

Informatik-Bericht Nr. 2007-7

Schriftenreihe Fachbereich Informatik, Fachhochschule Trier

Effiziente Erstellung und praktischer Einsatz von N-A-Modellen mittels Fuzzy Logic und automatisierter Entwicklungsverfahren

Peter Gemmar, Oliver Gronz, Manfred Stüber

Zusammenfassung

Die Prozesse, die in einem Gebiet zur Abflussbildung beitragen sind hochkomplex und nur teilweise konzeptionell beschreibbar. Vorhandene Modelle bilden diese Prozesse nur unvollständig in unterschiedlichen Abstraktionsgraden mit verschiedenen systembedingten Schwächen nach.

Der Einsatz unkonventioneller Methoden wie der Fuzzy Logic ermöglicht die effiziente Erstellung operationell einsetzbarer N-A-Modelle zur Hochwasservorhersage.

In dem vorliegenden Paper werden die unterschiedlichen Fuzzy-Logic-Ansätze, ihre Vor- und Nachteile in der N-A-Modellbildung und die Modellierung selbst detailliert beschrieben. Ein Ansatz zur nahezu vollständig automatisierten Modellbildung und -Optimierung mithilfe von Messdaten wird erläutert. Abschließend werden die Anwendung des Modells in der Praxis und Testergebnisse präsentiert.

Mit diesem Verfahren wurden bisher etwa 50 Modelle für verschiedene Einzugsgebiete unterschiedlicher Größe erstellt. Diese Modelle befinden sich im operationellen Betrieb.

1. Motivation

Bei der Hochwasservorhersage spielen Niederschlags-Abfluss-Modelle (N-A-Modelle) eine zentrale Rolle. Sie beschreiben den hydrologischen Prozess, der ausgehend vom Niederschlag oder schmelzenden Schnee zum Abfluss in dem Gewässer eines betrachteten Gebiets führt. Je besser und zielgerichteter dieser Prozess beschrieben wird, umso besser werden die Qualität der Hochwasservorhersage und damit der Schutz vor Schäden sein.

Naturgemäß ist der dynamische Prozess der Abflussbildung sehr komplex, die Beziehungen zwischen den Eingangs-, Zustands- und Ausgangsgrößen sind nichtlinear und örtlich verteilt. Durch Aufspaltung in geeignete Teilprozesse auf Basis einer hauptsächlich hydrologisch begründeten Systemstruktur, die Bewegung und Speicherung von Niederschlagswasser in der Fläche und im Gewässer beschreibt, werden meist analytisch deterministische Modelle für die Abflussbildung entwickelt. Da die benutzten und vereinfachten Elementarprozesse nicht axiomatisch zu beschreiben sind, wird eine experimentelle Systemidentifikation des Abflussmodells benötigt. Hierbei werden im Wesentlichen die (ungenauen) Parameter bei fester Systemstruktur geschätzt. Eine adaptive Parameteroptimierung im Betrieb wird oft zur Kalibrierung der herkömmlichen Modelle eingesetzt.

Der Erfolg der Parameteridentifikation hängt davon ab, wie gut die gewählten und meist nicht messbaren Zustandsgrößen mit den inneren Größen des Prozesses in eindeutiger Beziehung stehen. Die meist stark variierenden und schwer erfassbaren geologischen und topologischen Randbedingungen eines modellierten Gebiets erschweren diese Problematik. Betrachtet man den Prozess der Abflussbildung systemisch abstrakt mit den Eingangsgrößen $u(t)$, den Zustandsgrößen $q(t)$ und der Ausgangsgröße $x(t)$ und verzichtet zunächst auf einen theoretischen analytischen Modellansatz, dann liegt es nahe, beobachtbare Zustandsgrößen

(aus $u(t)$ und $x(t)$ folgt $q(t)$) und darüber steuerbare Ausgabegrößen (aus $u(t)$ und $q(t)$ folgt $x(t)$) mit nutzbaren systemtechnischen Zusammenhängen zu wählen. Liegen diese Größen zudem als Messdaten vor, was bei vielen Gebieten durch eine große Anzahl historischer Daten gewährleistet ist, dann existiert eine vielversprechende Situation für eine experimentelle Entwicklung und Identifikation von Systemmodellen für den N-A-Prozess.

Hierbei sind grundsätzlich zwei Aufgaben zu lösen: 1) die Strukturidentifikation und 2) die Parameteridentifikation. Für erstere ist es hilfreich und effektiv, wenn vorliegende Erfahrung als a priori Wissen über die Beziehung der gewählten Systemgrößen in die Strukturfindung eingebracht werden kann. Die Parameteridentifikation liefert dann bei festgelegter Struktur eine statische Kennfunktion zwischen den Eingangs- und Ausgabegrößen.

Zur Lösung der beiden oben genannten Aufgaben bietet sich der Einsatz unkonventioneller Methoden des Soft Computing an: Fuzzy Logic, Neuronale Netze oder Evolutionäre Algorithmen. Wir haben uns für den Einsatz der Fuzzy Logic entschieden, die hier entscheidende Vorteile bietet: 1) mit WENN ... DANN – Regeln lässt sich Erfahrungswissen über den N-A-Prozess direkt formulieren und in die Systemstruktur einbringen. 2) Für die Strukturfindung werden Fuzzy Mengen über Eingangs- oder Zustandsgrößen definiert. 3) Für die Parameteridentifikation lässt sich ein statistisches Optimierungsverfahren einsetzen, das eine Kennfunktion für die Abflussgröße mit minimalem Fehler über die Optimierungsmenge liefert. 4) Das Systemmodell ist transparent bis hin zu einer quasi physikalischen Interpretation, z.B. bei Benutzung restringierter Optimierungsmethoden. 5) Alle genannten Schritte lassen sich rechnergestützt und weitestgehend automatisiert durchführen, so dass eine sehr zeit- und kostengünstige Erzeugung von N-A-Modellen ermöglicht wird. Diese Form der Modellbildung und die erzielten Ergebnisse werden im Folgenden beschrieben.

2. Fuzzy Modellansatz

Fuzzy Modelle werden aufgebaut, indem mittels WENN-DANN Regeln Fuzzy Relationen über die Eingangs-, Zustands- und einer Ausgabegröße hergestellt werden. Hierzu werden für diese Systemgrößen (linguistische Variablen) zunächst Fuzzy Mengen (linguistische Terme) definiert, um die Zugehörigkeit von bestimmten Wertebereichen zu sprachlich handhabbaren Eigenschaften wie z.B. „klein“, „mittel“ oder „groß“ zu definieren. Damit kann nun ein Systemmodell in linguistischer Form über einen Satz von WENN-DANN-Regeln (Regelbasis) erzeugt werden, wobei auf das Wissen (fachliche Expertise) über das typische Prozessverhalten (z.B. qualifizierte logische Beziehung der Eingangs- und Ausgabevariablen in bestimmten Systemsituationen bzw. -zuständen) explizit zurückgegriffen werden kann. Die Auswertung erfolgt dann über ein Fuzzy Inferenzsystem, bei dem die scharfen Eingangswerte in die Zugehörigkeitswerte der betroffenen Fuzzy Mengen „fuzzifiziert“ werden, die Regeln damit fuzzy-logisch ausgewertet und in einen „scharfen“ Ergebniswert überführt werden. Beim Aufbau von Fuzzy Inferenzsystemen unterscheidet man grundsätzlich zwischen zwei unterschiedlichen Vorgehensweisen: Fuzzy Systeme nach Mamdani oder nach Takagi-Sugeno-Kang (TSK, SUGENO & KANG 1985). Im Folgenden werden die beiden Systeme und ihre Vor- und Nachteile beim Einsatz zur Hochwasservorhersage kurz erläutert.

Der typische Aufbau einer Regel R_i bei einem Mamdani-System lautet beispielsweise:

R_i : WENN Q_{t_0} IST niedrig UND $Q_{t-1, \dots, t-n}$ IST niedrig DANN Q_{t+1} = niedrig

Bei diesem System ergibt sich der Ausgabewert der Zielgröße zunächst als Fuzzy Menge. Jede Regel R_i ($i=1,2, \dots$) hat eine bestimmte Fuzzymenge der Zielgröße als Ergebnis in der Konklusion. Der scharfe Ausgabewert ergibt sich über die Auswertung der einzelnen Regelprämissen (Aggregation) und Abbildung auf die Regelkonklusion und anschließende Zusammenfassung aller Konklusionsmengen (Akkumulation) und deren Defuzzifizierung (SCHULZ 2002).

Dieser Aufbau bietet den Vorteil, dass das vorhandene Wissen über die Abflussbildung direkt in die Modellierung (Strukturfindung und Parameteroptimierung) einfließen kann. Das so entstandene System ist sowohl im Aufbau als auch in der mathematischen Beschreibung einfach. Es entstehen jedoch praktische Nachteile:

- Jede Systemgröße muss in Fuzzy Mengen unterteilt werden, wodurch die Anzahl der verwendbaren Größen wegen der sonst zu großen Regelbasis stark eingeschränkt wird.
- Es ist keine Unterscheidung zwischen dem qualitativen und dem quantitativen Einfluss der Systemgrößen möglich.
- Das entstandene System ist nur schwierig und (zeit)aufwändig zu optimieren.

Aus den genannten Gründen betrachten wir zur Modellierung ein TSK Fuzzy System. In diesen Systemen besteht die Konklusion aus einer beliebigen linearen Funktion mit einem Parametersatz p_i . Eine Beispielregel R_i lautet hier:

R_i : WENN Q_{t0} IST niedrig UND $Q_{t-1, \dots, t-n}$ IST niedrig DANN $Q_{t+1} =$
 $p_1 E_1 + \dots + p_n E_n + \dots + p_{n+k} E_{n+k}$

Damit ist eine Trennung der Zustandsbeschreibung, die in der Prämisse (WENN-Teil) enthalten ist, von dem quantitativen Einfluss der Systemgrößen, der in der Konklusion enthalten ist, möglich. Je nach Zustand des Gesamtsystems findet eine bestimmte Gewichtung der Systemgrößen statt und es können auch Systemgrößen in der Konklusion verwendet werden, die in der Zustandsbeschreibung nicht berücksichtigt wurden. Auf diese Weise ergibt sich ein weiterer Vorteil, indem die Anzahl der Regeln bei TSK Systemen im Allgemeinen niedriger als bei Mamdani Systemen ist.

Die Auswertung einer Fuzzy Regelbasis z.B. zur konkreten Bestimmung eines Vorhersagewertes ist in Abbildung 1 prinzipiell gezeigt. Links erkennt man die Fuzzifizierung der (scharfen) Eingangswerte e_i und die Ermittlung der Inferenzwerte der einzelnen Regeln (Erfüllungsgrad w^i) z.B. über einen Produkt-Operator für die Prämissenkonjunktion. Die Erfüllungsgrade beschreiben, wie stark sich das System in dem von der i -ten Regel beschriebenen Zustand befindet. Für jede Regel wird ein Resultat $Q_{i,a}$ der linearen Konklusionsgleichung bestimmt. Das Gesamtergebnis über alle Regelkonklusionen wird berechnet, indem die Resultate der einzelnen Regeln entsprechend ihres Erfüllungsgrades gewichtet gemittelt werden.

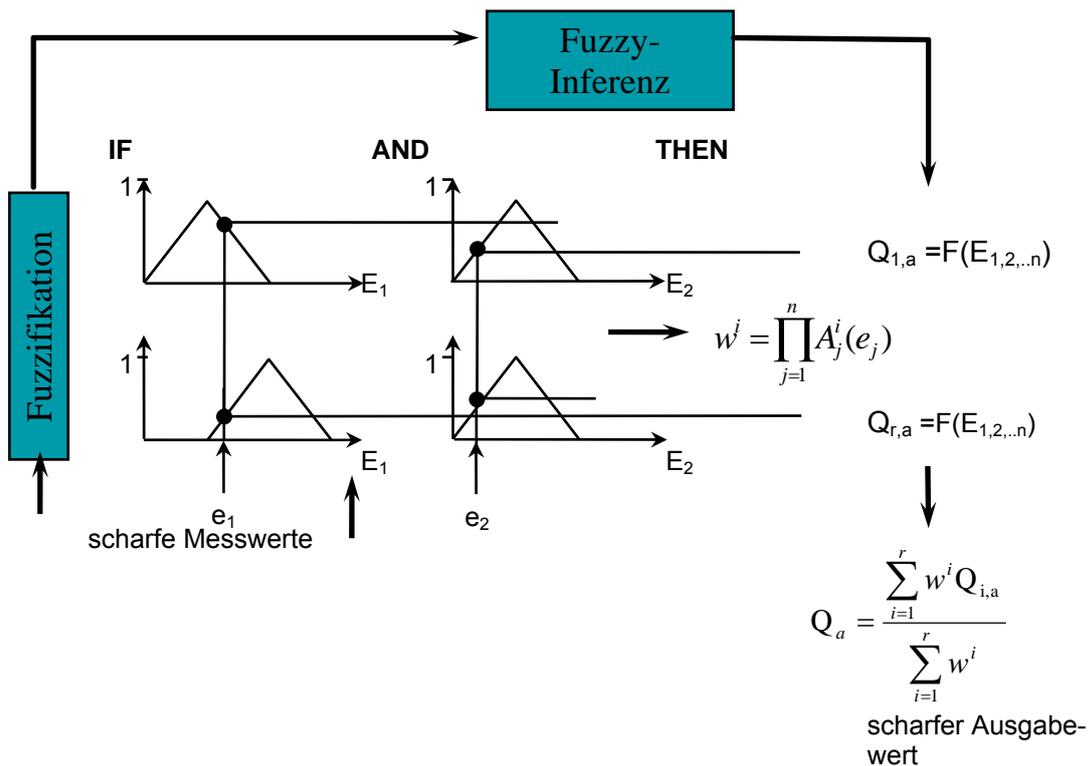


Abb 1: Fuzzy Inferenzsystem (TSK-System) zur Berechnung von Abflussvorhersagen Q_a aus Messwerten e_i

3. Modellierungsverfahren

Die Modellierung des Fuzzy-Systems umfasst wie bereits erwähnt zwei wesentliche Aspekte, die Strukturfindung und die Parameteridentifikation. Bei der Strukturfindung kommen verschiedene Informationsquellen zum Einsatz. Eine Möglichkeit ist es, a priori Wissen über Beziehungen der Systemgrößen zu verwenden. Hilfreich ist die Analyse der Vorgehensweise, die erfahrene Experten speziell in diesem Gebiet zur Vorhersage verwenden. Häufig können diese über Jahre gesammelten Erfahrungen in einfacher Form in Regeln formuliert werden. Auch die Unterscheidung der verschiedenen Zustände und wovon es abhängt, dass ein Gebiet sich in einem bestimmten Zustand befindet, ist hilfreich. Als weiteres Hilfsmittel können unterschiedliche statistische und analytische Methoden aus dem Bereich des Data Minings verwendet werden.

Welche Systemgrößen nun konkret zur Beschreibung des Systemzustandes bei der Strukturidentifikation verwendet werden - also welche Größen in die Prämisse eingehen und welche WENN-DANN-Beziehungen aufgestellt werden - ist von unterschiedlichen, meist gebietsspezifischen Faktoren abhängig. Denkbar ist es, bei größeren Einzugsgebieten zunächst nur den aktuellen Pegel zu verwenden, da dieser viel über den Gesamtzustand aussagt. Auch die Integration oder die Differenziation dieses Wertes über bestimmte Zeitbereiche liefert weitere aussagekräftige Informationen. In kleineren Gebieten ist eventuell eine Einbeziehung der Bodenfeuchte notwendig, da diese einen Hinweis darauf gibt, welcher Anteil vom aktuellen Niederschlag als Abfluss zu erwarten ist.

Auch der Aufbau der Konklusion zur expliziten Berechnung des zu erwartenden Abflusses kann nicht allgemein gültig formuliert werden. In unseren Modellen hat sich eine Linearkombination bewährt, in der der aktuelle Abfluss und die Niederschläge der vergangenen Stunden gewichtet aufsummiert werden. Die Länge der Konklusion, also das Zeitfenster, in dem die Niederschläge noch eine Auswirkung auf die Abflussbildung haben, ist aber auch hier wieder gebietsabhängig.

Liegt die Struktur des Systems fest, sind die Parameter p_i der Regelkonklusionen zu bestimmen. Bei TSK-Systemen erhält man bei k Konklusionsgrößen einen Parametervektor der Dimension k . Bei n Regeln ergibt sich somit ein Parametersatz mit $n \cdot k$ Parametern, die so zu bestimmen sind, dass die Zielgröße optimal approximiert wird. Dies lässt sich für reale Vorhersagesystemen praktisch nur noch automatisiert erreichen.

4. Modelloptimierung

Im vorherigen Abschnitt wurde deutlich, dass bei der Verwendung von TSK-Systemen eine große Anzahl von Parametern entsteht. Gehen beispielsweise in der Konklusion der aktuelle Abfluss und die Stundenmittel der Niederschläge der letzten 40 Stunden ein, so enthält jede Konklusion 41 Parameter. Werden 6 Regeln aufgestellt, dann ergeben sich 246 Parameter. Eine manuelle Optimierung des Systems ist somit nahezu unmöglich. Verwendet wird stattdessen ein statistisches Optimierungsverfahren, welches historische Messwerte verwendet, um Parametersätze mit minimalem quadratischem Fehler der Zielgröße für diese Trainingsmenge zu berechnen.

Betrachten wir nun dieses Beispielmodell mit seinen 6 Regeln. In den Regeln wird der aktuelle Abfluss Q_t und die Bodenfeuchte B_t verwendet; Q_t wird in drei Fuzzy-Mengen eingeteilt, B_t in zwei. Das komplette Modell wäre demnach:

$$R_1: \quad \text{WENN } Q_t \text{ IST klein UND } B_t \text{ IST klein DANN } Q_{t+1} = p_{1,1}Q_t + p_{1,2}N_t + \dots + p_{1,41}N_{t-39}$$

$$R_2: \quad \text{WENN } Q_t \text{ IST mittel UND } B_t \text{ IST klein DANN } Q_{t+1} = p_{2,1}Q_t + p_{2,2}N_t + \dots + p_{2,41}N_{t-39}$$

...

$$R_6: \quad \text{WENN } Q_t \text{ IST groß UND } B_t \text{ IST groß DANN } Q_{t+1} = p_{6,1}Q_t + p_{6,2}N_t + \dots + p_{6,41}N_{t-39}$$

Zu optimieren sind also die Parameter $p_{i,j}$, $1 \leq i \leq 6$, $1 \leq j \leq 41$. Dazu kann unter Verwendung von Messdaten aus einer Zeitreihe mit je n Einträgen für alle Systemgrößen ein überbestimmtes lineares Gleichungssystem aufgestellt werden: $y = F(E)p$. y ist ein Vektor mit den gemessenen Abflüssen Q_t , p der gesuchte Vektor aller Parameter $p_{i,j}$ und $F(E)$ eine Matrix, in der die normierten Erfüllungsgrade $w_{n,j}$ aller Regeln multipliziert mit den zugehörigen Messwerten enthalten sind. Das vollständige Gleichungssystem für das betrachtete Beispiel lautet demnach:

$$\begin{pmatrix} Q_1 \\ Q_2 \\ \vdots \\ Q_n \end{pmatrix} = \begin{pmatrix} \overbrace{w_{1,1}Q_0 \quad w_{1,1}N_1 \quad \cdots \quad w_{1,1}N_{-39}}^{41\text{Werte}} \quad \overbrace{w_{2,1}Q_0 \quad w_{2,1}N_1 \quad \cdots \quad w_{2,1}N_{-39}}^{41\text{Werte}} \quad \cdots \quad w_{6,1}N_{-39} \\ w_{1,2}Q_1 \quad w_{1,2}N_2 \quad \cdots \quad w_{1,2}N_{-38} \quad w_{2,2}Q_1 \quad w_{2,2}N_2 \quad \cdots \quad w_{2,2}N_{-38} \quad \cdots \quad w_{6,2}N_{-38} \\ \vdots \quad \vdots \\ w_{1,n}Q_{n-1} \quad w_{1,n}N_n \quad \cdots \quad w_{1,n}N_{n-39} \quad w_{2,n}Q_{n-1} \quad w_{2,n}N_n \quad \cdots \quad w_{2,n}N_{n-39} \quad \cdots \quad w_{6,n}N_{n-39} \end{pmatrix} \begin{pmatrix} P_{1,1} \\ P_{1,2} \\ \vdots \\ P_{1,41} \\ P_{2,1} \\ P_{2,2} \\ \vdots \\ P_{2,41} \\ \vdots \\ P_{6,41} \end{pmatrix}$$

Es müssen also wenigstens so viele Messwerte vorliegen, wie Parameter in dem Modell enthalten sind. In der praktischen Anwendung hat sich gezeigt, dass n ungefähr $10|p|$ sein sollte. Zusätzlich sollten die Messdaten auch alle Regeln abdecken. In der Praxis liegen meistens weit umfangreichere Messreihen vor, in denen auch Messwerte unterschiedlicher Jahreszeiten vorkommen. Dabei können auch große Gleichungssysteme mit $n > 50.000$ (mehrere Jahre) mit geringem Zeitaufwand (wenige Minuten) auf heute üblichen PCs (1,6 GHz CPU, 1GByte Arbeitsspeicher) gelöst werden.

Zur Lösung des Gleichungssystems wird ein Verfahren (LS) verwendet, das die Parameter so berechnet, dass die Summe der Quadrate der Differenzen zwischen den gemessenen und den berechneten Abflüssen minimal wird (Least-Squares-Kriterium, WEINHOLD 2003). Neben dieser Lösung, die im möglichen Lösungsraum das globale Minimum im Sinne des Least-Squares-Fehlers bestimmt, können auch restringierte Lösungen berechnet werden. Der Grund für diese Restriktionen ist, dass bei der global optimalen Lösungen negative Werte für einzelne Parameter entstehen können, die z.B. als Parameter für den Abfluss zum Zeitpunkt t , schwierig zu interpretieren sind. Zur Vermeidung negativer Parameter kann das Non-Negative-Least-Squares-Verfahren (NNLS) benutzt werden. Eine weitere Einschränkung ist die Vorgabe eines möglichen Wertebereiches für jeden einzelnen Parameter. Dies ist insbesondere dann sinnvoll, wenn dieser Parameter direkt hydrologisch interpretiert werden kann und sein möglicher Wertebereich bekannt ist.

Mit Hilfe dieser automatischen Modelloptimierung erhält man ein Modell, das unter Verwendung der vorgegebenen Regeln (Modellstruktur), das Verhalten des Gebietes unter Beachtung des Gütekriteriums (LS) optimal approximiert. Die einzelnen Parameter, die durch die Lösung des Gleichungssystems berechnet werden, enthalten insbesondere bei restringierten Modellen Informationen über das Abflussgeschehen in dem Gebiet. Wird ein Modell verwendet, in dem wie oben beschrieben eine Linearkombination aus dem vorherigen Abfluss und den Niederschlägen der letzten Stunden verwendet wird, so sind die Parameter direkt inter-

pretierbar. Bei den Faktoren, mit denen die Niederschläge multipliziert werden, handelt es sich um den Anteil des Niederschlags, der zu der entsprechenden Stunde zum Abfluss führt.

5. Praktische Anwendung und Testergebnisse

Das beschriebene Verfahren zur Modellgenerierung eignet sich für die Abflussvorhersage in unterschiedlich großen Einzugsgebieten mit unterschiedlichen Eingangsgrößen. Dies wurde unter anderem mit Modellen für den Pegel Trier/Mosel und den Pegel Saarburg/Leuk erfolgreich nachgewiesen. Abbildung 2 zeigt einen Teil der Pegel (rot) und Teileinzugsgebiete (schwarz) im Einzugsgebiet der Mosel. Für alle rot dargestellten Pegel sind Fuzzy Vorhersagemodelle im operationellen Einsatz. Die schwarzen Verbindungslinien kennzeichnen die Modellabhängigkeiten bzw. die Eingangsgrößen für das jeweilige Modell. Der Pegel Saarburg/Leuk besitzt neben dem Abfluss am Pegel selbst lediglich die Niederschlagsdaten der Teileinzugsgebiete Leuk1-Leuk3 als Eingangsgrößen. Der Pegel Trier besitzt neben dem Pegel selbst nur die Oberliegerpegel Pegel Perl/Mosel, Fremersdorf/Saar, Bollendorf/Sauer, Prümzurlay/Prüm, Alsdorf/Nims und Saarburg/Leuk als Eingangsgrößen.

Beispielhaft werden nachfolgend das Vorhersagemodell für den Pegel Saarburg/Leuk (57 km²) genauer betrachtet und die Modellergebnisse vorgestellt. In das Pegelmodell Saarburg gehen der Abfluss am Pegel und die regionalisierten Niederschlagsdaten der drei Teilgebiete des Gesamteinzugsgebietes (Abbildung 3) ein.

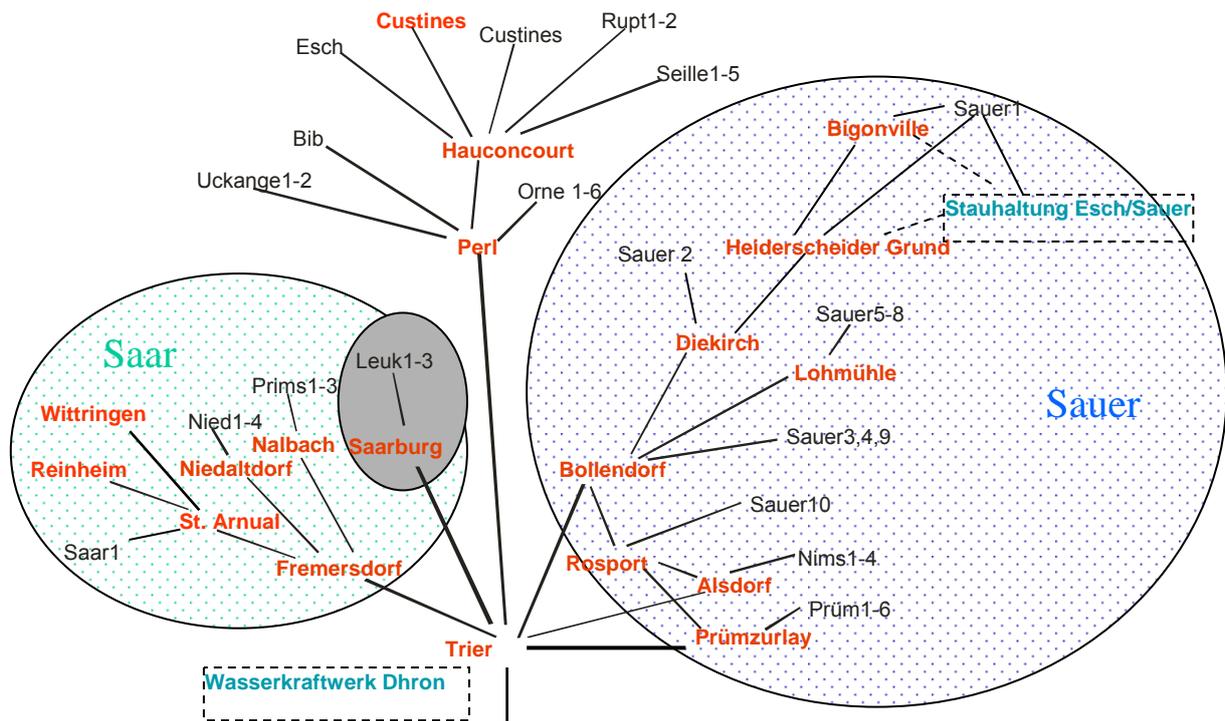
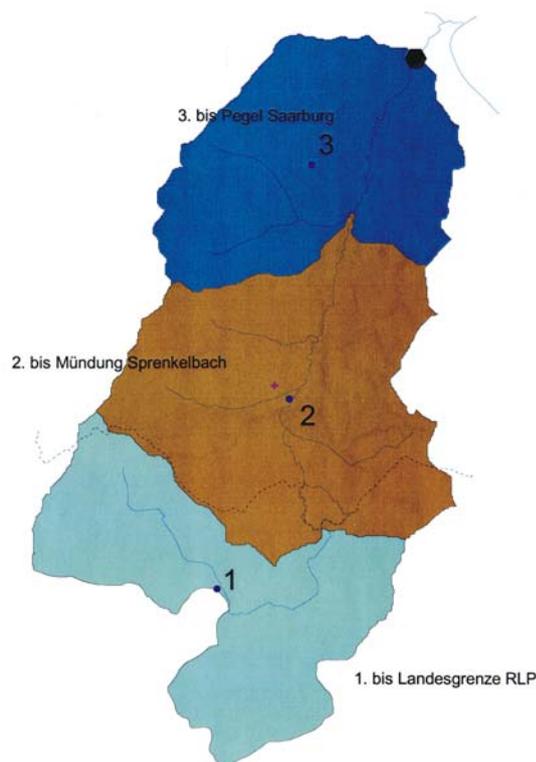


Abb. 2: Ausschnitt aus dem Einzugsgebiet der Mosel mit Pegeln und dazugehörigen Fuzzy Vorhersagemodellen sowie deren Abhängigkeiten



ID	Rechtswert	Hochwert	Fläche
1	2535754.39	5486947.67	16.91
2	2537025.61	5490287.39	23.09
3	2537420.48	5494418.67	17.01

Abb. 3: Teileinzugsgebiete der Leuk bis zum Pegel Saarburg

Modellaufbau

Die Fuzzy-Modelle zur Abflussvorhersage beschreiben in den Regelprämissen unterschiedliche Situationen bzw. Zustände des Einzugsgebietes. Die mögliche Anzahl der beschreibbaren Situationen richtet sich nach der Unterteilung der Eingangsgrößen in Fuzzy Mengen und deren Kombinationsmöglichkeiten. Die jeweiligen Regelkonklusionen setzen sich beim Ansatz nach Takagi-Sugeno-Kang aus den Eingangsgrößen und deren jeweiligen Gewichtungparametern zusammen.

Aus verfügbaren Messgrößen (Pegelabfluss, Niederschlag in Teileinzugsgebieten) lassen sich einige Kombinationen zur Bildung von Situationsbeschreibungen ableiten. Berücksichtigt man dazu noch die unterschiedlichen zeitlichen Abhängigkeiten oder generiert zusätzliche Systemgrößen, dann ergeben sich enorme Kombinationsmöglichkeiten für die Beschreibung der im Gesamteinzugsgebiet vorkommenden Situationen. Da jede Situationsbeschreibung eine Regelprämisse darstellt, würde somit ein Fuzzy Modell entstehen, das aufgrund seiner Größe nicht mehr handhabbar wäre. Die Untersuchungen an 20 Einzugsgebieten, an denen neben dem Pegel lediglich Niederschlagsdaten zur Verfügung stehen, haben gezeigt, dass Modelle mit einer sehr guten Vorhersagegüte erzeugt werden können, wenn die Situationsbeschreibung sich auf zwei Größen stützt:

1. Abfluss am Pegel zum Vorhersagezeitpunkt und

2. mittlerer Abfluss am Pegel über einen gebietsabhängigen Zeitraum vor dem Vorhersagezeitpunkt.

Beim Pegel Saarburg/Leuk wurde als gebietsabhängiger Zeitraum zur Mittelung des Abflusses ein Zeitbereich von 13 bis 48 Stunden automatisiert ermittelt. Die Einteilung der beiden Eingangsgrößen in Fuzzy-Sets geschieht automatisch nach einer Heuristik, die sich nach der Anzahl der verfügbaren Daten, nach der Anzahl der zu optimierenden Parameter in der Konklusion und nach der Verteilung der Eingangsdaten im Eingangsraum richtet.

In den Konklusionen werden für die unterschiedlichen Situationen die jeweiligen Einflüsse der Eingangsgrößen durch deren Gewichtung dargestellt. Neben den Niederschlagsdaten wird in den Konklusionen der Abfluss zum Vorhersagezeitpunkt als zusätzliche Eingangsgröße betrachtet. Eine Datenvoranalyse zeigte am Pegel Saarburg, dass sich der Niederschlag in jeder Situation mit einer maximalen Reaktionszeit von 12 Stunden am Pegel auswirkt. Für den Pegel Saarburg ergab sich somit eine Regelbasis mit lediglich sechs Regeln. Die Situationen (Regelprämissen) lauten:

WENN $Q_Saarburg_{t0}$ IST klein UND $Abflussmittel_in_Saarburg_{t-13..t-48}$ IST klein DANN ...
WENN $Q_Saarburg_{t0}$ IST mittel UND $Abflussmittel_in_Saarburg_{t-13..t-48}$ IST groß DANN ...
WENN $Q_Saarburg_{t0}$ IST groß UND $Abflussmittel_in_Saarburg_{t-13..t-48}$ IST klein DANN ...
WENN $Q_Saarburg_{t0}$ IST klein UND $Abflussmittel_in_Saarburg_{t-13..t-48}$ IST groß DANN ...
WENN $Q_Saarburg_{t0}$ IST mittel UND $Abflussmittel_in_Saarburg_{t-13..t-48}$ IST klein DANN ...
WENN $Q_Saarburg_{t0}$ IST groß UND $Abflussmittel_in_Saarburg_{t-13..t-48}$ IST groß DANN ...

Die jeweiligen Regelkonklusionen beschreiben den quantitativen Einfluss der Eingangsgrößen und gehen umso stärker in das Gesamtergebnis ein, je stärker die jeweilige Regelprämisse zutrifft. Der Aufbau der Regelkonklusionen sieht für alle Regeln folgendermaßen aus:

$$Q_Saarburg(t+1) = Q_Saarburg_{t0} +$$
$$p_{i,0} * TG1_{t0} + p_{i,1} * TG1_{t-1} + \dots + p_{i,11} * TG1_{t-11} +$$
$$p_{i,12} * TG2_{t0} + p_{i,13} * TG2_{t-1} + \dots + p_{i,23} * TG2_{t-11} +$$
$$p_{i,24} * TG3_{t0} + p_{i,24} * TG3_{t-1} + \dots + p_{i,35} * TG3_{t-11}$$

wobei i die Regel bezeichnet.

Somit ergeben sich $6 \cdot 36 = 216$ zu optimierende Parameter in den Regelkonklusionen. Dem stehen 6450 stündlich gemessene Datensätze gegenüber. Somit lässt sich, wie in Kapitel 4 beschrieben, ein stark überbestimmtes lineares Gleichungssystem bilden. Je größer das Verhältnis zwischen den verfügbaren Daten und den zu optimierenden Parametern ist, umso besser können Hochwasserereignisse mit einer von den historischen Ereignissen abweichenden Charakteristik vorhergesagt werden.

Für die praktische Optimierung der Parameter werden mathematische Lösungsverfahren (LS, NNLS usw.) angewendet. Dabei werden die Parameter so bestimmt, dass das Modell eine 1h-Vorhersage hinsichtlich des quadratischen Fehlers möglichst genau berechnet. Hierbei können auch negative Parameter ermittelt werden, die hinsichtlich der Interpretation der Abhängigkeiten der Eingangsgrößen nicht verwertbar sind. Eine Einschränkung des Eingaberaums dahingehend, dass die Parameter nicht negativ sein dürfen bzw. dass Parameterwertebereiche vorgegeben werden ist möglich (siehe Kapitel 4). Erste Untersuchungen, bei denen Parameterwertebereiche aufgrund von Wissen über das Einzugsgebiet eingeschränkt wurden, zeigen dass die Qualität der Vorhersagen nahezu unverändert bleibt, die

Interpretierbarkeit der Modelle jedoch stark zunimmt. Hinzukommt, dass dabei einige Parameter durch die Optimierung auf Null gesetzt werden, was bedeutet, dass die Eingangsgröße, die dem Modell zur Optimierung angeboten wurde, auf die Vorhersageberechnung keinen Einfluss hat. Die Zahl der Modellparameter kann somit weiter verringert werden.

Zur Optimierung des Pegelmodells Saarburg standen Daten von 6450 stündlichen Messzeitpunkten über 10 Hochwasserereignisse zur Verfügung. Eine dem Optimierungsverfahren vorgegebene grobe Einschränkung der Parameterwertebereiche, mit dem Wissen über den möglichen Einfluss der Eingangsgrößen eingebracht wurde, hatte zur Folge, dass bei der Optimierung die Gewichtung der Niederschläge, die mehr als neun Stunden vor dem Vorhersagezeitraum liegen, auf Null gesetzt wurden. Dies bedeutet, dass die geschätzte Reaktionszeit des Einzugsgebietes von den grob ermittelten 12 Stunden auf neun Stunden verkürzt werden konnte. Dies zeigt auch, dass neben der automatischen Ermittlung des quantitativen Einflusses der Eingangsgrößen die zeitlich variablen Abhängigkeiten je nach Situation automatisch ermittelt werden können.

Modelleigenschaften

Die Bewertung der Vorhersageberechnungen erfolgt anhand unterschiedlicher Fehlergrößen. Eine wichtige Fehlergröße für die Beurteilung der Modellvorhersagequalität ist der modifizierte Nash-Sutcliffe-Koeffizient (ME), der sich wie folgt berechnet:

$$ME_{\text{mod}} = 1 - \frac{\sum_{i=1}^n (Q_{\text{gemessen}_i} - Q_{\text{berechnet}_i})^2}{\sum_{i=1}^n (Q_{\text{gemessen}_i} - Q_{(\Delta W(10\text{cm}))_i})^2}$$

Der modifizierte Koeffizient ME_{mod} gibt ein Maß für die Übereinstimmung des berechneten Abflusses mit dem gemessenen Abfluss bezogen auf eine reale Abweichung des Wasserstands um $\Delta W = 10\text{cm}$ an. Der ideale Wert $ME_{\text{mod}} = 1$ zeigt eine optimale Güte der Vorhersage an. Werte von $ME_{\text{mod}} \geq 0$ zeigen eine mittlere Abweichung des berechneten Wasserstands von $\Delta W \leq |\pm 10|\text{cm}$ an. Bei Werten $ME_{\text{mod}} < 0$ liegt die Abweichung des aus dem berechneten Abfluss ermittelten Wasserstands bei $\Delta W > |\pm 10|\text{cm}$.

Die Abbildungen 4 und 5 beschreiben die Kurvenverläufe der Mess- und Vorhersagewerte für die Hochwasserereignisse vom Februar 1997 und Januar 2003. Tabelle 1 zeigt die modifizierten Nash-Sutcliffe-Koeffizienten für die beiden Ereignisse zusammengefasst bis zu 24 Vorhersagestunden.

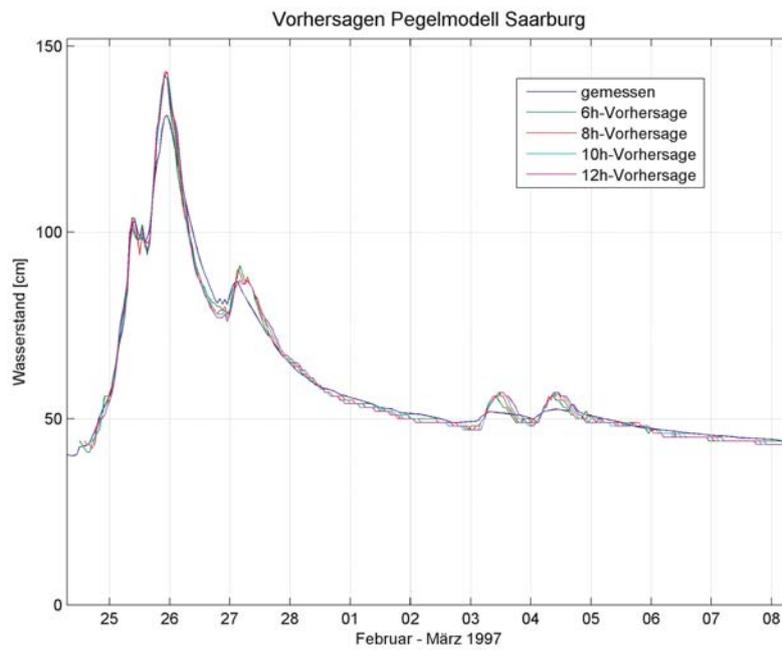


Abb. 4: Vorhersageergebnisse des Hochwasserereignisses von Februar-März 1997 am Pegel Saarburg

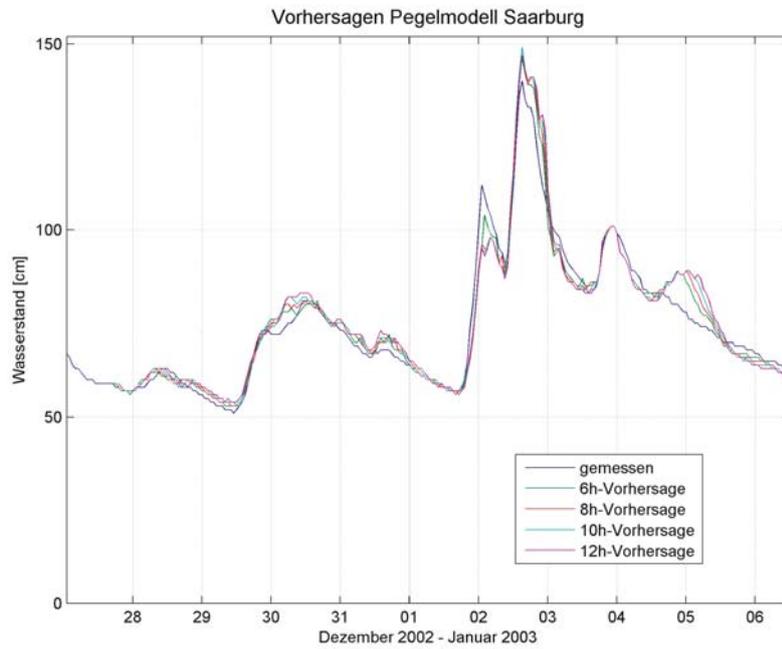


Abb. 5: Vorhersageergebnisse des Hochwasserereignisses vom Dezember 2002- Januar 2003 am Pegel Saarburg

Tab. 1: Modifizierte Nash-Sutcliffe-Koeffizienten für Pegel Saarburg, Ereignisse 2002 und 2003

Vorhersage [h]	Ansteigender Bereich	Scheitelpbereich	Abfallender Bereich	Gesamt
1	0,9906	0,9446	0,9923	0,9878
2	0,9796	0,8651	0,9824	0,9718
3	0,9693	0,7973	0,9691	0,9548
4	0,9573	0,7255	0,9574	0,9379
5	0,948	0,6003	0,9464	0,9176
6	0,9458	0,5048	0,9365	0,9025
7	0,9386	0,4438	0,926	0,8885
8	0,9325	0,3867	0,9162	0,8756
9	0,9247	0,342	0,9127	0,8670
10	0,9224	0,301	0,9067	0,8589
11	0,9113	0,2869	0,8993	0,8496
12	0,9077	0,2709	0,8947	0,8440
13	0,904	0,2865	0,8905	0,8414
14	0,8989	0,2857	0,8851	0,8363
15	0,8915	0,2934	0,8854	0,8350
16	0,8872	0,2905	0,8821	0,8312
17	0,8838	0,2953	0,882	0,8305
18	0,8763	0,3028	0,8815	0,8288
19	0,8745	0,3069	0,8802	0,8276
20	0,8658	0,3253	0,8774	0,8251
21	0,8635	0,3278	0,8785	0,8253
22	0,8554	0,3156	0,8776	0,8215
23	0,8527	0,3612	0,8777	0,8250
24	0,8429	0,3678	0,8765	0,8224

Fazit

Die entwickelten Fuzzy Vorhersagemodelle sind nach der Optimierung mit festen Parametern eingestellt. Man kann sie als hochdimensionale Kennfelder betrachten, die im operationellen Betrieb keine Anpassung, also keine Eichung/Kalibrierung benötigen. Neben einer schnellen, automatisierten Modellerstellung bietet dieser Entwicklungsansatz weitere Vorteile:

- Eine schnelle Berechnung der Vorhersagen im operationellen Betrieb (Millisekundenbereich). Dies ermöglicht im Hochwasserfall eine ständige Aktualisierung der Vorhersageberechnung, sobald neue Messdaten verfügbar sind.
- Eine Interpretation der Gewichtungparameter im Modell. Daraus kann der quantitativen und zeitlichen Einflüsse der Eingangsgrößen auf den Abfluss ermittelt werden.

Aus dem für die Parameteroptimierung verwendeten Verfahren ist die Eigenschaft ableitbar, dass das Integral der vorhergesagten Abflüsse über alle zur Modellierung verwendeten Daten nahezu identisch mit dem Integral der gemessenen Daten ist. In Verbindung mit der hohen Vorhersagegüte lässt sich zeigen, dass dieses Verfahren auch im Bereich der Wasserhaushaltsmodellierung einsetzbar ist. Die Leistungsfähigkeit dieses Fuzzy Ansatzes hat sich auch bei der Entwicklung einer automatischen Wasserkraftwerkssteuerung zur optimierten Stromerzeugung (FERRING 2006) und zur Korrektur von systematischen Fehlern in den Niederschlagsdaten des Lokalmodells gezeigt (GEMMAR 2006).

Literatur

- FERRING, D. & M. RUMPF (2006): Entwicklung eines plattformunabhängigen Systems für die operationelle Steuerung des Generatoreinsatzes im Dhronkraftwerk. - Abschlussarbeit (Betr. Prof. Dr. P. Gemmar), Fachbereich Informatik, FH Trier
- GEMMAR, P., M. GREVING & M. STÜBER (2006): Verbesserte Hochwasservorhersage durch Korrektur von Niederschlagsvorhersagen. - Wasserwirtschaft 7-8
- SCHULZ, G (2002): Regelungstechnik. - Oldenbourg Wissenschaftsverlag GmbH, München
- STÜBER, M. & P. GEMMAR (1997): Untersuchungen zur Prognose der Pegelentwicklung bei Hochwasserereignissen mit Fuzzy-Logic am Beispiel zweier Moselpegel. - Schriften zur Angewandten Informatik der FH Trier, Bericht 9703, Trier
- STÜBER, M., P. GEMMAR & M. GREVING (2000): Machine supported Development of Fuzzy-Flood Forecast Systems. - In: PIK Report No. 65, European Conference on Advances in Flood Research, Proceedings Vol. 2, Hrsg. A. Bronstert et al., Potsdam
- SUGENO, M. & G.T. Kang (1985): Fuzzy Identification of Systems and its Applications to Modeling and Control. - In: IEEE Trans. Syst. Man Cybern. 15(1), S. 116-132
- SUGENO, M. & G.T. KANG (1988): Structure Identification of Fuzzy Models. - Fuzzy Sets and Systems, Vol. 28, Amsterdam, S. 15-33
- WEINHOLD, J. O. (2003): Effiziente Lösung restringierter Least-Squares Probleme zur Parameteroptimierung in TSK-Fuzzy-Systemen. - Diplomarbeit, Universität zu Lübeck