

# Datenanalyse und Prozessoptimierung am Beispiel Kläranlagen

Prof. Dr. Thomas Bartz-Beielstein  
thomas.bartz-beielstein@fh-koeln.de, Tel. 02261/8196-6391

Prof. Dr. Wolfgang Konen  
wolfgang.konen@fh-koeln.de, Tel. 02261/8196-6275

## **Abstract:**

Ziel dieses Projekts ist die Erforschung von Prognosemodellen für Füllstände in Kläranlagen aufgrund von Regeneintrag und Bodenbeschaffenheit. Wir vergleichen verschiedene Prognoseverfahren und nutzen die Sequentielle Parameteroptimierung (SPO), um für jedes Prognosemodell in vergleichbarer Weise bestmögliche Parameter zu finden. Mit SPO wird in konsistenter Weise der Einfluss von Parametern analysiert. So kommt es zu einer zielgerichteten Vereinfachung und Verbesserung des Modellentwurfs.

**Ansprechpartner:** Prof. Dr. Thomas Bartz-Beielstein

**Institut:** Institut für Informatik

**Kooperationspartner:** Technische Werke Emmerich am Rhein GmbH, Gelsenwasser AG

## **Zielsetzung**

Die Prognose von Füllständen in Regenüberlaufbecken aufgrund von Regeneintrag und Bodenbeschaffenheit steht im Mittelpunkt unserer Untersuchungen. Böden und Kanalnetze stellen ein komplexes dynamisches System dar. Offensichtlich haben die aktuellen Zustände der Böden einen wichtigen Einfluss auf den Füllstand der Überlaufbecken, da z.B. trockene Böden ein anderes Abflussverhalten zeigen als feuchte oder gar gefrorene Böden. Wichtige Faktoren sind ferner die unterschiedlichen Einflüsse räumlich verschiedener Bodensegmente auf den Füllstand aufgrund von Geländeterrainverläufen oder Bodenbeschaffenheit (Wasserdurchlässigkeit). Hinzu kommen noch weitere Einflussfaktoren wie die Sonnenscheindauer, die landwirtschaftliche Nutzung, Temperatur usw. Im Forschungsprojekt KANNST (KANalNetz-Steuerung) wird die Modellierung und Prognose von Füllstandshöhen in Regenüberlaufbecken Basis einzelner Regenmessungen untersucht [1].



*Abbildung 1: Regenmessfeld*

Der wasserwirtschaftliche Stand der Technik besteht heute meistens darin, dass Kanalnetze noch als unregelmäßige Systeme betrieben werden, bei denen z.B. am Ablauf von Regenüberlaufbecken feste Drosselmengen eingestellt werden [3, 4]. Im Vordergrund einer Weiterentwicklung der Kanalnetzbewirtschaftung steht der Gewässerschutz, da so wenig Schmutzfracht wie irgend möglich aus dem Kanal in die Gewässer gelangen soll. In jüngster Zeit werden die hierfür erforderlichen vernetzten Regelungssysteme im Rahmen des Projektes KANNST [1] unter Nutzung moderner Messtechnik und neuer Verfahren der Computational Intelligence (CI) entwickelt. Ziel ist es, die Kapazität des Kanalnetzes möglichst optimal auszunutzen, um die Entlastungen (Überträge in Gewässer) zu minimieren. Hierfür ist eine vorausschauende Prognose der Füllstände aufgrund von Regeneintrag und Bodenbeschaffenheit unerlässlich. Für die Modellierung von Kanalnetzen und Kläranlagen setzen wir komplexe Simulationsprogramme wie SIMBA oder SWMM ein. Bild 1 zeigt exemplarisch einige Vorrichtungen zur Datenerhebung in KANNST, nämlich ein Niederschlagsmessfeld.

## **Methodenspektrum**

Aus unseren Untersuchungen geht hervor, dass einige CI-Verfahren wie z.B. Neuronale Netze, die auf simulierten oder hinreichend glatten Daten eine sehr gute Performance gezeigt haben [5], auf den intermittierenden Daten der realen Messstellen ohne Datenvorverarbeitung keinen kleinen Prognosefehler erzielten und/oder instabiles Verhalten zeigten. Dies liegt aber unseres Erachtens nicht in den Verfahren an sich, sondern in einer noch fehlenden Datenvorverarbeitung begründet. Um den Einsatzbereich der CI-Verfahren möglichst breit zu gestalten, rückt die Bedeutung einer geeigneten, allerdings noch zu erforschenden Datenvorverarbeitung in den Mittelpunkt unserer weiteren Forschung.

Zusammenfassend konnten wir mit dem systematischen Ansatz der sequentiellen Parameteroptimierung (SPO) [3-5] folgende wichtige Aussagen gewinnen:

- Vergleichbarkeit: Es liegt nicht an ungünstig gewählten Parametern, dass bestimmte Modelle eine schlechtere Performance zeigen. Mit hoher Wahrscheinlichkeit sind diese Modelle dann nicht für Modellierungsaufgaben in der Wasserwirtschaft nicht gut geeignet.
- Performanz: Die manuellen Ergebnisse konnten durch SPO nochmals um 30-40% gesteigert werden.
- Teilautomatisierung: Der oft sehr mühsame und zeitaufwändige Prozess des Parametertuning lässt sich teilweise automatisieren.
- Relevanz: Die Zusatzinformationen, die SPO liefert, erlauben es, die Wichtigkeit einzelner

Parameter und ihrer Wertebereiche abzuschätzen. Daraus erschließen sich oftmals überraschende Einsichten über die Modellierung sowie Ansätze, wie sich möglicherweise ein verbessertes Modell bauen lässt.

Gerade zum letzten Punkt konnten im Rahmen unserer Untersuchungen erste Ansätze gewonnen werden, die wir planen, zukünftig für eine genauere Modellierung zu nutzen. Die Prognosen über längere Beobachtungszeiträume sind noch nicht perfekt, dies kann aber basierend auf nur einem Input auch nicht unbedingt erwartet werden. Es ist geplant, weitere Datenquellen hinzuzuziehen (mehr Regenmessstellen, Temperatur und Luftfeuchte-Daten, Sonnenscheindauer, Tageslänge etc.) und deren Parameter schrittweise durch SPO zu optimieren.

Hierzu werden praxisbewährte Methoden der Computational Intelligence (CI) und des Data Mining am Institut für Informatik der FH Köln gebündelt zum Einsatz gebracht. In Kooperationsprojekten mit Partnern aus Industrie und Wirtschaft werden die Methoden auf Einsetzbarkeit und Leistungsfähigkeit geprüft. Der Einsatz in diesen konkreten Anwendungsfällen ermöglicht es, die Reichweite und die Grenzen verschiedener, oftmals komplexer CI und Data Mining Methoden auch für Praktiker aus Industrie und Wirtschaft gut fassbar darzustellen. Das Institut für Informatik der FH Köln unterstützt Unternehmen beim Einsatz dieser Methoden. Zum gegenwärtigen Themen- und Methodenspektrum gehören

- Evolutionäre Algorithmen
- Genetic Programming
- Neuronale Netze
- Fuzzy Logic
- Entscheidungsbäume und Random Forests
- Modellierung dynamischer Systeme, Echo State Networks
- Experimentelles Design, Versuchsplanung
- Sequentielle Parameteroptimierung
- Standardverfahren des Data Minings
- Hybride Kombinationen der o.g. Methoden

Inwieweit diese Verfahren sinnvoll kombiniert werden können, soll anhand weiterer Forschungen untersucht werden. Weitere aktuelle Forschungsprojekte, die mit CI-Methoden bearbeitet werden, sind z.B. in [1,2,5] beschrieben.

## **Danksagung**

Teile dieser Arbeit wurden durch die FH Köln im Rahmen des anerkannten Forschungsschwerpunktes COSA gefördert. Für viele interessante Diskussionen und die Bereitstellung anwendungsspezifischer Daten danken wir unserem Kollegen Prof. Dr. Michael Bongards und seinem Team, besonders Dipl.-Ing. Tanja Hilmer und Dipl.-Ing. Andreas Stockmann.

## **Literatur**

- [1] Bongards, M., Online-Konzentrationsmessung in Kanalnetzen - Technik und Betriebsergebnisse, Forschungsbericht FH Köln, 2007.
- [2] Westenberger et al., Business Intelligence an Hochschulen, Forschungsbericht FH Köln, 2007.
- [3] Bartz-Beielstein, T.; Parsopoulos, K. E.; Vrahatis, M. N.: Design and analysis of optimization algorithms using computational statistics. Applied Numerical Analysis and Computational Mathematics (ANACM) 1 (2004) 2, S. 413–433.
- [4] Bartz-Beielstein, T.; Lasarczyk, C.; Preuß, M.: Sequential Parameter Optimization. In: Proceedings 2005 Congress on Evolutionary Computation (CEC'05), Edinburgh, Scotland (McKay, B.; et al., Hg.), Bd. 1, S. 773–780. Piscataway NJ: IEEE Press. 2005.
- [5] Bartz-Beielstein, T.: Experimental Research in Evolutionary Computation—The New

Experimentalism. Natural Computing Series. Berlin, Heidelberg, New York: Springer. 2006.

[4] Bartz-Beielstein, T.; Bongards, M.; Claes, C.; Konen, W.; Westenberger, H.: Datenanalyse und Prozessoptimierung für Kanalnetze und Kläranlagen mit CI-Methoden. In: Proc. 17th Workshop Computational Intelligence (Mikut, R.; Reischl, M., Hg.), S. 132–138. Universitätsverlag, Karlsruhe. 2007.

[5] Bartz-Beielstein, Thomas ; Zimmer, Tobias ; Konen, Wolfgang: Parameterselktion für komplexe Modellierungsaufgaben der Wasserwirtschaft – Moderne CI-Verfahren zur Zeitreihenanalyse. In: Mikut, R. (Hrsg.) ; Reischl, M. (Hrsg.): Proc. 18th Workshop Computational Intelligence, Universitätsverlag, Karlsruhe, 2008. – im Druck