ieeexplore.ieee.org/document/6015564/>
- Herrera, M., Torgo, L., Izquierdo, J., Pérez-García, R.: Predictive models for forecasting hourly urban water demand. *Journal of Hydrology* 387(1–2), 141–150 (2010). <https://doi.org/10.1016/j.jhydrol.2010.04.005>. URL <https://www.sciencedirect.com/science/article/pii/S0022169410001861>
- House-Peters, L.A., Chang, H.: Urban water demand modeling: Review of concepts, methods, and organizing principles. *Water Resources Research* 47(5) (2011). <https://doi.org/10.1029/2010WR009624>. URL <http://doi.wiley.com/10.1029/2010WR009624>
- Kumar, A., Rizvi, S.A.A., Brooks, B., Vanderveld, R.A., Wilson, K.H., Kenney, C., Edelstein, S., Finch, A., Maxwell, A., Zuckerbraun, J., et al.: Using Machine Learning to Assess the Risk of and Prevent Water Main Breaks. In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 472–480. ACM (2018). <https://doi.org/10.1145/3219819.3219835>
- Lecun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., Jackel, L.: Backpropagation applied to handwritten zip code recognition. *Neural Computation* 1(4), 541–551 (1989)
- LeCun, Y.A., Bengio, Y., Hinton, G.E.: Deep learning. *Nature* 521(7553), 436–444 (2015). <https://doi.org/10.1038/nature14539>. URL <https://www.nature.com/articles/nature14539>
- Lobrecht, A.H., Solomatine, D.P.: Machine learning in realtime control of water systems. *Urban Water* 4(3), 283–289 (2002). [https://doi.org/10.1016/S1462-0758\(02\)00023-7](https://doi.org/10.1016/S1462-0758(02)00023-7). URL <https://www.sciencedirect.com/science/article/pii/S1462075802000237>
- Mosavi, A., Ozturk, P., Chau, K.w.: Flood prediction using machine learning models: Literature review. *Water* 10(11), 1536 (2018)
- Niu, W.J., Feng, Z.K., Cheng, C.t., Zhou, J.z.: Forecasting Daily Runoff by Extreme Learning Machine Based on Quantum-Behaved Particle Swarm Optimization. *Journal of Hydrologic Engineering* 23(3), 04018002 (2018). [https://doi.org/10.1016/S0370-2693\(01\)00961-3](https://doi.org/10.1016/S0370-2693(01)00961-3). URL [https://ascelibrary.org/doi/10.1061/\(ASCE\)HE.1943-5584.0001625](https://ascelibrary.org/doi/10.1061/(ASCE)HE.1943-5584.0001625)
- Abu Qdais, H., Bani Hani, K., Shatnawi, N.: Modeling and optimization of biogas production from a waste digester using artificial neural network and genetic algorithm. *Resources, Conservation and Recycling* 54(6), 359–363 (2010). <https://doi.org/10.1016/j.resconrec.2009.08.012>. URL <https://www.sciencedirect.com/science/article/pii/S092134490900202X>
- Reinecke, S.E., Höffmann, A.K., Stachowske, M., Hampel, U., Ehrhard, P.: Effizienzsteigerung von Kläranlagen-Vorstellung Maschinelles Lernen in der Siedlungswasserwirtschaft 11 von Mess- und Simulationstechnischen Methoden. *Labor-Fachzeitschrift, GIT* (2018)
- Rosenblatt, F.: *The Perceptron—A Perceiving and Recognizing Automaton*. Cornell Aeronautical Laboratory (1957)
- Sanchez-Marre, M., Cortes, U., R-Roda, I., Poch, M., Lafuente, J.: Learning and Adaptation in Wastewater Treatment Plants Through Case-Based Reasoning. *Computer-Aided Civil and Infrastructure Engineering* 12(4), 251–266 (1997). <https://doi.org/10.1111/0885-9507.00061>. URL <http://doi.wiley.com/10.1111/0885-9507.00061>
- Shalev-Shwartz, S., Ben-David, S.: *Understanding Machine Learning: From Theory to Algorithms*. Cambridge university press (2014)
- Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., et al.: Mastering the game of Go with deep neural networks and tree search. *nature* 529(7587), 484 (2016)
- Song, M.L., Fisher, R., Wang, J.L., Cui, L.B.: Environmental performance evaluation with big data: Theories and methods. *Annals of Operations Research* 270(1–2), 459–472 (2018)
- Souza, J., Francisco, A., Piekarski, C., Prado, G.: Data Mining and Machine Learning to Promote Smart Cities: A Systematic Review from 2000 to 2018. *Sustainability* 11(4), 1077 (2019). <https://doi.org/10.3390/su11041077>. URL <https://doi.org/10.3390/su11041077>
- Steinkraus, D., Buck, I., Simard, P.Y.: Using GPUs for machine learning algorithms. In: *Eighth International Conference on Document Analysis and Recognition (ICDAR'05)*, pp. 1115–1120 Vol. 2 (2005). <https://doi.org/10.1109/ICDAR.2005.251>
- Tiwari, M.K., Adamowski, J.E.: Medium-Term Urban Water Demand Forecasting with Limited Data Using an Ensemble Wavelet—Bootstrap Machine-Learning Approach. *Journal of Water Resources Planning and Management* 141(2), 04014053 (2015). [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000454](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000454)
- Vinyals, O., Ewalds, T., Bartunov, S., Georgiev, P., Vezhnevets, A.S., Yeo, M., Makhzani, A., Küttler, H., Agapiou, J., Schrittwieser, J., et al.: *Starcraft II: A New Challenge for Reinforcement Learning*. arXiv preprint arXiv:1708.04782 (2017)
- Winkler, D., Haltmeier, M., Kleidorfer, M., Rauch, W., Tschekner-Gratl, F.: Pipe failure modelling for water distribution networks using boosted decision trees. *Structure and Infrastructure Engineering* 14(10), 1402–1411 (2018). <https://doi.org/10.1080/15732479.2018.1443145>. URL <https://www.tandfonline.com/doi/full/10.1080/15732479.2018.1443145>
- Yang, M.D., Su, T.C.: Automated diagnosis of sewer pipe defects based on machine learning approaches. *Expert Systems with Appli-*

---

cations 35(3), 1327–1337 (2008). <https://doi.org/10.1016/J.ESWA.2007.08.013>. URL <https://www.sciencedirect.com/science/article/pii/S0957417407003405>

**Zeng, G., Qin, X., He, L., Huang, G., Liu, H., Lin, Y.:** A neural network predictive control system for paper mill wastewater treatment. *Engineering Applications of Artificial Intelligence* 16(2), 121–129 (2003). [https://doi.org/10.1016/S0952-1976\(03\)00058-7](https://doi.org/10.1016/S0952-1976(03)00058-7). URL <https://www.sciencedirect.com/science/article/pii/S0952197603000587#FIG1>

**Hinweis des Verlags** Der Verlag bleibt in Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutsadressen neutral.