

# EVOLUTIONARY MULTIOBJECTIVE DIRECT POLICY SEARCH (EMODPS) – EINE HEURISTISCHE ENTSCHEIDUNGSHILFE IN DER ENERGIE-WASSERWIRTSCHAFT

Christoph LIBISCH-LEHNER<sup>1</sup>

## Einleitung

Talsperren und Wasserspeicher werden seit mehr als 2000 Jahren errichtet, sind heute Teil der modernen Infrastruktur und werden in der Regel im komplexen sozio-ökonomischen Kontext beschrieben. Sie dienen der Energieerzeugung, der Trinkwasserversorgung, der Bewässerung, dem Hochwasserschutz oder der Niederwasseraufhöhung in Trockenperioden. Der Einfluss externer Größen wie hydro-meteorologische Rahmenbedingungen oder Veränderungen im Strommarkt, die Bedürfnisse mehrerer Nutzer gleichzeitig zu bedienen, oder das Zusammenspiel zwischen wasserwirtschaftlichen Anlagen im Verbundsystem erhöhen die Komplexität der Einsatzplanung einzelner Anlagen. Ein effektiver und optimierter Betrieb der Anlagen ist von großer Bedeutung, um die Ressource Wasser bestmöglich zu nutzen.

Der Betrieb von wasserwirtschaftlich Anlagen kann mittels der „Theorie zur optimalen Steuerung“ (Optimal Control Theory) beschrieben werden. Die Theorie der optimalen Steuerung setzt sich mit dem Problem auseinander, Kontrollregeln für ein System zu finden, um dieses im Sinne einer Zielfunktion bestmöglich zu betreiben. Traditionell werden optimale Betriebsregeln mittels direkter Optimierungsmethoden (z.B., Dynamische Programmierung) ermittelt. Für komplexe, „real-world“ Probleme weisen diese Methoden jedoch Limitierungen auf. Dazu zählen der „Fluch der Dimension“, der „Fluch der Modellierung“ und der „Fluch der Mehrzielplanung“. Mit Methoden des „Bestärkenden Lernen“ (Reinforcement Learning) können die genannten Probleme aber überwunden werden. Durch drastische Reduktion von Entscheidungsvariablen und durch die getrennte Betrachtung von wasserwirtschaftlichen System und Optimierungsstrategie (indirekte Optimierung) kann die Systemkomplexität ohne Einschränkung erhalten bleiben.

## EMODPS – Evolutionary Multi-Objective Direct Policy Search

Bei Reinforcement Learning (RL) handelt es sich um die Problemstellung, dass ein Agent, durch Interaktion mit seiner dynamischen Umgebung, selbständig eine optimale Strategie erlernt. Entscheidungen des Agenten basieren auf Trial-and-Error, wobei die getroffenen Entscheidungen je nach Auswirkung belohnt oder bestraft werden. RL kann in Methoden unterteilt werden die entweder den Raum der Nutzenfunktion (space of value function, z.B. Dynamische Programmierung) oder den Raum der Betriebsregeln (space of policies) absuchen.

EMODPS (Giuliani et al. 2014) zählt zu den Methoden die direkt den Raum der Betriebsregeln nach optimalen Lösungen durchsuchen. Dabei wird die Betriebsstrategie einer wasserwirtschaftlichen Anlage parametrisiert und mit einer geeigneten Funktion abgebildet (Direct Policy Search, Rosenstein und Barto, 2001). Flexible und nichtlineare Funktionen müssen angewendet werden, um zu verhindern, dass sich die Suche nach der optimalen Strategie auf einen Teil des Entscheidungsraums beschränkt, der möglicherweise nicht die optimale Lösung enthält. Untersuchungen (Giuliani et al. 2016) haben gezeigt, dass künstliche Neuronale Netze (kNN), deren Aktivierungsfunktion Radiale Basis Funktionen sind, die besten Ergebnisse liefern. Die Lösung des Optimierungsproblems beschränkt sich auf die Suche nach der optimalen Parametrisierung des kNN's und erfolgt mittels einer Evolutionsstrategie (Moriarty et al., 1999). Bei Evolutionsstrategien handelt es sich um stochastische Suchmethoden, die, in Anlehnung an die biologische Evolution, einen intelligenten Suchprozess nutzen. Für eine Trainingsperiode (Zeitraum beobachteter oder simulierter Daten (z.B. Zuflüsse oder Strompreise)) werden durch evolutionäre Prozesse neue Lösungsstrukturen erschlossen und mittels einer Fitnessfunktion bewertet.

---

<sup>1</sup> Pöyry Energy GmbH, Kranichberggasse 4, 1120 Wien, Tel.: +43 664 828 72 21, [christoph.libisch@poyry.com](mailto:christoph.libisch@poyry.com), [www.poyry.at](http://www.poyry.at)

Gute Lösungen werden durch die Prozesse Selektion und Vererbung an die nächste Generation von Parametern übergeben. Durch die Vermischung mit einer Zufallskomponente (Mutation) werden die Parameter verändert und neu bewertet. Abhängig von der Komplexität des Optimierungsproblems (Anzahl der Reservoirs, Anzahl von Nutzern, Mehrzielplanung, etc.) müssen geeignete Evolutionsstrategien ausgewählt werden, um eine optimale Lösung zu finden. Zatarain Salazar et al., 2016 haben gezeigt, dass für die Erstellung von Pareto-optimalen Lösungen auf moderne Evolutionsalgorithmen (z.B. Borg MOEA) zurückgegriffen werden sollte. Neben der Parametrisierung der Betriebsregeln und der Kopplung einer geeigneten Evolutionsstrategie ist die Auswahl der Variablen, auf deren Basis eine Entscheidung (z.B. Abgabe von Wasser) getroffen wird, von Bedeutung. Diesbezüglich werden Variablen, die den Systemzustand (z.B. Speicherinhalt und Zuflüsse) beschreiben, verwendet. Zusätzlich kann das kNN mit exogenen Informationen (z.B. Zuflussprognosen, Zeit, Bedarf, Klimavariablen, Strompreisen, etc.), die auf zukünftige Systemzustände schließen lassen, betrieben werden. Dadurch wird der Suchraum vergrößert und die Wahrscheinlichkeit die optimale Lösung zu finden steigt.

An dieser Stelle sei erwähnt, dass EMODPS das globale Optimum eines Problems nicht exakt findet. Die gefundene Lösung ist dem globalen Optimum jedoch immer sehr nahe. Insbesondere für Studien, die darauf abzielen die Performanz eines wasserwirtschaftlichen Systems für einen abgeschlossenen Zeitraum zu analysieren (Machbarkeitsstudien, Szenario-Analysen, etc.), kann EMODPS eine optimale Lösung approximieren. Betriebsregeln für den operationellen Betrieb eines Wasserspeichers können ebenfalls mittels EMODPS ermittelt werden. In diesem Fall muss jedoch bedacht werden, dass nicht alle möglichen Systemzustände (z.B. Kombination aus Systemvariablen wie Zufluss und Speicherinhalt) im Beobachtungszeitraum abgebildet werden. Um die erlernte Strategie gegen unbeobachtete Zustände robust zu gestalten, müssen in diesem Fall Zeitreihen durch stochastische Simulationen erweitert werden. Somit erlernt eine Strategie ein robustes Verhalten gegenüber Extremzuständen, die außerhalb der Beobachtungen liegen.

## Ergebnisse

Im Zuge des Vortrages werden ausgewählte Beispiele für die erfolgreiche Anwendung von EMODPS in der Wasserwirtschaft gezeigt und diskutiert.

## Literatur

- [1] Giuliani, M., Mason, E., Castelletti, A., & Pianosi, F. (2014). Universal approximators for multi-objective direct policy search in water reservoir management problems: a comparative analysis, 16, 10333.
- [2] Giuliani, M., Ph, D., Castelletti, A., Pianosi, F., Mason, E., Reed, P. M., & Asce, A. M. (2016). Curses, Tradeoffs, and Scalable Management: Advancing Evolutionary Multiobjective Direct Policy Search to Improve Water Reservoir Operations, 142(2). [http://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000570](http://doi.org/10.1061/(ASCE)WR.1943-5452.0000570).
- [3] Moriarty, D. E., Schultz, A. C., & Grefenstette, J. J. (1999). Evolutionary Algorithms for Reinforcement Learning, 11, 241–276.
- [4] Rosenstein, M. T., & Barto, A. G. (2001). Robot weightlifting by direct policy search. IJCAI International Joint Conference on Artificial Intelligence, 839–844.
- [5] Zatarain Salazar, J., Reed, P. M., Herman, J. D., Giuliani, M., & Castelletti, A. (2016). A diagnostic assessment of evolutionary algorithms for multi-objective surface water reservoir control. Advances in Water Resources, 92, 172–185. <http://doi.org/10.1016/j.advwatres.2016.04.006>